SEVENTH EDITION

ECONOMICS OF DEVELOPMENT



PERKINS RADELET LINDAUER BLOCK



Income status, 2011

Low income | Lower-middle income | Upper-middle income | High income

Source: World Bank, http://data.worldbank.org/about/country-classifications/country-and-lending-groups

LATIN AMERICA



Income status, 2011

Low income | Lower-middle income | Upper-middle income | High income

Source: World Bank, http://data.worldbank.org/about/country-classifications/country-and-lending-groups

AFRICA AND THE MIDDLE EAST

ECONOMICS OF DEVELOPMENT

Seventh Edition

ECONOMICS OF DEVELOPMENT

Seventh Edition

Dwight H. Perkins
Harvard University

Steven Radelet
U.S. Agency for International Development

David L. Lindauer
Wellesley College

Steven A. Block
Tufts University



W. W. Norton & Company
New York • London

W. W. Norton & Company has been independent since its founding in 1923, when William Warder Norton and Mary D. Herter Norton first published lectures delivered at the People's Institute, the adult education division of New York City's Cooper Union. The firm soon expanded its program beyond the Institute, publishing books by celebrated academics from America and abroad. By midcentury, the two major pillars of Norton's publishing program—trade books and college texts—were firmly established. In the 1950s, the Norton family transferred control of the company to its employees, and today—with a staff of four hundred and a comparable number of trade, college, and professional titles published each year—W. W. Norton & Company stands as the largest and oldest publishing house owned wholly by its employees.

Editor: Jack Repcheck
Assistant Editor: Hannah Bachman
Manuscript Editor: Candace Levy
Project Editor: Rachel Mayer
Electronic Media Editor: Nicole Sawa
Marketing Manager, Economics: Sasha Levitt
Production Manager: Sean Mintus
Permissions Manager: Megan Jackson
Permissions Clearing: Bethany Salminen
Text Design: JoAnn Simony
Art Director: Rubina Yeh
Composition: Jouve North America—Brattleboro, VT
Manufacturing: Quad Graphics—Taunton, MA

Copyright © 2013, 2006, 2001, 1996, 1992, 1987, 1983 by W. W. Norton & Company, Inc.

All rights reserved.
Printed in the United States of America.

Library of Congress Cataloging-in-Publication Data
Economics of development / Dwight H . Perkins . . . [et al.].—7th ed.
p. cm.
Rev. ed. of: Economics of development / Dwight H. Perkins, Steven Radelet, David L. Lindauer. 6th ed. c2006.
Includes bibliographical references and index.
ISBN 978-0-393-93435-9 (hardcover)—ISBN 0-393-93435-7 1. Developing countries—Economic policy. 2. Economic development. I. Perkins, Dwight H. (Dwight Heald), 1934- II. Perkins, Dwight H. (Dwight Heald), 1934- Economics of development.
HC59.7.E314 2013
338.9—dc23

2012029688

W. W. Norton & Company, Inc., 500 Fifth Avenue, New York, NY 10110-0017
wwnorton.com

W. W. Norton & Company Ltd., Castle House, 75/76 Wells Street, London W1T 3QT

1 2 3 4 5 6 7 8 9 0

Brief Contents

PART ONE Development and Growth

1	Patterns of Development	3
2	Measuring Economic Growth and Development	23
3	Economic Growth: Concepts and Patterns	55
4	Theories of Economic Growth	89
5	States and Markets	129

PART TWO Distribution and Human Resources

6	Inequality and Poverty	165
7	Population	217
8	Education	257
9	Health	299

PART THREE Macroeconomic Policies for Development

10	Investment and Savings	349
11	Fiscal Policy	391
12	Financial Development and Inflation	421
13	Foreign Debt and Financial Crises	455
14	Foreign Aid	499
15	Managing Short-Run Crises in an Open Economy	545

PART FOUR Agriculture, Trade, and Sustainability

16	Agriculture and Development	583
17	Agricultural Development: Technology, Policies and Institutions	619
18	Trade and Development	665
19	Trade Policy	709
20	Sustainable Development	757

Contents

Preface xv
International Development Resources on the Internet xxiii

PART ONE
Development and Growth

1 Patterns of Development 3

Three Vignettes 3 | *Malaysia* 3 | *Ethiopia* 5 | *Ukraine* 6 | *Development and Globalization* 8 | Rich and Poor Countries 10 | Growth and Development 13 | *Diversity in Development Achievements* 15 | Approaches to Development 16 | *The Study of Development Economics* 18 | Organization 19 | Summary 20

2 Measuring Economic Growth and Development 23

Measuring Economic Growth 24 | *Measuring GDP: What Is Left Out?* 25 | *Exchange-Rate Conversion Problems* 27 | Economic Growth around the World: A Brief Overview 32 | JARED DIAMOND: GUNS, GERMS, AND STEEL 34 | *Economic Growth, 1970–2010* 36 | What Do We Mean by *Economic Development*? 38 | *Measuring Economic Development* 40 | HUMAN DEVELOPMENT DEFINED 41 | WHY USE LOGARITHMS? 43 | *What Can We Learn from the Human Development Index?* 44 | *Millennium Development Goals* 46 | TARGETS OF THE MILLENNIUM DEVELOPMENT GOALS 47 | Is Economic Growth Desirable? 50 | Summary 53

3 Economic Growth: Concepts and Patterns 55

Divergent Patterns of Economic Growth since 1960 56 | BOTSWANA'S REMARKABLE ECONOMIC DEVELOPMENT 59 | Factor Accumulation, Productivity, and Economic Growth 60 | CALCULATING FUTURE VALUES, GROWTH RATES, AND DOUBLING TIMES 61 | Saving, Investment, and Capital Accumulation 64 | Sources of Growth Analysis 66 | Characteristics of Rapidly Growing Countries 74 | *1. Macroeconomic and Political Stability* 75 | *2. Investment in Health and Education* 77 | *3. Effective Governance and Institutions* 79 |

INSTITUTIONS, GOVERNANCE, AND GROWTH 80 | *4. Favorable Environment for Private Enterprise* 82 | *5. Trade, Openness, and Growth* 83 | *6. Favorable Geography* 84 | Summary 87

4 Theories of Economic Growth 89

The Basic Growth Model 91 | The Harrod-Domar Growth Model 94 | *The Fixed-Coefficient Production Function* 94 | *The Capital–Output Ratio and the Harrod-Domar Framework* 96 | *Strengths and Weaknesses of the Harrod-Domar Framework* 98 | ECONOMIC GROWTH IN THAILAND 101 | The Solow (Neoclassical) Growth Model 103 | *The Neoclassical Production Function* 103 | *The Basic Equations of the Solow Model* 104 | *The Solow Diagram* 108 | *Changes in the Saving Rate and Population Growth Rate in the Solow Model* 109 | POPULATION GROWTH AND ECONOMIC GROWTH 112 | *Technological Change in the Solow Model* 113 | *Strengths and Weaknesses of the Solow Framework* 116 | Diminishing Returns and the Production Function 117 | EXPLAINING DIFFERENCES IN GROWTH RATES 118 | *The Convergence Debate* 121 | Beyond Solow: New Approaches to Growth 125 | Summary 127

5 States and Markets 129

Development Thinking after World War II 130 | MARKET FAILURE 133 | Fundamental Changes in the 1970s and 1980s 137 | GHANA AFTER INDEPENDENCE 139 | THE DECLINING EFFECTIVENESS OF GOVERNMENT INTERVENTION IN THE MARKET: KOREA, 1960S–2010 143 | Structural Adjustment, the Washington Consensus, and the End of the Soviet Model 144 | *Soviet Command Model to Market Economies: The Great Transition* 148 | *Was the Washington Consensus a Success or Failure?* 153 | Summary 159

PART TWO

Distribution and Human Resources

6 Inequality and Poverty 165

Measuring Inequality 166 | Patterns of Inequality 172 | *Growth and Inequality* 174 | *What Else Might Cause Inequality?* 177 | Why Inequality Matters 178 | Measuring Poverty 180 | *Poverty Lines* 181 | NATIONAL POVERTY LINES IN BANGLADESH, MEXICO, AND THE UNITED STATES 183 | *Why $1.25 a Day?* 186 | *Dissenting Opinions on the Extent of Absolute Poverty* 191 | WHO IS *NOT* POOR? 192 | *Poverty Today* 192 | *Who Are the Poor?* 193 | *Living in Poverty* 195 | Strategies to Reduce Poverty 197 | *Growth is Good for the Poor* 198 | *Sometimes Growth May Not Be Enough* 200 | *Pro-Poor Growth* 201 |

WHY SHOULD DEVELOPMENT STRATEGIES HAVE A POVERTY FOCUS? 202 | *Improving Opportunities for the Poor* 205 | *Income Transfers and Safety Nets* 206 | Global Inequality and the End of Poverty 208 | Summary 214

7 Population 217

A Brief History of World Population 218 | *The Demographic Transition* 220 | *The Demographic Situation Today* 224 | TOTAL FERTILITY RATES 225 | *The Demographic Future* 227 | POPULATION MOMENTUM 229 | The Causes of Population Growth 231 | *Thomas Malthus, Population Pessimist* 232 | *Why Birth Rates Decline* 233 | Population Growth and Economic Development 236 | *Population and Accumulation* 237 | POPULATION GROWTH, AGE STRUCTURE, AND DEPENDENCY RATIOS 239 | *Population and Productivity* 241 | *Population and Market Failures* 243 | Population Policy 245 | *Family Planning* 246 | *Authoritarian Approaches* 249 | MISSING GIRLS, MISSING WOMEN 251 | Population Issues for the Twenty-First Century 253 | Summary 254

8 Education 257

Trends and Patterns 258 | *Stocks and Flows* 259 | *Boys versus Girls* 263 | *Schooling versus Education* 264 | Education as an Investment 267 | *The Rate of Return to Schooling* 269 | *Estimated Rates of Return* 272 | *First-Generation Estimates* 273 | ESTIMATING RATES OF RETURN FROM WAGE EQUATIONS 275 | *Second-Generation Estimates* 276 | *Puzzles* 278 | RETURNS TO SCHOOLING AND INCOME OPPORTUNITIES 279 | *Making Schooling More Productive* 281 | *Underinvestment* 282 | *Misallocation* 282 | *Improving Schools* 286 | *Reducing the Costs of Going to School* 287 | MEXICO'S PROGRESA 288 | *Inefficient Use of Resources* 290 | *It Is about More than the Money* 293 | COMBATING *TEACHER* ABSENCE 294 | Summary 298

9 Health 299

What Is Health? 302 | LIFE EXPECTANCY 305 | *Transitions in Global Health* 307 | *The Epidemiologic Transition* 308 | The Determinants of Improved Health 310 | Health, Income, and Growth 313 | *Income and Health* 314 | HOW BENEFICENT IS THE MARKET? A LOOK AT THE MODERN HISTORY OF MORTALITY 318 | *Health and Productivity* 319 | *Health and Investment* 320 | Three Critical Diseases 321 | MALARIA, YELLOW FEVER, AND THE PANAMA CANAL 322 | *HIV/AIDS* 323 | HIV/AIDS, MALARIA, AND TUBERCULOSIS: SOME BASICS 324 | *Malaria* 329 | MAKING MARKETS FOR VACCINES 330 | *Tuberculosis* 333 | What Works? Some Successes in Global Health 335 | *Preventing HIV/AIDS in Thailand* 336 | *Controlling Tuberculosis in China* 336 | *Eradicating Smallpox* 337 | *Eliminating Polio in Latin America* 338 | *Preventing Deaths from Diarrheal Disease* 340 | *Lessons Learned* 342 | Health Challenges 343 | Summary 344

PART THREE
Macroeconomic Policies for Development

10 Investment and Savings 349

Using Investment Productively: Cost-Benefit Analysis 351 | *Present Value* 351 | *Opportunity Costs* 355 | *Shadow Prices* 355 | *Welfare Weights* 356 | Barriers to Productive Public and Private Investment 357 | Barriers to Doing Business 359 | Foreign Direct Investment 363 | *FDI Patterns and Products* 364 | *Benefits and Drawbacks of FDI* 365 | *FDI and Growth* 369 | *Policies Toward Foreign Direct Investment* 370 | Savings 374 | Household Saving and Consumption 376 | Corporate Saving 379 | Government Saving 380 | Foreign Saving 384 | Summary 388

11 Fiscal Policy 391

Government Expenditures 393 | *Categories of Government Expenditures* 394 | REINING IN FISCAL DECENTRALIZATION IN BRAZIL AND CHINA 398 | *Government Revenue and Taxes* 399 | TAX RATES AND SMUGGLING: COLOMBIA 401 | *Taxes on International Trade* 401 | *Sales and Excise Taxes* 402 | *Personal and Corporate Income Taxes* 404 | *New Sources of Tax Revenues* 404 | *Changes in Tax Administration* 405 | *Fundamental Tax Reform* 405 | TAX ADMINISTRATION IN INDIA AND BOLIVIA IN THE 1980S 406 | INDONESIAN TAX REFORM 407 | Taxes and Income Distribution 411 | *Personal Income Taxes* 412 | *Taxes on Luxury Consumption* 413 | *Corporate Income and Property Taxes: The Incidence Problem* 414 | Economic Efficiency and the Budget 417 | *Sources of Inefficiency* 417 | *Neutrality and Efficiency: Lessons from Experience* 418 | Summary 420

12 Financial Development and Inflation 421

The Functions of a Financial System 423 | *Money and the Money Supply* 423 | *Financial Intermediation* 426 | *Transformation and Distribution of Risk* 426 | *Stabilization* 427 | Inflation 427 | *Inflation Episodes* 428 | HYPERINFLATION IN PERU, 1988–90 431 | Monetary Policy and Price Stability 432 | *Monetary Policy and Exchange-Rate Regimes* 433 | *Sources of Inflation* 435 | *Controlling Inflation through Monetary Policy* 438 | *Reserve Requirements* 439 | *Credit Ceilings* 439 | *Interest-Rate Regulation and Moral Suasion* 440 | *International Debt and Combating Recessions* 441 | Financial Development 442 | *Shallow Finance and Deep Finance* 443 | *Shallow Financial Strategy* 444 | *Deep Financial Strategy* 447 | *Informal Credit Markets and Micro Credit* 449 | DOES MICRO CREDIT REDUCE POVERTY? 451 | Summary 453

13 Foreign Debt and Financial Crises 455

Advantages and Disadvantages of Foreign Borrowing 458 | Debt Sustainability 459 | Debt Indicators 460 | From Distress to Default 463 | A SHORT HISTORY OF SOVEREIGN LENDING DEFAULT 465 | *The 1980s Debt Crisis* 466 | *Causes of the Crisis* 467 | *Impact on the Borrowers* 469 | *Escape from the Crisis, for Some Countries* 470 | The Debt Crisis in Low-Income Countries 473 | *Debt Reduction in Low-Income Countries* 474 | *The Heavily Indebted Poor Country Initiative* 475 | ODIOUS DEBT 476 | DEBT RELIEF IN UGANDA 479 | Emerging Market Financial Crises 480 | *Domestic Economic Weaknesses* 482 | *Short-Term Capital Flows* 484 | *Creditor Panic* 486 | MODEL OF SELF-FULFILLING CREDITOR PANICS 486 | *Stopping Panics* 489 | *Lessons from the Crises* 494 | Summary 496

14 Foreign Aid 499

Donors and Recipients 501 | *What Is Foreign Aid?* 501 | Who Gives Aid? 503 | THE MARSHALL PLAN 503 | THE COMMITMENT TO DEVELOPMENT INDEX 507 | *Who Receives Foreign Aid?* 512 | *The Motivations for Aid* 514 | CHINA'S FOREIGN AID 515 | Aid, Growth, and Development 518 | *View 1. Although Not Always Successful, on Average, Aid Has a Positive Impact on Economic Growth and Development* 520 | CONTROLLING RIVER BLINDNESS IN SUB-SAHARAN AFRICA 524 | *View 2. Aid Has Little or No Effect on Growth and Actually May Undermine Growth* 526 | FOOD AID AND FOOD PRODUCTION 528 | *View 3. Aid Has a Conditional Relationship with Growth, Stimulating Growth Only Under Certain Circumstances, Such as in Countries with Good Policies or Institutions* 533 | Donor Relationships with Recipient Countries 535 | *The Principal-Agent Problem* 536 | *Conditionality* 537 | *Improving Aid Effectiveness* 540 | Summary 543

15 Managing Short-Run Crises in an Open Economy 545

Equilibrium in a Small, Open Economy 546 | *Internal and External Balance* 547 | REAL VERSUS NOMINAL EXCHANGE RATES 550 | *The Phase Diagram* 553 | *Equilibrium and Disequilibrium* 556 | PIONEERING STABILIZATION: CHILE, 1973–84 559 | *Stabilization Policies* 560 | Applications of the Australian Model 564 | *Dutch Disease* 564 | RECOVERING FROM MISMANAGEMENT: GHANA, 1983–91 566 | *Debt Repayment Crisis* 567 | *Stabilization Package: Inflation and a Deficit* 569 | THE GREEK DEBT CRISIS OF 2010–12 571 | *Drought, Hurricanes, and Earthquakes* 574 | Summary 575 | Appendix to Chapter 15: National Income and the Balance of Payments 576

PART FOUR
Agriculture, Trade, and Sustainability

16 Agriculture and Development 583

Unique Characteristics of the Agricultural Sector 584 | Structural Transformation 587 | Two-Sector Models of Development 590 | *The Labor Surplus Model* 591 | SURPLUS LABOR IN CHINA 598 | *The Neoclassical Two-Sector Model* 599 | DEBATES OVER SURPLUS LABOR 602 | Evolving Perspectives on the Role of Agriculture in Economic Growth and Poverty Alleviation 604 | *Agriculture and Economic Growth* 604 | THE NUTRITION LINKAGE TO ECONOMIC GROWTH 608 | *Agriculture and Poverty Alleviation* 610 | *Agricultural Growth as a Pathway out of Poverty* 613 | Summary 617

17 Agricultural Development: Technology, Policies, and Institutions 619

Characteristics of Traditional Agriculture and Agricultural Systems 620 | *Agricultural Systems* 621 | Diagnosing the Constraints to Agricultural Development 622 | Raising the Technical Ceiling 627 | *The Green Revolution* 628 | *Recent Trends in Agricultural Productivity* 632 | A MODEL OF INDUCED TECHNICAL CHANGE IN AGRICULTURE 635 | Raising the Economic Ceiling 637 | *Food Production Analysis* 638 | *What to Produce? The Product–Product Decision* 638 | *How to Produce It? The Factor–Factor Decision* 641 | *How Much to Produce? The Factor–Product Decision* 643 | FERTILIZER SUBSIDIES IN MALAWI 645 | *Market Access* 651 | CELL PHONES AND AGRICULTURAL DEVELOPMENT 651 | Institutions for Agricultural Development 653 | Land Reform 656 | The World Food Crisis of 2005–08 658 | *Causes of the Crisis* 659 | *Consequences of the Crisis* 661 | Summary 663

18 Trade and Development 665

Trade Trends and Patterns 667 | *Who Trades?* 671 | Comparative Advantage 674 | *The Benefits of Trade* 677 | *Winners and Losers* 681 | Trading Primary Products 683 | EMPIRICAL EVIDENCE ON PRIMARY EXPORT-LED GROWTH 687 | *Export Pessimism* 688 | *Declining Terms of Trade?* 690 | *Dutch Disease* 693 | DUTCH DISEASE: A GEOMETRIC PRESENTATION 697 | NIGERIA: A BAD CASE OF DUTCH DISEASE 700 | INDONESIA: FINDING A CURE 702 | *The Resource Trap* 703 | *Breaking the Resource Curse* 705 | Summary 707

19 Trade Policy 709

Import Substitution 711 | Protective Tariffs 713 | *Import Quotas* 714 | EFFECTIVE RATES OF PROTECTION 715 | Trade Protection and Politics 718 | THE TWO-COUNTRY MODEL WITH A TARIFF 719 | *Production Subsidies* 720 | *Exchange-Rate Management* 722 | *Outcomes of Import Substitution* 724 | Export Orientation 725 | *Removing the Bias against Exports* 727 | *Favoring Exports* 728 | *Building Export Platforms* 730 | IS CHINA'S EXCHANGE-RATE POLICY UNFAIR? 731 | Trade Strategy and Industrial Policy 734 | Trade, Growth, and Poverty Alleviation 736 | *Trade Reforms and Poverty Alleviation* 739 | Key Issues on the Global Trade Agenda 741 | *Increased Global Competition and the Rise of China (and India)* 741 | *Does Outward Orientation Create Sweatshops?* 743 | LABOR ACTIVISTS AND LABOR OUTCOMES IN INDONESIA 746 | *Expanding Market Access* 747 | *Multilateral Trade Negotiations and the WTO* 750 | *Temporary Migration: Another Dimension of International Trade* 753 | Summary 755

20 Sustainable Development 757

Will Economic Growth Save or Destroy the Environment? 759 | Concept and Measurement of Sustainable Development 761 | *Saving for a Sustainable Future* 765 | THE MALTHUSIAN EFFECT OF POPULATION GROWTH ON ADJUSTED NET SAVINGS IN GHANA 768 | Market Failures 769 | *Externalities and the Commons* 770 | Policy Solutions 773 | *Property Rights* 773 | *Government Regulation* 774 | *Taxes, Subsidies, and Payments for Environmental Services* 776 | TAXING WATER POLLUTION IN COLOMBIA 777 | *Marketable Permits* 779 | *Informal Regulation* 781 | Policy Failures 782 | POLICY FAILURES AND DEFORESTATION IN INDONESIA 783 | *Poverty-Environment Linkages* 785 | *Global Climate Change* 792 | Summary 800

Index 803

Preface

In 1983, when the first edition of this textbook was published, 50 percent of the world's population lived in nations the World Bank classified as low income. By 2010 the number had dropped to 12 percent. Much of that change is the result of rapid economic growth in China and India. Today, both are middle-income economies. But economic growth and development has not been limited to these two Asian giants. "Africa Rising" was the cover story of a 2011 issue of *The Economist*, reflecting more than a decade of rapid growth in a region *The Economist* 10 years earlier referred to as "The Hopeless Continent." Throughout Africa, East and South Asia, Latin America, and elsewhere, dramatic improvements have been taking place in the education, health, and living standards of billions of people.

The study of the economics of development has had to keep pace with these historic changes. We have tried to keep pace as well. In this as in previous editions, we have incorporated new ideas and new data and provide fresh insights from the experiences of the nations that make up the developing world. While there is much that is new in this seventh edition, the distinguishing features of this text remain the same:

- It is based primarily on the real-world experiences of developing countries. It explores broad trends and patterns and uses numerous real-country examples and cases to illustrate major points, many of which are drawn from the authors' own experiences.
- It draws heavily on the empirical work of economists who believe that attention to the data not only reveals what the development process entails but permits us to test our beliefs about how that process works.
- It relies on the theoretical tools of neoclassical economics to investigate and analyze these real-world experiences in the belief that these tools contribute substantially to our understanding of economic development.
- It highlights the diversity of development experience and recognizes that the lessons of theory and history can be applied only within certain institutional and national contexts.

As in previous editions, the seventh edition of *Economics of Development* is intended to be both accessible and comprehensive. The discussion is *accessible* to those students, whether undergraduates or those pursuing advanced degrees in international relations, public policy, and related fields, who have only an elementary background in economics. At the same time, the text provides a *comprehensive* introduction to all

students, including those with significant training in economics, who are taking their first course in development economics.

Major Changes for the Seventh Edition

The seventh edition of *Economics of Development* continues and extends the major revisions of the book initiated in previous editions. The substantial changes reflect the contributions of three new coauthors—Steven Radelet in the fifth edition, David Lindauer in the sixth, and Steven Block in the seventh—working alongside original coauthor Dwight Perkins. The seventh edition features fundamental revisions of many chapters. The revised chapters take full advantage of research on development economics over the past decade. In addition, there are more and better tables, charts, and other exhibits chronicling the lessons and remaining controversies of the development field. The chapter summaries that follow highlight the major changes in the seventh edition. Asterisks indicate chapters that are essentially or entirely new for this edition.

Chapter 1 (Patterns of Development) has been condensed. It begins with the three vignettes—on Malaysia, Ethiopia, and Ukraine—that were introduced in the previous edition and updated for this one. The chapter includes a new table on the classification of world economies and populations according to income status. Also added is a section on how the study of development economics differs from the study of economics as applied to developed nations.

Chapter 2 (Measuring Economic Growth and Development) reflects important updates by various agencies and authors, including the 2005 International Comparison Program (ICP) estimates of purchasing power parity; Angus Maddison's latest and last estimates of world economic growth; and the United Nations Development Programme's (UNDP) 2010 revision of the human development index (HDI). The analysis of economic growth and happiness has been fully revised and expanded, reversing some of the conclusions first reported in the sixth edition. A new box has been added on the determinants of long-run economic growth suggested by Jared Diamond's *Guns, Germs, and Steel*. There is also a new box on the use of logarithms, essential to understanding the HDI and other development measures.

Chapter 3 (Economic Growth: Concepts and Patterns) has been reorganized and updated to include the latest data and country examples. This edition features a new box on the calculation of growth rates, future values, and doubling times. Previous material on the characteristics of production functions and on growth convergence has been moved to Chapter 4. The previous discussion of structural change has been entirely rewritten and relocated to Chapter 16, where it supports the discussion of economic dualism.

Chapter 4 (Theories of Economic Growth) has also been streamlined and consolidated. This edition integrates into its presentation of growth theory the discussions of production functions and growth convergence previously found in Chapter 3. The previous discussion of dual-sector growth models has been eliminated from the discussion of growth theory and relocated to Chapter 16, where its primary purpose is to illustrate sectoral interactions as a foundation for discussing the role of agriculture in development. Chapter 4 also provides newly updated data and illustrative figures.

***Chapter 5 (States and Markets)** raises the central question, What makes economic development happen? For this edition, senior author, Dwight Perkins, takes a fresh look, tracing the evolution of thinking about development from Adam Smith, through notions of The Big Push advanced in the 1940s by economist Paul Rosenstein-Rodan, to more recent debates over Structural Adjustment and the Washington Consensus.

Chapter 6 (Inequality and Poverty) updates the analysis of poverty by examining the revision of the global poverty line from $1 per day to $1.25 per day. Combined with recent measures of purchasing power parity (PPP), the latest estimates on levels of both inequality and poverty are presented. A new section, "Living in Poverty," has been added that includes insights from the work by economists Abhijit Banerjee, Esther Duflo, and others on the economic lives of the poor. The use of conditional cash transfers also receives more attention.

Chapter 7 (Population) incorporates the United Nations 2010 Revision to its world population projections. The chapter now includes more discussion of the demographic dividend and a new section on population issues for the twenty-first century.

Chapter 8 (Education) benefits from recent revisions to the Barro-Lee data set on school attainment and from recent results of the Organisation for Economic Co-Operation and Development's (OECD) Programme for International Student Assessment (PISA). Fuller use is made of econometric approaches, including natural experiments and randomized controlled trials (RCTs) in determining rates of return to schooling and the effectiveness of alternative interventions to improve learning outcomes. A new box on combating teacher absence has been added.

Chapter 9 (Health) now includes an extended discussion of the relationship between income and health. The Preston curve, showing the relationship between life expectancy and per capita income is presented and the debate over causality more fully developed. A box has been added on creating markets for vaccines for diseases that primarily affect populations in low-income settings.

***Chapter 10 (Investment and Savings)** draws on material from two chapters in the sixth edition. The chapter assumes students have had some background in the principles of macroeconomics and focuses on topics central to developing nations. These include barriers to both public and private investment and alternative sources of

savings to finance productive investments. Special attention is given to foreign direct investment and its role in promoting economic growth.

Chapter 11 (Fiscal Policy) continues to focus on the key components of government expenditures and revenues. Data have been updated and a box added on the challenges of fiscal decentralization in Brazil and China.

Chapter 12 (Financial Development and Inflation) reorders some of the material on monetary policy and price stability, providing a more intuitive flow to the material. The discussion of microcredit has been expanded and includes a new box based on an RCT in India.

Chapter 13 (Foreign Debt and Financial Crises) is updated to provide the most recent data on foreign debt and provides current examples of financial crises (including the 2010–11 Euro Zone crisis, though the focus remains on developing countries).

Chapter 14 (Foreign Aid) has been updated to reflect recent trends in official development assistance, which, in part, are the results of events in Afghanistan, Iraq, and Pakistan. New boxes have been added on the commitment to development index and on Chinese foreign aid.

Chapter 15 (Managing Short-Run Crises in an Open Economy), formerly Chapter 21, now concludes the section on macroeconomic policies for developing countries. Presentation of the Australian model of short-run macroeconomic management has been clarified and illustrated with updated examples. New boxes cover real versus nominal exchange rates and the Greek debt crisis of 2010–11 (with an application of the Australian model). This chapter also features a new appendix for this edition, providing a review of national income and balance of payments accounting.

***Chapter 16 (Agriculture and Development)** is the first of two entirely new chapters on agriculture. This chapter places agriculture in its broad developmental context, emphasizing the potential contributions of agriculture to both growth and poverty alleviation. Specific topics include structural transformation, dual-sector growth models, agriculture and growth, and agriculture as a pathway out of poverty.

***Chapter 17 (Agricultural Development: Technology, Policies and Institutions)** builds on the broad discussion of agriculture and development in the previous chapter, concentrating on policies and institutions to promote agricultural development. Specific topics introduced in this chapter include a typology of agricultural systems common in developing countries, a broad framework for analyzing constraints to agricultural production growth, the role of technical change and the green revolution, the role of institutions and land reform in agricultural development, and a review of the global food price crisis of 2005–08.

***Chapter 18 (Trade and Development)** presents an overview of trends and patterns in world trade. It reviews the theory of comparative advantage, discussing both

the benefits of trade and its distributional consequences. Special attention is paid to trade in primary products, including export pessimism and the terms of trade, Dutch disease and the real exchange rate, and the resource curse and responses to it.

***Chapter 19 (Trade Policy)** builds on the broad discussion of trade and development in the previous chapter. It reviews import substitution as a trade strategy and the consequences of trade protection. This is followed by discussion of export orientation, including experience with export processing zones. Evidence is presented on trade, growth, and poverty alleviation. The chapter concludes with an examination of key issues on the global trade agenda, such as the impact of China and India on global trade competition, sweatshops and labor standards, the Doha Round of trade negotiations, and temporary labor migration as a strategy to alleviate world poverty.

***Chapter 20 (Sustainable Development)** is essentially a new treatment of the subject, retaining from the previous edition only the discussion of market and policy failures. New topics addressed in this edition are the environmental Kuznets curve hypothesis, an expanded and more analytical treatment of the concept and measurement of sustainable development, institutional perspectives on externalities (drawing on the work of Elinor Ostrom), payments for environmental services as a response to externalities, a substantially expanded treatment of poverty-environment linkages, and a new section on the economics of climate change. New boxes in this edition cover the Malthusian effect of population growth on adjusted net savings in Ghana, taxation of water pollution in Colombia, and, policy failures and deforestation in Indonesia.

About the Authors

Of the four original authors of *Economics of Development* only Dwight Perkins remains as an active contributor to this edition. Michael Roemer's death in 1996 took from the development field one of its most thoughtful and productive writers and practitioners. Mike, in many ways, was the single most important contributor to the earlier editions, and his legacy endures in this edition. Malcolm Gillis, an expert in issues of public finance and economic development, played the central role in getting this book started in the early 1980s. He later went on to a distinguished career as president of Rice University, from which he retired. Donald Snodgrass was responsible for Part Three, "Human Resources," for the first five editions. Both Malcolm Gillis's and Donald Snodgrass's strong contributions are evident in the current edition as well. The new authors are privileged to be part of a text that, thanks to the scholarship of the original authors, helped to define the field of development economics.

Dwight H. Perkins is the H. H. Burbank Professor of Political Economy Emeritus at Harvard University and former director of the Harvard Institute for International Development. Professor Perkins is a leading scholar on the economies of East and Southeast Asia. Professor Perkins's legacy is contained not only in the many chapters

he has contributed to *Economics of Development* and in his many scholarly books and articles but also in the thousands of students he has taught over his distinguished academic career (including all of his current coauthors!).

Steven Radelet joined *Economics of Development* for its fifth edition. At the time he was a fellow at Harvard's Institute for International Development and taught in both Harvard's economics department and the Kennedy School of Government. He subsequently was deputy assistant secretary of the U.S. Treasury for Africa, the Middle East, and South Asia; a Senior Fellow at the Center for Global Development; and Senior Advisor on Development for Secretary of State Hillary Clinton. He is an expert on foreign aid, developing country debt and financial crises, and economic growth and has extensive experience in West Africa and Southeast Asia. He currently serves as Chief Economist for the U.S. Agency for International Development (USAID). In that capacity he was unable to contribute to this edition but his prior work on the textbook significantly informs this edition as well.

David L. Lindauer is the Stanford Calderwood Professor of Economics at Wellesley College, where he has taught since 1981. He has frequently served as a consultant to the World Bank and was a faculty associate of the Harvard Institute for International Development. Professor Lindauer's area of expertise is in labor economics. His research and policy advising has included work on industrial relations, labor costs and export potential, minimum wages, poverty and unemployment, public sector pay and employment, and racial affirmative action. He has worked on labor market issues in East and Southeast Asia, Sub-Saharan Africa, and elsewhere. Professor Lindauer, an award-winning teacher of economics, brings his considerable experience teaching undergraduates to this edition.

Steven A. Block is Professor of International Economics and head of the International Development Program at the Fletcher School of Law and Diplomacy, Tufts University. He joins *Economics of Development* beginning with this edition and has been teaching development economics at the Fletcher School since 1995. Professor Block also holds a faculty appointment at the Friedman School of Nutrition Science and Policy at Tufts University, and has been a visiting scholar at the Harvard University Center for International Development and at Harvard's Weatherhead Center for International Affairs. He has published numerous scholarly articles in the areas of agricultural development and political economy, and worked extensively on policy advisory teams across Sub-Saharan Africa and in Southeast Asia.

Acknowledgments

Any textbook that makes it to a seventh edition accumulates many debts to colleagues who read chapters, provided feedback, or contributed in some way to the success and longevity of the work. We owe many thanks to many people. In these acknowledgments, we wish to thank those individuals who contributed to this edition.

Dwight H. Perkins is grateful to the hundreds of colleagues and students from around the developing world and at Harvard and other universities in the United States and elsewhere who, over the past five decades, have taught him what he knows about development economics and to his wife, Julie, who has joined him on many of his trips to developing countries.

David L. Lindauer thanks his research assistants, Yue Guan and Teju Velayudhan. They did a tremendous job creating the charts and figures in many of the chapters. Dana Lindauer, Pasinee Panitnantanakul, and Anisha Vachani provided additional assistance. Thanks are owed David Johnson and Joseph Stern for help in drafting several of the text Boxes. He also received excellent comments from Akila Weerapana (Wellesley College), Jere Behrman (Pennsylvania), Lant Pritchett (Harvard), Martin Ravallion (World Bank), and Paul Glewwe (Minnesota). He greatly appreciates the sabbatical leave provided by Wellesley College that provided the time needed to produce this edition and wishes to thank his family for all their support.

Steven A. Block thanks his research assistant, Bapu Vaitla, for his impeccable support and useful suggestions. He is particularly grateful to Peter Timmer (emeritus, Harvard University) and Jeffrey Vincent (Duke University) for their critical reading and constructive suggestions for Chapters 16, 17, and 20 of this edition. In addition, he is grateful to his family—Avi; Ruthie; and wife, Maria—for their love and patience.

All three of us wish to thank everyone at W. W. Norton and Company for their continued support. We are especially grateful for the continued guidance and efforts of our editor, Jack Repcheck.

D.H.P. *Cambridge*
S.R. *Washington, D.C.*
D.L.L. *Wellesley*
S.A.B. *Tufts*

International Development Resources on the Internet

International Organizations

1. The World Bank (*www.worldbank.org*) hosts specific sites dedicated to country information (*www.worldbank.org/html/extdr/regions.htm*), data on a range of development indicators (*www.worldbank.org/data*), specific development themes (*www.worldbank.org/html/extdr/thematic.htm*), poverty reduction (*www.worldbank.org/poverty*), and governance and anticorruption (*www.worldbank.org/wbi/governance*).
2. The International Monetary Fund (*www.imf.org*) hosts individual country information (*www.imf.org/external/country/index.htm*) and an index of over 100 economic, commodity, and development organizations (*www.imf.org/np/sec/decdo/contents.htm*).
3. African Development Bank (*www.afdb.org*), Asian Development Bank (*www.adb.org*), and Inter-American Development Bank (*www.iadb.org*).
4. The United Nations development organizations, including the United Nations Development Programme (*www.undp.org*), the Food and Agriculture Organization (*www.fao.org*), the World Health Organization (*www.who.org*), the United Nations Children's Fund (*www.unicef.org*), the Joint United Nations Programme on HIV/AIDS (*www.unaids.org*), and the United Nations Millennium Project with information on the Millennium Development Goals (*www.unmillenniumproject.org*).

Independent Research Organizations

5. The Center for Global Development (*www.cgdev.org*).
6. The Center for International Development at Harvard University (*www.hks.harvard.edu/centers/cid*).
7. The Earth Institute at Columbia University (*www.earthinstitute.columbia.edu*).
8. The Overseas Development Institute (*www.odi.org.uk*).
9. The World Institute for Development Economics Research (*www.wider.unu.edu*).
10. The World Resources Institute (*www.wri.org*).

Information Gateways

11. The Development Gateway (*www.developmentgateway.org*).
12. Institute of Development Studies (*www.ids.ac.uk*).
13. The International Development Research Centre (*www.idrc.ca*).
14. Netaid.org (*www.netaid.org*).
15. Oneworld.net (*www.oneworldgroup.org*).
16. International Economics Network, Development Resources (*www.internationaleconomics.net/development.html*)

Data Sources

In addition to the other sites listed, the following offer useful data.

17. The Center for International Comparisons (Penn World Tables, *www.pwt.econ.upenn.edu*).
18. The Development Assistance Committee of the Organization for Economic Cooperation and Development (*www.oecd.org/dac*).
19. The Living Standards Measurement Study of Household Surveys (*www.worldbank.org/LSMS*).
20. The Roubini Global Economics Monitor (*www.rgemonitor.com*).
21. The World Factbook (*www.cia.gov/cia/publications/factbook*).

Foundations

22. The Bill and Melinda Gates Foundation (*www.gatesfoundation.org*).
23. The Ford Foundation (*www.fordfound.org*).
24. The Open Society Institute (*www.soros.org*).
25. The Rockefeller Foundation (*www.rockfound.org*).
26. The William and Flora Hewlett Foundation (*www.hewlett.org*).

PART ONE

Development and Growth

1 Patterns of Development

THREE VIGNETTES[1]

MALAYSIA

During the early 1980s, when she was 17 years old, Rachmina Abdullah did something no girl from her village had ever done before. She left her home in a beautiful but poor part of the state of Kedah in Malaysia, where people grew rice in the valleys and tapped rubber trees in the nearby hills, and went to work in an electronics plant in Penang, 75 miles away. Rachmina's family was poor even by the modest standards of her village, and her parents welcomed the opportunity for their daughter to earn her own keep and possibly even send money back to help them feed and clothe the family, deal with recurrent emergencies, and raise their five younger children. With these benefits in mind, they set aside their reservations about their unmarried daughter's unheard-of plan to go off by herself to work in the city.

[1]The three narratives that follow are fictional. The vignette on Malaysia is loosely based on Fatimah Daud, *Minah Karan: The Truth about Malaysian Factory Girls* (Kuala Lumpur: Berita Publishing, 1985), and Kamal Salih and Mei Ling Young, "Changing Conditions of Labour in the Semiconductor Industry in Malaysia," *Labour and Society* 14 (1989), 59–80. The narratives on Ethiopia and Ukraine are based on discussions with experts who have lived and worked in these nations. The named individuals are constructs rather than actual people. Data used in all three vignettes are from *World Development Indicators Online*.

Rachmina got a job assembling integrated circuits in a factory owned by a Japanese company. Every day, she patiently soldered hundreds of tiny wires onto minute silicon chips. It was tedious, repetitive work that had to be performed at high speed and with flawless accuracy. From a long day of hard work with few breaks, Rachmina could earn the equivalent of a few dollars. Because their wages were low, Rachmina and her colleagues welcomed opportunities to work overtime. Often they put in two or three extra hours a day, for up to seven days a week. They particularly liked working Sundays and holidays, when double wages were paid. Rachmina shared a small house in a squatter area with seven other factory workers. By living simply and inexpensively, most of the young women managed to send money back home each month and generally enjoyed the unfamiliar freedom of living apart from their families and villages.

Five years later, Rachmina, who had accumulated some savings, decided it was time to return to her village, where she soon married a local man and settled down. She later had two children—fewer than her friends who had stayed in the village and married earlier. Her savings helped provide for her family, and she was able to enroll her children in the local school.

Rachmina's chance to work in an electronics factory came about because, starting in the 1970s, American and Japanese electronics manufacturers were moving into export processing zones (EPZs) established by the Malaysian government. The national unemployment rate was high, and the government was particularly anxious to find more urban, nonagricultural jobs for the indigenous Malay population. In the mid-1970s, demand for electronic devices was growing by leaps and bounds, and international firms were looking for overseas locations where they could carry out parts of their operations at lower cost. The first beneficiaries of this migration were the newly industrializing nations of East Asia: Hong Kong, Singapore, South Korea, and Taiwan. Malaysia, with its good infrastructure, English-speaking workforce and stable political environment, also attracted foreign investors. Although the wages were lower than those paid in Japan and the United States, they were much higher than most Malays could earn through farm work, and people lined up for the opportunity to secure these prized jobs.

Malaysia, which previously had been known mainly for the export of rubber, tin, and palm oil, became one of the world's largest exporters of electronic components and other labor-intensive manufactured goods. Partly because of these exports, Malaysia emerged as one of the fastest growing economies in the world and a leading development success story. The income of the average Malay more than *quadrupled* in real terms between 1970 and 2010, infant mortality fell from 41 to 6 infants per thousand, and life expectancy rose from 61 to 75 years. Adult literacy jumped from 58 to 92 percent and the ratio of girls to boys enrolled in school increased from 83 to 103 percent. (In other words, if Rachmina had grandchildren, today her granddaughters would be slightly *more likely* to attend school than would her grandsons.)

The Malaysian economy changed as well. Agriculture accounted for about one third of national output in 1970; today it accounts for less than 10 percent. Households in Kedah still grow rice, and young women still work in electronics assembly

in Penang as they did in the 1970s. (If you own a Dell laptop, it very well may have been assembled in Penang.) But Penang must now look to the future. Low wage competition from Vietnam and elsewhere in Southeast Asia is attracting the electronic assembly plants that once came to Malaysia. Penang is hoping to develop a more knowledge-based economy, which might include biotechnology, business-process outsourcing, and medical tourism.[2] Rachmina's grandchildren in all likelihood will live in a Malaysia far different from the one she knew.

ETHIOPIA

On another continent and at about the same time that Rachmina was on her way to begin working in Penang, Getachew was born in Ethiopia. Getachew's family and many of his relatives lived in a rural area outside of Dese, a drought-affected area in Amhara region and a day's bus ride from the capital city of Addis Ababa. The family lived in a thatched hut and had few possessions. They owned cooking utensils, some blankets and clothing, a radio, and a bicycle. Getachew's sisters spent two hours a day fetching water from a small stream outside their village. The village did not have a paved road or electricity. In addition to growing tef, a cereal crop similar to millet, the family grew vegetables and relied on its own production for most of its consumption needs. The family was especially proud of their livestock. Getachew's father raised and traded oxen, which earned the family most of their meager cash income.

Getachew was the fifth of eight children, one of whom died at birth and another before her third birthday. Getachew received five years of schooling, but the years were not in succession. In some years, he needed to tend the family's crops and livestock with his father and brothers. In other years, the family did not have enough money to pay for uniforms and other school fees and could afford to send only one or two of their children to school. Priority was given to Getachew's older brothers. By the time he was 16, Getachew was able to read and write, although not well.

Getachew and his family have known hard times. His mother died shortly after the birth of her last child. This was due, in part, to her weakened state as a result of the drought and famine in 1984, compounded by multiple births and lack of emergency postpartum care. Despite worldwide attention to Ethiopia's plight that year, relief came too late to help them. The area has been affected by drought and shortages of food since then, but none as severe. Political transition in 1991 brought a lot of uncertainty, even to the countryside. The village school remained closed that year, as the teacher returned to live in the capital. Prices for most commodities went up at the same time that Getachew's father earned little for his oxen. In 1998, war broke out between Ethiopia and neighboring Eritrea. Getachew was staying with his brother in Addis at the time and escaped the draft, but several of his friends were conscripted to serve. One lost a

[2]Homi Kharas, Albert Zeufack, and Hamdan Majeed, *Cities, People & the Economy: A Study on Positioning Penang*, Khazanah Nasional Berhad and the World Bank (Kuala Lumpur, 2010).

leg during the war and returned home but was no longer much help tending livestock. Another contracted HIV/AIDS and, without treatment, died soon after.

Getachew's second oldest brother is a truck driver and occasionally provides the family with goods and cash. Getachew went with his brother to Addis Ababa and lived there for a time, finding only occasional casual day labor. Life was hard in the city, in some ways harder than in the countryside. Back home everyone lived similarly. In Addis, many people had money to spend while Getachew did not. When his father fell ill from tuberculosis, Getachew went back to help out on the farm. He would like to get married but land has become increasingly scarce in his village, and it is unclear when he will be able to support a family of his own.

Getachew's life has been much like his father's and parallels that of most Ethiopians and many Africans. Per capita incomes in 2004 were at about the same levels as in 1981. In the intervening years, incomes at times increased and at other times declined, but overall, economic stagnation characterized the nation. Since 2004, however, economic growth has been faster and more consistent, averaging 6.6 percent per year. This is much faster than at any time over the past three decades. Despite the global recession of 2008–09, Ethiopia's economy continues to grow rapidly, although it is hard to know if this will be sustained.

Looking at other indicators of living standards, infant mortality rates fell from an estimated 136 per thousand in 1970 to 67 per thousand in 2009, reflecting the potential for health outcomes to improve even when income does not. Life expectancy, at 56 years, is 13 years more than in 1970 but still well below the levels in Malaysia and other more affluent economies. Adult literacy is less than one out of three, but this will improve in the future. Four out of every five of Ethiopia's children of school age are now enrolled in primary school—double the level of a decade ago. The economy is changing too, albeit slowly. In 1970, 61 percent of national output was derived from agriculture; today this figure is 51 percent. Growing crops and tending livestock remain the main economic activities for Getachew and three quarters of all Ethiopians.

With more than 80 million people, Ethiopia is one of the most populous poor nations in the world. But it is also one of Africa's emerging economies. Exports of primary products (including coffee), remittances from Ethiopians working abroad, and foreign direct investment are all fueling the nation's recent growth. Underlying these trends are better economic policies and management; the spread of new technologies, including cell phones; and increasing economic relations with China, India, and the Middle East.

UKRAINE

Unlike Getachew or Rachmina, Viktor and Yulia are relatively well educated. Both were born in L'viv in western Ukraine, about 300 miles from the capital, Kyiv. They graduated from secondary school in 1980 and went on to study for several more years

at the local polytechnic institute, which is where they met. Viktor studied engineering, and Yulia architectural drawing. After finishing their studies, they married, and Viktor began work at a local glass factory. Yulia was hired by a municipal agency. As was common during the Soviet era, the couple moved into the one-bedroom apartment where Yulia's parents lived. Viktor and Yulia owned a refrigerator and other kitchen appliances, a television, furniture, some musical instruments, many books, and a telephone. They went on vacations, often going to a state-subsidized "sanatorium" in the Carpathian Mountains in Ukraine's southwest. Their daughter, Tetiana, was born in 1986, and Yulia was able to take a paid maternity leave.

Viktor's and Yulia's lifestyle in the 1980s was certainly modest by American or western European standards. They enjoyed few luxuries but most of their everyday needs were met. Often this required standing in long lines at government stores for staples such as bread, cooking oil, milk, and sugar. They also had their own garden plot where they grew flowers, fruit, and vegetables. Occasionally they purchased Polish goods in the gray market (technically illegal but not enforced). Healthcare and daycare were publicly provided.

Like many other ethnic Ukrainians in L'viv, Viktor, Yulia, and Yulia's parents longed for national independence. They spoke and maintained their native language even though the official language of the Soviet Union was Russian. Beyond their nationalist leanings, they thought their lives would be better in a less-centralized economy but were unaware of the severe consequences the breakup of the Soviet Union would entail.

Ukraine became independent in December 1991, with 90 percent of voters in support of a referendum on independence. Street celebrations and emotional speeches about freedom marked the event. But independence also had negative consequences. Trade with Russia collapsed and, with it, so did orders at the glass factory where Viktor was employed. Viktor was paid less and less often. A common refrain heard throughout the former Soviet Union was "they pretend to pay us and we pretend to work." Ukraine relied on energy supplies from Russia, but without foreign exchange to pay for them, Russian fuel exports dwindled, and many Ukrainians had to endure cold winters without much heat in their homes or offices. Mismanagement of the domestic economy led to hyperinflation in 1993–94, with prices rising almost 5,000 percent. The hyperinflation destroyed the purchasing power of pensioners, like Yulia's aging parents, and others living on fixed incomes. The healthcare system fell apart. Medicine, at times, had to be obtained on the black market, and one could never be sure of its efficacy. Life became harsher, with more anxiety, stress, and uncertainty about the future. Life expectancy for Ukrainian men was 66 years in 1989 but fell to 62 years by 1995. In 2009, it was still lower than two decades earlier, at 64 years.

Ukrainians had hoped that after independence foreign investment would flow into their country. It did not. Foreigners looked at Ukraine's situation and found existing technology backward, products of poor quality, and corruption rife. Instead of foreign purchase of factories, company officials often stripped factories of whatever assets they

had and kept the proceeds themselves. Per capita income in 1998 was only 40 percent of its pretransition peak in 1989. All of this was happening in a society in which adult literacy was universal and where the nation's boys *and* girls were all well educated.

Viktor was one of many Ukrainian men who had a difficult time with the transition. He could not adjust to changing circumstances and never found a new job. He spent a lot of time at home, doing some carpentry and other odd jobs now and then. His health was poor as a result of smoking too much and, Yulia believes, the environmental hazards of the glass factory. Many of Viktor's friends from the factory have similar health problems; a few have died prematurely. Yulia is holding the family together. She has reinvented herself. She still is employed at the municipal agency, although her wages are not always paid. She spends much of her time at work drawing plans for some of the newly rich Ukrainians who are building summer villas and renovating apartments. She prefers not to discuss where the money to pay for these villas, and for her services, is coming from.

Ukraine's economy rebounded during the 2000s, and Viktor and Yulia were starting to feel more optimistic about the prospects for their daughter. By 2008 income per capita had risen to three quarters of its level in 1989. But the global recession hit Ukraine hard, in part because of a sharp decrease in demand for steel, one of the nation's major exports. A weak domestic banking sector also proved vulnerable to the world financial crisis. The economy shrank by 15 percent in 2009. (By way of comparison, during the Great Recession the U.S. economy contracted by only 2.5 percent.) Although the economy improved again in 2010, the nation is plagued by political turmoil and poor governance. Problems of corruption, cronyism, and unresponsive public institutions remain unresolved.

DEVELOPMENT AND GLOBALIZATION

These three development vignettes, about Malaysia, Ethiopia, and Ukraine, are meant to capture the range of experiences of individual nations over the past three to four decades. Some nations, including Malaysia, have experienced historically unprecedented rates of economic growth, which have dramatically changed the lives of their populations. In other parts of the world, including Ethiopia and much of sub-Saharan Africa, economic growth, at least until fairly recently, has been minimal, and standards of living have changed far less from one generation to the next. A third group of nations experienced a fundamental transition from one economic system to another. In many nations, including Ukraine, this resulted in an abrupt and steep decline in living standards. Recovery occurred in some but not all areas during the 2000s; the global financial crisis in 2008–09 hit some nations harder than others. Understanding the causes and consequences of these different patterns of economic development is the central goal of this textbook.

Economic growth, stagnation, and transition have had a profound impact on the lives of Rachmina, Getachew, and Viktor and Yulia, respectively, and on the more

than 5.6 billion people of the developing countries these individuals are meant to personify. As different as the outcomes have been across nations, all countries have been affected by dramatic changes both within their borders and outside of them.

- Political systems have undergone profound changes, especially since the end of the cold war. Many low-income countries have adopted more democratic political systems since the early 1990s. The relationship between these political changes and the process of economic development and poverty reduction remains a matter of considerable debate.
- Substantial demographic shifts have led to a fall in population growth rates in many countries, with a decline in the number of dependent children and a corresponding growth in the share of workers in the population. Looking forward, many low-income countries will soon see large segments of the population reaching retirement age, with important implications for savings, tax revenues, pension systems, and social programs.
- The spread of endemic disease, including the HIV/AIDS pandemic, threatens development progress in many countries. In more than half a dozen African countries, more than a quarter of the adult population is HIV-positive. HIV/AIDS, malaria, tuberculosis, and other diseases bring a heavy human toll and substantial economic costs.
- Global trade has grown rapidly in line with sharply falling transportation and communication costs, giving rise to far more sophisticated global production networks. Instead of products being made start to finish in one location, firms in one country specialize in one part of the production process, while firms in another country play a different role. There has been a dramatic shift away from producing goods for the local market under government protection toward greater integration with global markets.[3]
- Capital moves much more quickly across borders than it did several decades ago. More sophisticated financial instruments and a greater emphasis on private capital have opened opportunities for low-income countries to access foreign capital for local investment. In some countries rapid financial liberalization resulted in deep financial crises when local financial institutions were weak and foreign capital was quickly withdrawn. On the other hand, the financial crisis of 2008–09 that originated in the United States and other advanced economies had far less of an impact on many developing economies than was expected.

[3]Malaysia has been most obviously helped by this process, and, at least in the short run, Ukraine has been hurt. Even rural Ethiopians who engage in subsistence agriculture are not insulated from global economic events. The Asian financial crisis of the late 1990s provides an example. The crisis resulted in a sharp decline in global demand for shoes, handbags, and other goods made from leather. This in turn lowered the demand and price for animal hides, a traditional Ethiopian export, and reduced the cash income of rural Ethiopians who might have had no idea why the prices they received for their animal skins had fallen.

- Information and ideas spread much faster around the globe than in earlier times. Cell phones, the Internet, and other communications technology have created new opportunities for low-income countries. Farmers can get pricing information that previously was unavailable, and family members can send money without the need of traditional banks. The new technology has created jobs that provide services via satellite and through the Internet, such as accounting, data entry, and telephone help lines.

Many forces are at work behind these changes. One of the most important is the process of **globalization**. *Globalization* is a term used by different people to mean many different things. Columbia University economist Jagdish Bhagwati defines *economic globalization* as the integration of national economies into the international economy through trade in goods and services, direct foreign investment, short-term capital flows, international movements of people, and flows of technology. Globalization has important noneconomic aspects as well, including integration of cultures, communications, and politics. It is not a new phenomenon: The early voyages of Ferdinand Magellan, Christopher Columbus, Zheng He, Marco Polo, and others opened an early epoch of globalization, and the late nineteenth and early twentieth centuries saw increased global integration until the process abruptly ended with the onset of World War I. But the current era has included more parts of the world and affected far more people than earlier episodes.

These broad global trends and the individual stories of Rachmina, Getachew, Viktor and Yulia raise many issues central to the process of economic development addressed in this book. How do governments promote investment, industrialization, and exports? How do countries educate their citizens and protect their health, enabling them to become productive workers? Who benefits from foreign investment and integration with global trading networks, and who loses? How does the shift from agriculture to manufacturing affect the lives of the majority of people in developing countries who still are rural and poor? How will climate change affect the lives of those who already face extreme poverty? This book explores the economics of these and other issues in an attempt to understand why some countries develop rapidly, whereas others seem not to develop at all. Remember that within each nation are people like Rachmina, Getachew, Viktor and Yulia, whose lives are deeply affected by the progress their nations make along the path toward economic development.

RICH AND POOR COUNTRIES

The countries with which this book is concerned have been labeled with many different terms. A term in vogue during the 1980s, especially in international forums, was the **third world**. Perhaps the best way to define it is by elimination. Take away the

industrialized economies of western Europe, North America, and the Pacific (the first world, although it was rarely called that) and the industrialized, formerly centrally planned economies of eastern Europe (the second world), and the remaining countries constitute the third world. This terminology is used much less frequently today. The geographic configuration of the third world has led to a parallel distinction of **North** (first and second worlds) versus **South**, which still has some currency.

The more popular classifications used today implicitly put all countries on a continuum based on their degree of development. Therefore, we speak of the distinctions between developed and underdeveloped countries, more and less developed ones, or—to recognize continuing change—**developed countries** and **developing countries**. The degree of optimism implicit in the words *developing countries* and the handy acronym **LDCs** (less-developed countries), make these widely used terms, although they suffer from the problem that *developed* implies the process is fully complete for wealthier countries. The United Nations employs a classification scheme that refers to the poorest nations as the **least-developed countries**. Some Asian, eastern European, and Latin American economies, whose industrial output is growing rapidly, are sometimes referred to as **emerging economies**. Richer countries are frequently called **industrialized countries**, in recognition of the close association between development and industrialization. The highest-income countries are sometimes called postindustrial countries or service-based economies because services (finance, research and development, medical services, etc.), not manufacturing, account for the largest and most rapidly growing share of their economies.

The rich–poor dichotomy, based simply on income levels, has been refined by the World Bank[4] to yield a four-part classification:

- **Low-income economies**, with average incomes less than $1,005 per capita in 2010, converted into dollars at the current exchange rate.
- **Lower-middle-income economies**, with incomes between $1,006 and $3,975.
- **Upper-middle-income economies**, with incomes between $3,976 and $12,275.
- **High-income economies**, with incomes over $12,275.

The World Bank's classification system dates back to the 1970s. The Bank wanted poorer countries to receive better lending terms. To do so required some way to distinguish economic capacity for repaying loans. Gross national product

[4]The World Bank, formally the International Bank for Reconstruction and Development (IBRD), borrows funds on private capital markets in developed countries and lends to developing countries; through its affiliate the International Development Association (IDA), it receives contributions from the governments of developed countries and lends to low-income countries at very low interest rates with long repayment periods. The Bank, as it often is called, is perhaps the world's most important and influential development agency. Its role is explored in more detail in the discussion of foreign aid in Chapter 14.

(GNP) per capita, also referred to as GNI (gross national income) per capita, was adopted as such a measure. The actual cutoffs used to distinguish between low-, middle-, and high-income economies were based on natural gaps among countries. The threshold income levels are updated annually to account for international price inflation.[5]

Table 1–1 divides the world according to the Bank's classification scheme. It may surprise you that the largest number of countries, 70, falls into the high-income category. This is because in addition to the well-known rich nations such as France, Japan, and the United States, there are a large number of small rich nations, including Aruba, Brunei, Isle of Man, Liechtenstein, and Qatar. Despite the number, the high-income economies represent only 16 percent of the world's population. The middle-income economies represent 72 percent. China and India alone account for almost half of the population of the middle-income economies. The 35 low-income economies, the poorest nations in the world, represent 12 percent of humanity. These economies can be found mostly in sub-Saharan Africa. Haiti is the only nation in the Western Hemisphere that is still a low-income country. Other low-income nations are located throughout Asia, with Bangladesh being the most populous of these. As a group, the low-income nations in 2010 averaged just over $1,200 GNI per capita, measured in terms of **purchasing power parity (PPP)**. PPP is a way of accounting for the difference in prices between nations and gives a more accurate comparison of incomes among countries. (PPP is discussed at greater length in Chapter 2.) Average GNI per capita in low-income nations in 2010 was a mere 3.4 percent of the average GNI per capita of high-income nations, $37,183. We will have more to say about global income inequality in Chapter 6.

Table 1–1 provides a snapshot of the world in 2010. It tells us that one out of every two people in the world live in low- or lower-middle-income countries where the average standard of living is well below that in the upper-middle-income and high-income economies. But it does not tell us much about how things have changed over time. In 1983, when this textbook was first published, 16 percent of the world's population lived in high-income countries, the same percentage as today. But 50 percent of the world's population in 1983 lived in low-income countries, compared to only 12 percent today. This is a dramatic change, the result of rapid economic growth in China, India, and many other previously very poor nations.[6]

[5]The inflation rate used today is an average of inflation in the Euro Zone, Japan, the United Kingdom, and the United States.

[6]According to World Bank estimates, India graduated from low-income to lower-middle-income status in 2009. China became a lower-middle-income economy in the late 1990s and graduated into the upper-middle-income group in 2010. Because of systematic undervaluation of its currency, China actually graduated from one income category to the next earlier than reported by the World Bank. When China officially became an upper-middle-income economy in 2010, the average GNI per capita of both the lower and the upper-middle-income groups fell. Do you see why?

TABLE 1–1 Classification of World Economies, 2010

COUNTRY CLASSIFICATION	GNI PER CAPITA* (US$)	COUNTRIES† (NO.)	POPULATION IN MILLIONS (% OF WORLD TOTAL)	AVERAGE GNI PER CAPITA‡ (US$, PPP)	REGIONAL EXAMPLES§
Low-income	≤ $1,005	35	817 (12%)	$1,247	Ethiopia, Bangladesh, Cambodia, Haiti, Tajikistan
Lower-middle-income	$1,006–3,975	57	2,466 (36%)	$3,701	Senegal, Sri Lanka, Philippines, Ecuador, Jordan, Ukraine
Upper-middle-income	$3,976–12,275	54	2,449 (36%)	$9,904	Gabon, Malaysia, Brazil, Iran, Romania
High-income	>$12,275	70	1,123 (16%)	$37,183	Australia, France, Japan, Norway, Saudi Arabia, Taiwan, United States
World**	$9,097	216	6,855 (100%)	$11,058	

*Gross national income (GNI) per capita expressed in terms of current market exchange rates.
†Countries with populations of 30,000 or more people are included.
‡Average GNI per capita by income group in terms of current purchasing power parity (PPP).
§For the low- and middle-income groups, the examples are listed by World Bank geographical regions in the following order: Sub-Saharan Africa, South Asia, East Asia and Pacific, Latin America and the Caribbean, Middle East and North Africa, and Europe and Central Asia.
**World GNI per capita values are based on the population weighted average of all 216 countries.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

GROWTH AND DEVELOPMENT

While the labels used to distinguish one set of countries from another can vary, one must be more careful with the terms used to describe the development process itself. The terms **economic growth** and **economic development** are sometimes used interchangeably, but they are fundamentally different. *Economic growth* refers to a rise in national or per capita income. If the production of goods and services in a country rises, by whatever means, and along with it average income increases, the country has achieved economic growth. Economic growth explains why the percentage of the world's population living in low-income countries, defined in terms of GNI per capita, has fallen so rapidly over the past three decades. *Economic development*

implies more—particularly, improvements in health, education, and other aspects of human welfare. Countries that increase their income but do not also raise life expectancy, increase schooling, and expand individual opportunities are missing out on some important aspects of development. The extent to which economic growth supports these broader criteria for development is related to **income distribution** within countries. The average income figures cited earlier tell us nothing about how widely (or narrowly) the benefits of growth are shared within countries. If all of the increased income is concentrated in the hands of a few or spent on monuments or a military apparatus, there has been very little development in the sense that we mean.

Development is also usually accompanied by significant shifts in the structure of the economy, as more and more people typically shift away from rural agricultural production to urban-based and higher-paying employment, usually in manufacturing or services. Economic growth without structural change is often an indicator of the new income being concentrated in the hands of a few people. Situations of growth without development are the exceptions rather than the rule, but they do happen. Take the case of Equatorial Guinea, a small nation of fewer than 700,000 people on the west coast of Africa. The discovery and development of vast oil deposits off the nation's coast raised its GNI per capita income from an estimated US$330 in 1990 to US$12,420 in 2009. During the 2000s, Equatorial Guinea was the fastest-growing economy in the world, averaging growth rates of 25 percent per year, far greater than China, India, or any other successful economy. With growth rates of this magnitude, Equatorial Guinea moved from being a low- to a high-income economy in about a decade.

Does this also mean that Equatorial Guinea became a developed economy? By 2009, Equatorial Guinea had a per capita income comparable to Hungary's, but this is where the similarity between the two nations ends. Life expectancy in Equatorial Guinea stands at 50 years. In Hungary it is 74 years. About 90 percent of school-aged Hungarian children are enrolled in primary school; for Equatorial Guinea it is closer to 50 percent. Despite Equatorial Guinea's sudden high level of per capita income there has been little transformation in the low levels of education and poor health care of most Equatorial Guineans. Nor has there been much change in their economic activity. Rapid economic growth has not brought economic development to most of the population of Equatorial Guinea. But again, this case is the exception rather than the rule. In most cases, increases in per capita incomes and economic development have moved together.

Modern economic growth, the term used by Nobel laureate Simon Kuznets, refers to the current economic epoch as contrasted with, say, the epoch of merchant capitalism or the epoch of feudalism. The epoch of modern economic growth still is evolving, so all its features are not yet clear, but the key element has been the application of science to problems of economic production, which in turn has led to industrialization, urbanization, and even explosive growth in population. Finally, it should always be kept in mind that, although economic development and modern economic growth involve much more than a rise in per capita income or product, no sustained development can occur without economic growth.

DIVERSITY IN DEVELOPMENT ACHIEVEMENTS

A large number of less-developed countries have experienced growth in income over the past four decades and many have enjoyed substantial growth. The most rapidly growing economies have been in Asia and include China, India, Indonesia, Korea, Malaysia, and Thailand. But several non-Asian countries also are among the fast growers, such as Botswana, Chile, Estonia, and Mauritius. Since 1970, Botswana, a landlocked country in southern Africa, has been one of the fastest-growing economies in the world and one that has used its increased income to improve the lives of its citizens. Botswana's experience challenges the stereotype that all African countries have been stuck with little growth and development. At the same time, several Asian countries have grown slowly or not at all, including Myanmar (Burma), North Korea, and Papua New Guinea.

There are many examples of countries that have had an income growth exceeding 2 percent a year over the past four decades. At 2 percent annual growth, average income doubles in 35 years; at 4 percent, it doubles in 18 years. In most of these countries, manufacturing grew more rapidly than the gross domestic product and thus moved these economies through the inevitable structural change that reduces the share of income produced and labor employed in agriculture. Many other countries experienced slower (albeit positive) growth and development, with incomes growing 1 or 2 percent per year. In still others, incomes stagnated or declined. Most of the countries in this latter group are in Africa, although income also fell elsewhere, including in many of the transition economies of eastern Europe and Central Asia.

Perhaps the most remarkable changes in low-income countries in recent decades have been the virtually universal improvement in health conditions and the availability of schooling. From 1970 to 2009, the infant mortality rate in today's low-income nations fell dramatically from 147 to 76 per thousand births. This means that within this group of nations an additional 71 children out of every 1,000 lived to see their first birthday. For today's middle-income nations, which in 1970 included many countries that were still low-income, the results are similarly dramatic. Primary school enrollment became nearly universal in middle-income economies and rose substantially in most of the low-income countries. With few exceptions, more than three-quarters of eligible children attend primary school in poor countries. Despite this good news, more than 1 billion people in the developing world continue to live in extreme poverty.

The study of economic development is not mainly a review of what has and has not been accomplished in the past. It is a field concerned most of all with the future, particularly the future of the least-advantaged people in the world. To comprehend the future, one must first try to understand how we got to the point where we are now. But the future will not be just a replay or a projection of trends of the past, as new forces that will shape that future also are at work. Some of these forces can be seen clearly today, whereas others are only dimly perceived, if they are seen at all.

Any list of changes that will make the future of economic development different from the past should probably start with the information revolution. Greatly enhanced communication around the world as represented by the Internet has sped up the flow of ideas across oceans and borders to an unprecedented degree. Lower transport costs, together with better information, contribute to global production networks and the expansion of global trade and investment. The rapid flow of information is also having an impact on politics by making it harder for authoritarian regimes to control what their people are allowed to know. Partly for this reason, democratic regimes are becoming more the norm than the exception in developing countries, and there is reason to expect this trend to continue.

Not all the foreseeable trends of the future are positive. Despite the benefits of advanced technology and the information revolution, some groups in society, notably the better educated, may capture most of the gains while other large groups are left behind. The experience with HIV/AIDS must keep us vigilant for whatever the next new infectious disease might be. Natural disaster—droughts, earthquakes, hurricanes, and tsunamis—hit rich and poor nations alike, but the effects are usually far more severe the poorer the country. Environmental degradation is much more serious today than it was a century ago, when Europe and North America were in the early stages of economic growth. Global warming, as a result, is a problem likely to play an important role in our future. The one positive change in the environmental sphere is that people around the world are becoming aware of the danger at a much faster pace than in the past, though international cooperation in limiting climate change remains a challenge.

We probably are not even aware of many of the forces that will shape the future economic development of nations. No one at the end of the nineteenth century had heard of nuclear energy, DNA, or integrated circuits. No one in the 1970s had heard of cell phones, laptop computers, or antiretroviral (ARV) drugs used to treat HIV/AIDS. Given the pace of change in the current world of the new millennium, similar and possibly greater discoveries will profoundly influence how economies develop. That said, we cannot rely on future discoveries to solve the problems of economic development and poverty among nations. We must try to understand how the nations of the world got to where they are today so that we can do a better job of raising living standards for all in the future.

APPROACHES TO DEVELOPMENT

This book is not for readers looking for a simple explanation of why some countries are still poor or how poverty can be overcome. Library shelves are full of studies explaining how development will occur if only a country will increase the amount it saves and invests or intensify its efforts to export, among other prescriptions.

For two decades in the mid-twentieth century, industrialization through import substitution—the replacement of imports with home-produced goods—was considered by many to be the shortest path to development. In the 1970s, labor-intensive techniques, income redistribution, and provision of basic human needs to the poor gained popularity as keys to development. More recently, economists have counseled governments to avoid high protective barriers and to depend substantially on markets to set prices and allocate resources. A different theme for some analysts is that development will be possible only with a massive shift of resources, in the form of foreign aid and investment, from the richest countries to the poorest. Others call for debt forgiveness for poor nations that have found it difficult to repay earlier loans.

No single factor is responsible for poverty, and no single policy or strategy can set in motion the complex process of economic development. Each of the various explanations and solutions to the development problem makes sense if placed in the proper context and makes no sense at all outside that set of circumstances. Import substitution has carried some countries toward economic development, but export promotion has helped others when import substitution bogged down. Prices badly distorted from their free-market values can stifle initiative and hence growth, but removing those distortions leads to development only when other conditions are met as well. Finally, where leaders backed by interests hostile to development rule countries, those leaders and their constituents must be removed from power before growth can occur. Fortunately, the majority of developing countries have governments that want to promote development.

This book is not neutral toward all issues of development. Where controversy exists we shall point it out. Indeed, the authors of this book differ among themselves over some questions of development policy. But we share a common point of view on certain basic points.

This text extensively uses the theoretical tools of mainstream economics in the belief that these tools contribute substantially to our understanding of development problems and their solution. The text does not rely solely or even primarily on theory, however. For five decades and more, development economists and economic historians have been building up an empirical record against which these theories can be tested, and this book draws heavily on many of these empirical studies. We try to give real-country examples for the major points made in this book. In part, these examples come from the individual country and cross-country comparative studies of others, but they also are drawn extensively from our own personal experiences working on development issues around the world. The several authors who contributed to this textbook, both the current and past editions, have been fortunate enough to study and work over long periods of time in Bolivia, Chile, China, the Gambia, Ghana, Indonesia, Kenya, Korea, Malaysia, Nepal, Peru, Samoa, Sri Lanka, Tanzania, Vietnam, and Zambia. At one time or another, at least one nation from this group has exemplified virtually all the approaches to development now extant.

THE STUDY OF DEVELOPMENT ECONOMICS

If you are like most students taking a course in development economics this will not be your first course in economics. Most likely, you have taken courses in principles of microeconomics and macroeconomics. Some of you may have also studied intermediate economic theory, statistics, econometrics, and other economic subfields. Your study of these subjects will prove extremely useful in your examination of development economics. In your introductory microeconomics class you learned the importance of incentives and how markets tend to clear when quantity demanded equals quantity supplied. In macroeconomics you learned how expanding the money supply can lead to inflation. These insights are as applicable to poor nations as they are to rich ones. But there also are important differences between your earlier study of economics and the study of development. Context matters.

If you took your micro principles courses in a college or university in the United States or other high-income country, the examples you were given and the problems you studied reflected those of a rich nation. Rent control is a common example included in principles textbooks to explain price ceilings and how they can have unintended effects, including housing shortages and black market prices. In a developing country, price ceilings will have the same impact but are unlikely to be used to control apartment prices. A better example might be how governments have attempted to use price ceilings to *lower* urban food prices for urban consumers (usually at the expense of farmers, who are often much poorer than urban consumers). Rich nations are more prone to employ price floors to *raise* food prices and support farm incomes.

When discussing taxes, authors of principles textbooks in the United States will focus on how marginal tax rates can affect the supply of labor—the higher the tax, the less likely workers will want to work. In poor nations, the presence of such taxes is most likely to encourage growth in the informal sector in which people are employed, although those workers evade paying taxes on their earnings. This happens in rich nations too but is a much less common response. In low-income nations, the lack of secure property rights to land helps explain squatter settlements and urban slums. This is not much of a problem in rich nations and is less likely to be discussed.

In learning macroeconomics, you may have studied how an expansion in the money supply may cause a nation's exchange rate to depreciate. You were less likely to have learned how a country can fix its exchange rate intentionally to undervalue its currency. Such practices are far more common in developing nations (including China) than in developed ones. You might have read about the importance of the independence of the Federal Reserve System in the United States, which helps insulate decisions about monetary policy from domestic politics. But in developing nations, the lack of independence of central banks is far more likely.

The study of macroeconomics in high-income economies tends to focus on economic stabilization—that is, on how monetary and fiscal policy can be used to keep

unemployment down and inflation low. Economic growth is less of a focus, in part because of the success high-income economies have had in growing their economies. To the extent you studied economic growth, much of the focus was on technological change as a determinant of growth. In the development setting, economic growth and structural change are central issues in the field. Growth rates depend not only on the technological frontier but also on the ability to mobilize savings and engage in productive investment. These issues will capture much of our attention in the chapters that follow.

A final difference between the study of economics in a developed versus a developing nation context is the role of **institutions**. Economic theory tends to take institutions (the rules of the game that govern the functioning of markets, banking systems, enforcement of property rights, and so on) as a given. But development is concerned with how one creates and strengthens institutions that facilitate development in the first place. How, for example, does a country acquire a government interested in and capable of promoting economic growth? Can efficiently functioning markets be created in countries that currently lack them, or should the state take over the functions normally left to the market elsewhere? Is a fully developed financial system a precondition for growth, or can a country do without at least some parts of such a system? Is land reform necessary for development and, if so, what kind of land reform? What legal systems are needed to support market-based growth? These institutional issues and many others like them are at the heart of the development process and will reappear in different guises throughout this book.

The economics you have studied before is an important foundation for the study of development economics. Be prepared to build on it.

ORGANIZATION

This book is divided into four parts. Part 1 examines the main factors, both those suggested by economic theory and those supported by empirical investigations, that contribute to differing rates of economic growth. This discussion involves the deliberate choices by governments, including the debate over how economic development should be guided or managed.

Part 2 goes beyond issues of economic growth and focuses directly on inequality and poverty. Because economic development first and foremost is a process involving people, who are both the prime movers of development and its beneficiaries, Part 2 deals with how human resources are transformed in the process of economic development and how that transformation contributes to the development process itself. Individual chapters are devoted to population, education, and health.

The other major physical input in the growth process is capital. Part 3 is concerned with how capital is mobilized and allocated for development purposes. From

where, for example, do savings come and how are they transformed into investment? How does government mobilize the resources to finance development? What kind of financial system is consistent with rapid capital accumulation? Will inflation enhance or hinder the process, and what roles will foreign aid and investment play?

Especially in the early stages of development, countries depend heavily on agriculture and on the export of food, fuel, and raw materials. Part 4 discusses strategies to enhance the productivity of such primary industries as a first, and often a continuing, task in stimulating economic development. Part 4 also explores trade in primary products, in manufactured goods and increasingly in services, too. In a more globalized world economy, trade plays a larger role in low- and middle-income nations than ever before. Part 4 concludes with the all-important question of environmental sustainability and the challenges developing nations confront in the face of climate change.

SUMMARY

- The last 40 years have seen a wide diversity of development experiences around the world. Some countries, including some very large ones like China, India, and Indonesia, have experienced rapid growth and development. Others, particularly many African countries and some in eastern Europe, have experienced stagnation or even a decline in incomes. Understanding the differences in these experiences and the lessons for the future is the core purpose of this book.
- Many different terms are used to differentiate poor from rich countries, but this text mainly uses the terms *developing* and *low-* and *middle-income economies* to refer to those nations with incomes substantially lower than the *developed* and *high-income* nations.
- Only 12 percent of the world's population today lives in low-income economies, nations with a GNI per capita falling below US$1,005 (in 2010). Twenty-five years ago, *half* of the world's population lived in low-income nations. Economic growth in China, India, and many other previously poor nations accounts for this historic change. Of course, many very poor people still live in these economies, but their numbers have fallen significantly.
- *Economic growth* refers to an increase in per capita incomes, whereas *economic development* involves, in addition, improvements in health and education and major structural changes, such as industrialization and urbanization. Some countries may have economic growth, usually because of the discovery of great mineral wealth, but not development because they retain many of the structural features of a traditional society.
- No single factor is responsible for poverty, and no single policy or strategy can set in motion the complex process of economic development. We can

learn much from the past experience of other nations, especially those that have achieved rapid growth and experienced economic development in recent decades.

- We must also be aware that new forces, from new diseases to new technologies, will influence the path and opportunities facing today's developing nations. Changes in the global climate, including the planet's physical climate as well as its economic and political climate, will also impact the course nations follow.
- The economics of development bears a lot in common with the economics you may have studied in other courses. But it is also different. Context matters. A focus on long-term economic growth and structural change in the economy and on the role of institutions commands the attention of development economists.



Measuring Economic Growth and Development

A native American saying recommends, "One should not go hunting a bear unless one knows what a bear looks like." This is sound advice for bear hunters; it also has meaning for our inquiry. Understanding how to achieve economic development requires some agreement on what we want to achieve.

The previous chapter drew a distinction between economic growth and economic development. Economic growth refers to a rise in real national income per capita—that is, a rise in the inflation-adjusted, per person, value of goods and services produced by an economy. This is a relatively objective measure of economic capacity. It is widely recognized and can be computed with varying degrees of accuracy for most economies. There is far less of a consensus on how to define economic development. Most people would include in their definition increases in the material well-being of individuals as well as improvements in basic health and education. Others might add changes in the structure of production (away from agriculture toward manufacturing and services), improvement in the environment, greater economic equality, or an increase in political freedom. Economic development is a normative concept, one not readily captured by any single measure or index.

To understand the magnitude of the global challenge of development, it is essential to be able to track what has happened to an economy over time and make comparisons between countries. If we want to understand why some nations experienced more rapid growth and development than others, we need measures of economic performance that are relatively accurate and comparable. Poor countries are environments in which information is scarce and data can be of questionable quality, so we have to assure ourselves that our indicators, though imperfect, are sufficiently

robust to help us understand the outcomes we observe. The study of economic development requires us to combine our insights on how economies work with an appeal to the evidence to check if our insights are consistent with experience. Measurement is central to this process and will be an issue we return to throughout this book.

To get started, this chapter introduces measures of national income and considers the problem of making cross-country comparisons when national incomes are expressed in different currencies. Equipped with a means of making comparisons of national income levels, we examine the record both over time and across countries. These data highlight the enormous differences in economic growth that have characterized different regions of the world over the past 500 years as well as over the more recent past. Much of the rest of this book is devoted to understanding what has caused these differences.

Economic growth may be central to achieving economic development, but there is much more to economic development than growth alone. Not only the level of per capita income but how that income is produced, spent, and distributed within and between countries determines development outcomes. There is much debate about how to define and measure economic development. We introduce two widely cited indicators of economic development, the human development index and the millennium development goals, and consider their strengths and weaknesses. The information presented in this chapter may not make you a better bear hunter but it will inform the rest of your study of development economics.

MEASURING ECONOMIC GROWTH

At the core of studies of economic growth are changes in national income. Two basic measures of national income are commonly employed. **Gross national product (GNP)** is the sum of the value of finished goods and services produced by a society during a given year. GNP excludes intermediate goods (goods used up in the production of other goods, such as the steel used in an automobile or the chips that go into a computer). GNP counts output produced by citizens of the country, including the value of goods and services produced by citizens who live outside its borders. GNP is one of the most common terms used in national income accounting. The World Bank and other multilateral institutions often refer to this same concept as gross national income (GNI). **Gross domestic product (GDP)** is similar to GNP, except that it counts all output produced within the borders of a country, including output produced by resident foreigners, but excludes the value of production by citizens living abroad. GNP or GDP divided by total population provides a measure of **per capita income**. Economic growth refers to changes in per capita income over time.

The distinction between GNP and GDP can be illustrated using examples from two very different economies, Angola and Bangladesh. More than three-quarters of

Angola's national income is derived from oil. Multinational companies drill for most of the oil and repatriate their profits. These profits count as part of Angola's GDP but not its GNP. In 2009, Angola's GDP was 12 percent higher than its GNP. By contrast, Bangladesh has few natural resources and little foreign investment. Large numbers of Bangladeshis work abroad, especially in the Persian Gulf: men often as construction workers and Bangladeshi women as domestics. The value of the output produced by these Bangladeshi workers counts as part of Bangladesh's GNP (since these workers are Bangladeshi nationals) but not as part of its GDP (because the work is performed outside of the country). In 2009, Bangladesh's GNP was 9 percent higher than its GDP. In most countries the differences between GNP and GDP are much smaller. In part because it is easier to track economic activity within a nation's borders, GDP has become the more widely used measure of national income by the International Monetary Fund (IMF), UN Development Programme, World Bank, and other multilateral agencies as well as by researchers engaged in analyzing cross-country data and trends. We follow this convention and refer primarily to GDP and GDP per capita as measures of national income from here on. Unless otherwise indicated, when discussing trends over time, we refer to **real GDP** and real GDP per capita—that is, per capita gross domestic product adjusted for domestic price inflation.[1]

The contribution of a sector or component of GDP, such as manufacturing or agriculture, is measured by the value added by that sector. **Value added** refers to the incremental gain to the price of a product at a particular stage of production. Therefore, the value added of the cotton textile industry is the value of the textiles when they leave the factory minus the value of raw cotton and other materials used in their production. At the same time, the value added is equal to the payments made to the factors of production in the textile industry: wages paid to labor plus profits, interest, depreciation of capital, and rent for buildings and land. Because the total value added at all stages of production equals total output, GDP is a measure of both total *income* and total *output.*

MEASURING GDP: WHAT IS LEFT OUT?

One way to calculate GDP is to add up the value of all the goods and services produced within a country and then sold on the market. The focus on goods and services sold in the market creates a measurement problem because many valuable contributions to society are excluded. When a farm household pays someone else to dig an irrigation ditch or repair a roof, such economic activity is included in GDP because these activities are purchased "in the market." However, when unpaid members of

[1]Real GDP is computed by deflating nominal GDP (GDP measured in current prices) by a price index. National statistical offices often calculate a variety of price indices, including the consumer price index (CPI), the GDP deflator, and others. What these indices share in common is an attempt to isolate any general increase (or decrease) in the price level for all goods.

the household perform these same tasks, they tend not to enter GDP. The scale of this problem tends to be larger in low-income countries and is evident in a poor nation like Cambodia, where about one-third of the labor force is classified as unpaid family workers, most of whom are engaged on family farms, producing food and other goods and services for their own consumption.

In most developing countries, a large number of activities do not enter the market. Much of what is produced by the agricultural sector is consumed by the farm household and never exchanged in the marketplace. To not include this production would seriously underestimate a nation's GDP. The usual practice is to include estimates from sample surveys of farm output consumed by the producer, which are then valued at the prices of marketed farm produce. This is done, for example, in Moldova and even includes the output of household garden plots. In India, estimates are made for the construction of traditional homes made out of mud, straw, and other local materials. Even illegal activity may be included, as in Afghanistan where estimates of poppy production, a banned crop, are part of the nation's GDP. Despite these adjustments, not all household production is accounted for. As economies grow, more output is transacted in the marketplace and gets included in GDP. The resulting estimates of GDP may overestimate the growth in economic activity because some of what is now captured is merely a transfer of production from within the household to the market.

An additional measurement problem for GDP arises from the need to compare apples with oranges in calculating the value of national output. A typical economy might produce thousands of different goods and services. Adding up the total value of goods and services that are traded in markets requires using their market prices. But accurate price information may not be available or may not be representative of market prices at the national level. Government agencies in poor countries may lack the means to conduct thorough market surveys of prices or may rely too heavily on information from major urban centers (where prices may be easier to track but are unrepresentative of markets around the country).

Another criticism of GDP is that it may be a measure of the goods and services produced by an economy, but does not account for the "bads" society produces. If a steel mill pollutes a river or the air, the value of the steel produced is included in GDP but the cost of pollution is not deducted. Should crime, congestion, and other social bads be deducted from estimates of GDP? *Gross* domestic product also does not account for the depreciation of goods (for example, when machinery or trucks wear out) or depletion of natural resources (when forests are cut, fisheries depleted, or mines exhausted). Proposals for making adjustments to GDP to account for these factors have been raised, but none has been widely adopted yet.[2] Although there are

[2]In 2008, President Nicolas Sarkozy of France established the international Commission on the Measurement of Economic Performance and Social Progress that included discussion of GDP as a useful measure. The commission, chaired by Joseph Stiglitz, a Nobel laureate in economics, raised many concerns about GDP as a measure of economic production and as a measure of the quality of life and of sustainability. More information is available at www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

obvious flaws in GDP as a measure of national income, there are also many benefits. Having a widely agreed on approach to measuring national income facilitates comparisons of nations' economic activity both over time and relative to other countries. Both types of comparisons are essential to understanding the process of economic development.

EXCHANGE-RATE CONVERSION PROBLEMS

Another measurement issue we need to consider is how to compare *levels* of GDP per capita across countries. The problem arises because each nation measures national income in its own currency: dinar in Tunisia, guarani in Paraguay, leu in Moldova, and so on. Economic growth rates can be computed in a nation's own currency, but if we want to understand better what is required to transform a nation from low to high income, it is useful to compare nations at different income levels. To do so requires converting GDP per capita into a common currency. The shortcut to accomplishing this goal is to use the market exchange rate between one currency, usually U.S. dollars, and each national currency. For example, to convert India's GDP per capita (2009) of about 57,000 rupees into U.S. dollars, use the appropriate exchange rate between the two currencies (about 49 rupees per US$1 in 2009), which in this case yields an estimate of about US$1,160.

A common reaction to this low figure by anyone who has lived in or visited India (or for that matter any developing nation) is that one U.S. dollar goes much further in India than it does in the United States. A basic woman's haircut in a less-affluent part of Mumbai, for example, might cost 200 rupees (US$4 at the official exchange rate), whereas a basic haircut in Boston might run US$40. If one can buy more for $1 in India than one can in the United States—in this example, 10 haircuts in Mumbai for the price of 1 in Boston—then India's true level of per capita income must be higher than the one given by converting currency using the official exchange rate.

There is considerable merit to this argument. One problem with converting per capita income levels from one currency to another is that exchange rates, particularly those of developing countries, can be distorted. Trade restrictions or direct government intervention in setting the exchange rate make it possible for an official exchange rate to be substantially different from a rate determined by a competitive market for foreign exchange.

But even the widespread existence of competitively determined market exchange rates would not eliminate the problem. The huge price difference in haircuts between Boston and Mumbai is not the result of trade restrictions or a managed Indian exchange rate. Instead, a significant part of national income is made up of what are called **nontraded goods and services**—that is, goods that do not and often cannot enter into international trade. Haircuts are one example. Internal transportation, whether by bus, taxi, or train, cannot be traded, although many transport inputs, such automobiles and rail cars, can be imported. Wholesale

and retail trade and elementary school education also are nontraded services. Land, homes, and office buildings are other obvious examples of goods that are not exchanged across national borders. Generally speaking, whereas the prices of traded goods tend to be similar across countries (because, in the absence of tariffs and other trade barriers, international trade could exploit any price differences), the prices of nontraded goods can differ widely from one country to the next. This is because the markets for nontraded goods are spatially separated and the underlying supply and demand curves can intersect in different places, yielding different prices.

Exchange rates are determined largely by the flow of traded goods and international capital and generally do not reflect the relative prices of nontraded goods. As a result, GDP converted to U.S. dollars by market exchange rates gives misleading comparisons of income levels if the ratio of prices of nontraded goods to prices of traded goods is different in the countries being compared. The way around this problem is to pick a set of prices for all goods and service prevailing in one country and to use that set of prices to value the goods and services of all countries being compared. In effect, one is calculating a purchasing power parity (PPP) exchange rate. Thus a cement block, a computer chip, or a haircut is assigned the same value whether it is produced in New Delhi or New York.

The essence of the procedure can be illustrated by the numerical exercise presented in Table 2–1. The two economies in the table are called the United States and India for illustrative purposes, and each economy produces one traded commodity (steel) and one nontraded service (retail sales). Each economy produces a different amount of each good. GDP, expressed in local currencies, is equal to the total value of production of steel plus retail sales. A ton of rolled steel sells for about $1,000 in the United States and Rs 50,000 in India. The value of the services of retail sales personnel is estimated in the most commonly used way, which is to assume the value of the service is equal to the wages of the worker providing the service. (For the United States we assume earnings of $10 per hour, working 40 hours per week for 50 weeks, for annual earnings of $20,000. In India, we assume annual earnings of Rs 60,000.) Wages are likely to differ widely across countries and to be determined almost exclusively by domestic labor supply and labor demand conditions. This is because workers cannot easily migrate from one country to another to take advantage of any differences in wages (partly because of immigration rules and partly because the cost of moving to a new country can be high, both financially and psychically). From the data in Table 2–1 we determine that GDP in the United States equals $240 billion and in India, Rs 1,490 billion.

One way of comparing the GDP levels in the two economies is to convert them into a single currency, say, the U.S. dollar. In this simple world of two goods and two nations, the exchange rate is determined solely by trade in steel. If steel is freely traded between the two countries, then the exchange rate settles where the price per ton of steel in the two countries is equal—that is, at the point at which the U.S. price of $1,000 per ton equals India's price of Rs 50,000 per ton or

TABLE 2–1 Market Exchange Rate Versus Purchasing Power Parity Methods of Converting GDP

	UNITED STATES			INDIA		
	QUANTITY	PRICE (US$)	VALUE OF OUTPUT (BILLION US$)	QUANTITY	PRICE (RUPEES)	VALUE OF OUTPUT (BILLION RUPEES)
Steel (million tons)	200	1,000 per ton	200	25	50,000 per ton	1,250
Retail sales personnel (millions)	2	20,000 per person per year	40	4	60,000 per person per year	240
Total GDP (local currency, billions)			240			1,490

Market exchange rate based on steel prices = Rs 50,000/$1,000 or Rs 50 = US$1.

1. India's gross domestic product (GDP) in U.S. dollars calculated by using the official exchange rate: Rs 1,490 billion/Rs 50 = US$29.8 billion.

2. India's GDP in U.S. dollars calculated by using U.S. prices for each individual product or service and applying that price to India's quantities (that is, using purchasing power parity [PPP]):
Steel: 25 million tons × $1,000/ton = $25 billion
Retail sales personnel: 4 million people × $20,000/person = $80 billion
GDP: $25 billion + $80 billion = $105 billion

3. Ratio of PPP calculation of India's GDP to official exchange rate calculation: $105 billion/$29.8 billion = 3.5

where US$1 = Rs 50.[3] Using this market-determined exchange rate, India's GDP of Rs 1,490 billion equals US$29.8 billion, or about 12 percent of U.S. GDP in this hypothetical example.

The problem with this comparison is that, although Rs 50 and US$1 purchase the same amount of steel in both countries, they purchase different amounts of the nontraded good. To compare the GDP levels of the two nations taking into account this difference in the purchasing power of the respective currencies, we cannot rely on market exchange rates. An alternative approach is to use a common set of prices applied to the output of both countries. We can calculate Indian GDP in U.S. dollars by applying U.S. prices for each product or service to India's quantities. (We could also compute U.S. GDP in terms of India's prices but the convention is to express PPP estimates in terms of U.S. dollars.) This PPP calculation results in India's steel production valued at US$25 billion and retail sales valued at US$80 billion, for an estimated India GDP of US$105 billion. In this example, the PPP calculation of

[3]At any other exchange rate, there would be profitable opportunities to buy more steel from one of the two countries, causing changes in the market for foreign exchange until the two steel prices were equivalent and the exchange rate settled at US$1 = Rs 50. This is sometimes referred to as *the law of one price,* reflecting how opportunities for arbitrage in traded goods lead to price convergence in these goods.

India's GDP is more than three times as large as the calculation that relied on market exchange rates. In terms of PPP, India's GDP is over 40 percent of the U.S. GDP.

Table 2–1 presents a hypothetical PPP conversion for two countries using two goods. The task becomes significantly more complicated in a world of tens of thousands of goods and more than 200 nations. The **International Comparison Program (ICP)**, which began in 1968 under the auspices of the United Nations and is now overseen by the World Bank, tackles this difficult task by deriving a set of **international prices** in a common currency. Detailed price data on a basket of hundreds of specific goods have been collected periodically for an ever-increasing number of nations. International prices are then derived by aggregating the price data from the individual countries and are used to determine the value of national output at these standardized international prices. The most recent round of international price comparisons released by the ICP was based on 2005 data and represented a significant quality improvement over the previous round of price data from 1993. Key elements of this improvement were coverage of a larger number of countries (146 countries in 2005 as compared to only 118 in 1993) and more careful comparison of specific goods and services across countries. (The next update of the ICP will be based on 2011 data, to be released in 2013.) Estimates of national income in terms of PPP are reported in the publications of the IMF, UN Development Programme (UNDP), World Bank, and other multilateral agencies. Researchers have made extensive use of these data.[4]

The ratio of GDP per capita based on international prices relative to GDP using official exchange rates ranged in 2009 from about 0.7 in Norway to 3.3 in the Gambia (Table 2–2). For high-income countries, like Germany, Japan, and the United Kingdom, the ratio is close to 1.0. This means market exchange rate conversion is a close approximation of what is obtained when converting German, Japanese, or UK GDP into international price dollars using the PPP method. This is to be expected because at similar levels of income the prices of nontraded goods tend to be similar as well. For low- and middle-income economies the ratio is greater than 1, consistent with the finding that the degree to which the official exchange rate conversion method understates GNP is related, generally, to the average income of the country. For China the ratio is 1.8, for Bolivia 2.5, for Vietnam and Ethiopia 2.7, and for India 2.8.

With differences of this magnitude, comparisons of per capita income levels using market exchange rate conversions can be misleading. Market exchange rates suggest that per capita incomes in the United States were about 40 times those in India in 2009. PPP calculations narrow the multiple to about 14 times—still a huge gap but maybe a more reasonable indicator of relative income levels. Another way of appreciating the difference between making comparisons of GDP using market exchange rates versus PPP is to think about world GDP as a whole. When world GDP

[4]If you read *The Economist* you may be familiar with another measure of PPP, the Big Mac index, which was introduced, lightheartedly, in 1986 and has been reported on annually ever since. The common basket of goods is a cheese hamburger with lettuce, onions, and pickles on a sesame bun. For a discussion of the Big Mac index as a measure of PPP, see M. Pakko and P. Pollard, "Burger Survey Provides Taste for International Economics," Federal Reserve Bank of St. Louis, *The Regional Economist* (January 2004), 12–13.

TABLE 2–2 Comparing GDP per Capita Using Market Exchange Rates and PPP in 2009 (US$)

COUNTRY	GDP AT MARKET EXCHANGE RATES	GDP AT PPP	RATIO OF PPP CALCULATION TO MARKET EXCHANGE RATE CALCULATION
Norway	79,089	56,214	0.7
Japan	39,738	32,417	0.8
Germany	40,670	36,378	0.9
United Kingdom	35,165	35,155	1.0
United States*	45,989	45,989	1.0
Hungary	12,868	20,312	1.6
Lebanon	8,175	13,070	1.6
China	3,744	6,828	1.8
Botswana	6,064	13,384	2.2
Bolivia	1,751	4,419	2.5
Vietnam	1,113	2,953	2.7
Ethiopia	344	934	2.7
India	1,192	3,296	2.8
The Gambia	430	1,415	3.3

*Gross domestic product (GDP) per capita in the United States is unchanged when measured in terms of purchasing power parity (PPP). This must be the case because the United States is used as the reference country by the International Comparison Program (ICP). As with any index number, the price index at the heart of the ICP must be compared relative to some base, and by convention, U.S. prices were selected.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

is calculated by converting each nation's GDP into a common currency using market exchange rates, the low- and middle-income economies account for 29 percent of world output. When the calculation is based on PPP, the low- and middle-income economies account for 44 percent of world output.

PPP allows for more valid comparisons of real income levels across economies. But PPP has its limits too. Trade and capital flows are transacted at market exchange rates and should be converted at those rates. The ICP provides a consistent set of PPP estimates of national income, but these are only estimates and critics have pointed out flaws in data collection and methodology.[5] PPP conversions cannot correct for

[5]Angus Deaton provides a useful introduction to the ICP in "Reshaping the World: The 2005 Round of the International Comparison Program," in Prasada Rao and Fred Vogel, eds., *Measuring the Size of the World Economy: The Framework, Methodology, and Results from the International Comparison Program* (Washington, DC, World Bank, in press). Problems associated with constructing PPP estimates are discussed in Angus Deaton and Alan Heston, "Understanding PPPs and PPP-Based National Accounts," *American Economic Journal: Macroeconomics* 2, no. 4 (2010), 36–45.

underlying problems in the measurement of GDP in a nation's own currency. Variations in the quality of goods cloud cross-country comparisons. In addition, the specific price index constructed by the International Comparison Project gives more weight to the goods consumed in rich nations and tends to bias upward the GDP of poorer nations that consume a different basket of goods. This index number problem occurs whenever one studies the aggregate performance of an economy over time or compares the performance of different economies. But, despite these problems, much can be learned from the data at hand, and PPP estimates of GDP per capita are central to the study of economic growth and development.

ECONOMIC GROWTH AROUND THE WORLD: A BRIEF OVERVIEW

We now turn from exploring the measurement of GDP to examining the actual performance of countries around the world in terms of the rate of growth of GDP per capita.[6] We begin by looking at the findings of economic historian Angus Maddison, who estimated income levels and corresponding rates of economic growth for the world economy as far back as the year 1 B.C.E. Such an exercise requires a lot of conjecture, especially the further back in time one goes. To perform the analysis, Maddison compiled estimates of population, GDP, and a price index for determining PPP.[7]

According to Maddison's calculations, average world income in 1000 was virtually the same as it had been 1,000 years earlier. In other words, growth in per capita income between 1 B.C.E. and 1000 was effectively zero. The next 820 years (from 1000 to 1820) were barely any better, with world income per capita growing, on average, by just 0.05 percent per year. (Note: This is not a growth rate of 5 percent; it is a growth rate of 0.05 percent.) During those 820 years, world GDP grew by only slightly more than the growth in world population. After eight centuries, world per capita income had increased by only 50 percent. To place this in some perspective, China today is one of the world's fastest-growing economies. With more than 1 billion people (about four times the entire world's population in 1000), economic growth in China averaged about 9.5 percent over the past decade, raising Chinese per capita incomes by 50 percent, not in 820 years but in just under 5 years!

[6]In this section we derive the growth rate of real GDP per capita in PPP using the formula for annual compound growth, $Y_t = Y_0(1 + r)^t$, where Y refers to real GDP per capita in PPP; t, the number of years under consideration; r, the rate of growth of real GDP per capita; and Y_0 refers to real GDP per capita in the base year and Y_t in the final year. Alternative ways of estimating growth rates are discussed in Box 3–2.

[7]Angus Maddison died in 2010. His work is being maintained by his colleagues. The data reported here are from Maddison's original web page titled "Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD." Links are available at www.ggdc.net/MADDISON/oriindex.htm, accessed February 2012.

Maddison's estimates indicate considerable uniformity in per capita incomes throughout the first millennium. The little bit of economic growth that did take place over the next 800 years was centered in western Europe and in what Maddison calls the western "offshoots" (Australia, Canada, New Zealand, and the United States). By 1820, these regions already had a decided advantage over the rest of the world. For example, whereas China and India may have been slightly ahead of the western European countries in 1000, average per capita incomes in western Europe and in their offshoots were already double those of China and India by 1820. Box 2–1 offers one explanation for some of the early and divergent trends and their consequences for the world we live in today.

Maddison's research suggests that rapid economic growth as we know it really began around 1820. He estimates that over the subsequent 190 years, the average growth in world income increased to 1.3 percent per year. Note that the difference between annual growth of 0.05 percent and 1.3 percent is huge. With the world economy growing at 0.05 percent per year, it would take more than 1,400 years for average income to double. With annual growth of 1.3 percent, average income doubles in just 55 years. The world had changed from no growth at all during the first millennium, to slow growth for most of the second millennium, to a situation in which, in the past two centuries, average real income began to double in less than every three generations.

Maddison's estimates of average income levels for the world's major regions since 1820 are shown in Figure 2–1.[8] Several features of these data are notable. First, economic growth rates clearly accelerated around the world since the early 1800s and especially after 1880. Second, and perhaps most striking, the richest countries recorded the fastest growth rates and the poorest countries recorded the slowest growth rates, at least until 1950. Per capita income in the Western offshoots grew by about 1.6 percent per year between 1820 and 1950, while in Asia it grew by only 0.16 percent. As a result, the ratio of the average incomes in the richest regions to those in the poorest regions grew from about 2:1 in 1820 to about 13:1 in 1950.

Between 1950 and 2008, the patterns of economic growth changed, at least in several regions. The gap between the Western offshoots and western Europe, which had been widening through 1950, narrowed significantly. The poorest region in 1950 (Asia) recorded the fastest subsequent growth rate (3.6 percent), thereby beginning to close the income gap with the richer regions of the world. By contrast, Latin America's growth stagnated during the 1980s and 1990s, and eastern Europe's collapsed after the fall of the Berlin Wall in 1989. Both regions resumed economic growth during the 2000s.

[8]Note that the *y*-axis in Figure 2–1 expresses GDP per capita in PPP using Geary-Khamis (GK) dollars, another PPP index. Figure 2–1 also expresses per capita income in logarithms. We will have more to say about the use of log scales later in this chapter. For now, it will be useful to know that when using a log scale, the slopes of the lines for each region are estimates of the growth rate of GDP per capita.

BOX 2-1 JARED DIAMOND: GUNS, GERMS, AND STEEL

Most of this textbook is devoted to understanding why over the past 50 years some countries have experienced rapid economic growth and development while others have not. Jared Diamond, a physiologist, geographer, and Pulitzer Prize–winning author, poses a related but different question. The world as we know it is the result of the historical dominance of Eurasians, especially people of Europe and East Asia. Why, Diamond asks, did history turn out this way? Why did things not work in reverse with Native Americans, Africans, and Aboriginal Australians conquering Europeans? We know that Hernán Cortés, the conquistador, overthrew the Aztec Empire and began Spanish colonization of the Americas. But Diamond wants to understand why the opposite did not happen. Why didn't Emperor Montezuma cross the Atlantic and conquer Europe? World history would have turned out differently had he done so.

For Diamond, much of the history of the last two centuries is the result of what happened in the previous 10,000 years. It was the advantages Europe and Asia had over other continents by 1500 that determined much of what followed. Diamond is interested in the early divergence of regional incomes. He identifies the "proximate causes" of Eurasian dominance over other regions: guns, germs, and steel. Eurasians had guns and steel for swords, which gave them their military advantage; they carried diseases, such as measles and smallpox, which decimated other populations; and they had political structures that could finance seaworthy ships and organize expeditions that led to the conquest of other lands.

But these are only proximate causes. Diamond digs deeper asking why Eurasians had these advantages more than 500 years ago. Diamond's explanation is geography. Eurasia has a land mass that has an east–west axis whereas Africa and the Americas have a north–south orientation. An east–west axis permitted the more rapid spread of both domesticated animals (cattle, chickens, horses) and edible grains, which in turn permitted more rapid development of settled farming communities. In time, these communities became productive enough to support craftsmen and others who developed the technologies that led to the guns, ships, and steel necessary for foreign adventures and conquest. If a continent has a north–south axis, plant and animal varieties cannot spread as rapidly because they need to adapt to different climates as they move from one area to the next, thus significantly limiting the opportunity for economic growth.

Domesticated agriculture also lies behind the germs that decimated indigenous populations in the Americas and elsewhere. Many infectious diseases result from microbes crossing over from animal populations to humans. Eurasians acquired these diseases and built up their immunity to them as a consequence

of developing settled agriculture early. As more densely populated communities came in close contact with their farm animals, diseases spread as did immunity to them. Eurasia also happened to have more animal species suitable for domestication than other continents. As Diamond writes, "Just think what the course of world history might have been if Africa's rhinos and hippos had lent themselves to domestication! If that had been possible, African cavalry mounted on rhinos and hippos would have made mincemeat of European cavalry mounted on horses."

Sources: Jared Diamond, *Guns, Germs and Steel: The Fates of Human Societies* (New York, W. W. Norton, 1998). A summary is contained in a 1997 talk by Jared Diamond, "Why Did Human History Unfold Differently on Different Continents for the Last 13,000 Years?" April 22, 1997, transcript available at http://edge.org/conversation/why-did-human-history-unfold-differently-on-different-continents-for-the-last-13000-years, accessed February 2012.

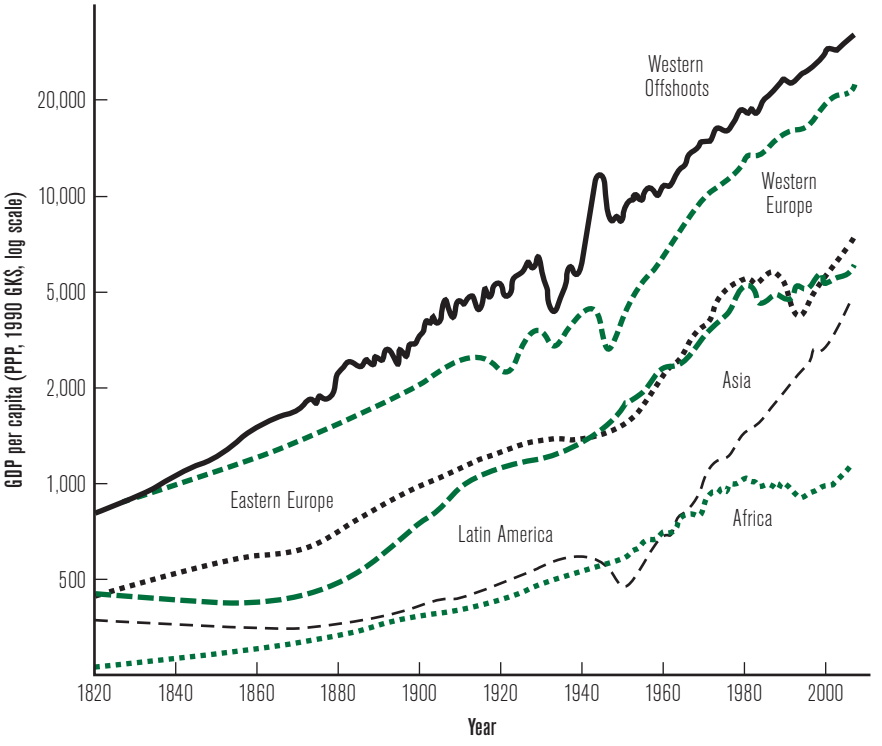


FIGURE 2-1 Levels of GDP per Capita by Region: 1820–2008

Notes: Western offshoots include Australia, Canada, New Zealand, and the United States. GDP, gross domestic product; GK$, Geary-Khamis dollars; PPP, purchasing power parity.
Source: Angus Maddison, "Statistics on World Population, GDP, and Per Capita GDP, 1–2008 AD," www.ggdc.net/MADDISON/Historical_Statistics/vertical-file_02-2010.xls.

In Africa, as elsewhere, average growth rates accelerated after 1820 and did so again after 1950, in the period associated with the end of the colonial era. But as in Latin America, economic growth in Africa faded after 1980, continued to stagnate in the 1990s, and rebounded only recently. As a result, the income gap between the world's richest regions (the Western offshoots) and the poorest in 2000 (Africa) reached 19:1. According to Maddison's work, this is the largest gap in income between rich and poor regions the world has ever known.[9] Because of resurgence in economic growth in Africa during the 2000s, this gap has narrowed but still remains huge by historical standards.

Maddison's broad sweep of world economic history indicates how differential rates of economic growth, especially over the past two centuries, have produced the divergence in income levels that characterizes the world's economy today.

ECONOMIC GROWTH, 1970–2010

Table 2–3 takes a closer look at the pattern of growth rates over the past four decades. The selection of decades as the unit of observation is somewhat arbitrary. The 1970s often are associated with two oil price shocks and other significant changes in commodity prices; the 1980s, with the first wave of international debt crises; the 1990s, with the major transition toward market economies, especially in eastern Europe and the republics of the former Soviet Union; and the 2000s, with the global consequences of the attacks of September 11, 2001, and of the financial crisis of 2008–09. The regional divisions in Table 2–3 differ from those in Maddison and conform to conventions used by the World Bank, a major source of data on economic development. Most of the regional definitions are self-explanatory; however, all high-income economies are combined in one category regardless of geographic location. Therefore, East Asia does not include Japan, Korea, Singapore, Taiwan, and a few small and affluent island economies. Similarly, Europe and Central Asia refers primarily to eastern Europe and Central Asia and excludes all (mostly western) European economies classified as high-income.

The growth rates in GDP per capita reported in Table 2–3 highlight major differences in economic growth both between regions and over time.[10] The 1970s were a decade in which *all* regions experienced positive growth. The 1980s often are referred to as "the lost decade" in Latin America because of the sharp downturn in regional

[9]Using a somewhat different methodology from that employed by Maddison, Lant Pritchett reached similar conclusions in "Divergence, Big Time," *Journal of Economic Perspectives* 11, no. 3 (1997), 3–17.

[10]The growth rates in Table 2–3 are based on constant US$. Although large differences in the *level* of GDP per capita are observed depending on whether PPP or market exchange rates are used, this is not the case when comparing *growth rates* of national income. The growth rates reported in the table are the differences between World Bank estimates of GDP growth and population growth by region and by decade.

TABLE 2–3 Rate of GDP per Capita Growth (Percent/Year)

	1970s	1980s	1990s	2000s
East Asia and Pacific	5.0	6.4	6.1	8.6
Europe and Central Asia	4.4	1.5	−2.9	5.7
Latin America and Caribbean	3.0	−0.3	1.7	2.5
Middle East and North Africa	3.0	−1.1	0.8	2.9
South Asia	1.2	3.5	3.8	5.7
Sub-Saharan Africa	1.1	−1.2	−0.2	2.6
High income	2.4	2.5	1.8	1.3

Sources: World Bank, *World Development Report 1995* (New York: Oxford University Press, 1995). World Bank, *World Development Report 2000/2001* (New York: Oxford University Press, 2001). World Bank, *World Development Report 2011* (Washington, DC: World Bank, 2011).

growth, from +3.0 percent in the 1970s to –0.3 percent in the following 10 years. Negative growth and falling per capita income also were features of the Middle East and North Africa region and sub-Saharan Africa in the 1980s. Growth in sub-Saharan Africa also remained negative, but by a smaller amount, throughout the 1990s. The economies of Europe and Central Asia collapsed in the 1990s after the transition from a planned to a market-based economic system. But the region rebounded during the 2000s, in some cases because of large improvements in commodity prices, especially oil, and in other cases because of the integration of these countries with the global economy. The poor performance of some regions during the 1980s and 1990s stands in sharp contrast to the accelerating growth rates—and the associated improvements in living standards—in both East and South Asia. The term *economic miracle* has been used to describe the historically unprecedented rates of growth achieved by some nations in these two regions.

The growth rates in Table 2–3 also reveal that the 2000s were the best decade since the 1970s for the low- and middle-income nations as a group. The political and economic events that rocked the high-income economies, especially the United States and the European Union, did not have the same impact elsewhere. Such resiliency on the part of low- and middle-income nations was not expected based on the experience of the 1980s and 1990s. Sub-Saharan Africa's improved growth performance is especially noteworthy. Some of it is due to much higher commodity prices. Emerging economies in the region are also benefiting from improved economic policies and management, more democratic and accountable governments, new technologies, and a new generation of development-oriented African leaders and entrepreneurs.[11]

[11]Steven Radelet, *Emerging Africa: How 17 Countries Are Leading the Way* (Washington, DC, Center for Global Development, 2010).

There are a few more points to take away from the growth rates presented in Table 2–3. First, remember that even small differences in growth rates imply huge differences in the potential for economic development. In the 1970s, economic growth in South Asia was 1.2 per year. At this rate, GDP per capita in the region expanded by a mere 12.5 percent in a decade. In the 1980s, South Asia achieved a growth rate a few percentage points higher, 3.5 percent, and ended the decade with a 41 percent increase in GDP per capita. In the 2000s, growth rates again grew by a few more percentage points, reaching 5.7 percent. That decade ended with per capita incomes 75 percent higher than they began. Second, the regional averages in Table 2–3 are weighted averages, where the weights are the population size of each nation in the region. Such averages are heavily influenced by the experience of the most populous country in the region, especially China in East Asia and India in South Asia, and disguise the wide range in individual country performance. For example, in East Asia, the Philippines' annual growth rate from 1979 to 2009 was only 0.7 percent, a fraction of its region's performance; Botswana grew at 4.4 percent per annum over these three decades, far exceeding not only the sub-Saharan Africa average but the performance of most nations worldwide.

The successful growth performance in Asia and a few countries in other regions relative to the high-income economies illustrates an observation by economic historian Alexander Gerschenkron. When Gerschenkron refers to "**the advantages of backwardness**," he is not suggesting that it is good to be poor. Instead, he means that being relatively poorer might allow low-income countries to grow more quickly. For the first nations to experience modern economic growth, in western Europe and its offshoots, growth rates were constrained by the rate of technological progress. That same constraint operates today. Growth rates in the high-income economies reported in Table 2–3 range from 1.3 to 2.5 percent, far lower than the growth rates of the successful regions in the developing world. Poor countries can borrow and adapt existing technology and have the potential to grow faster and to catch up to the more advanced economies. Over the past three decades, this is what enabled growth rates in Asia to exceed the average growth rate of the high-income nations. For development economists, the challenge is to understand why some countries have been able to realize the advantages of backwardness whereas others have fallen further behind.

WHAT DO WE MEAN BY *ECONOMIC DEVELOPMENT*?

As indicated in Chapter 1, economic growth is a necessary but not sufficient condition for improving the living standards of large numbers of people in countries with low levels of GDP per capita. It is necessary because, if there is no growth, individuals can become better off only through transfers of income and assets from others. In a

poor country, even if a small segment of the population is very rich, the potential for this kind of redistribution is severely limited. Economic growth, by contrast, has the potential for all people to become much better off without anyone becoming worse off. Economic growth has led to widespread improvements in living standards in Botswana, Chile, Estonia, Korea, and many other countries.

Economic growth, however, is not a sufficient condition for improving mass living standards for several reasons. First, governments promote economic growth not just to improve the welfare of their citizens but also, and sometimes primarily, to augment the power and glory of the state and its rulers. Governments in developing nations may use national income to expand their militaries or construct elaborate capital city complexes in deserts and jungles. Political leaders may be corrupt and expropriate income for personal gain, whether for conspicuous consumption at home or the accumulation of wealth in overseas bank accounts and property. When gains from growth are channeled in such ways, they often provide little benefit to the country's citizens. Second, resources may be heavily invested in further growth, with significant consumption gains deferred to a later date. In extreme cases, such as the Soviet collectivization drive of the 1930s, consumption can decline dramatically over long periods. When the Soviet Union fell in 1991, its consumers were still waiting for the era of mass consumption to arrive. Normally, the power to suppress consumption to this extent in the name of economic growth is available only to totalitarian governments. Third, income and consumption may increase, but those who already are relatively well off may get all or most of the benefits. The rich get richer, the old saw says, and the poor get poorer. (In another version, the poor get children.) This is what poor people often think is happening. Sometimes, they are right.

If economic growth does not guarantee improvement in living standards, then GDP per capita may not be a meaningful measure of economic development. In addition to problems associated with how income is spent and distributed, any definition of economic development must include more than income levels. Income, after all, is only a means to an end, not an end itself. More than 2,000 years ago, Aristotle wrote, "The life of money-making is one undertaken under compulsion and wealth is evidently not the good we are seeking, for it is merely useful for the sake of something else."

If economic growth and economic development are not the same thing, how should we define economic development? Amartya Sen, economist, philosopher, and Nobel laureate, argues that the goal of development is to expand the *capabilities* of people to live the lives they choose to lead. Income is one factor in determining such capabilities and outcomes, but it is not the only one. To be capable of leading a life of one's own choice requires what Sen calls "elementary functionings," such as escaping high morbidity and mortality, being adequately nourished, and having at least a basic education. Also required are more complex functionings, such as achieving self-respect and being able to take part in the life of the community. Income is but one of the many factors that enhance such individual capabilities.

In his 1998 Nobel address, Sen identified four broad factors, beyond mere poverty, that affect how well income can be converted into "the capability to live a minimally acceptable life":

- *Personal heterogeneities*: including age, proneness to illness, and extent of disabilities.
- *Environmental diversities*: shelter, clothing, and fuel, for example, required by climatic conditions.
- *Variations in social climate*: such as the impact of crime, civil unrest, and violence.
- *Differences in relative deprivation*: for example, the extent to which being impoverished reduces one's capability to take part in the life of the greater community.

According to Sen, economic development requires alleviating the sources of "capability deprivation" that prevent people from having the freedom to live the lives they desire.

Sen's seminal contributions played a key role in the formulation of the human development approach to economic development. This approach is also associated with the UNDP and the work of Pakistani economist Mahbub ul Haq. Part of the motivation of Haq and Sen was a concern over the focus by other development economists on economic growth. Like Aristotle, they wanted to explore how "money-making" was "useful for the sake of something else." They wanted to see the focus shift from the production of commodities to a focus on human lives, including the enhancement of individual capabilities and the enlargement of people's choices.

MEASURING ECONOMIC DEVELOPMENT

The UNDP published its first human development report in 1990 with "the single goal of putting people back at the center of the development process." Although the terminology is different, human development and economic development are the same idea. The distinction is intended to expand the perception of development as encompassing more than increases in per capita income (Box 2–2).

The UNDP attempted to quantify what it saw as the essential determinants of human development: to live a long and healthy life, acquire knowledge, and have access to the resources needed for a decent standard of living. For each of these elements, a specific measure was constructed and aggregated into an index, the **Human Development Index (HDI)**. Every year since 1990, the UNDP has calculated the value of the HDI for as many of the world's nations as the data permit and assessed the relative progress of nations in improving human development. Because the HDI combines outcomes with different units of measurement—years of life expectancy, years of schooling, and dollars of income—each outcome must be converted into an index number to permit aggregation into a composite measure. In response to criticisms of

BOX 2-2 HUMAN DEVELOPMENT DEFINED

Human development is a process of enlarging people's choices. In principle, these choices can be infinite and change over time. But at all levels of development, the three essential ones are for people to lead a long and healthy life, to acquire knowledge, and to have access to resources needed for a decent standard of living. If these essential choices are not available, many other opportunities remain inaccessible.

But human development does not end there. Additional choices, highly valued by many people, range from political, economic, and social freedom to opportunities for being creative and productive, and enjoying personal self-respect and guaranteed human rights.

Human development has two sides: the formation of human capabilities—such as improved health, knowledge, and skills—and the use people make of their acquired capabilities—for leisure, productive purposes, or being active in cultural, social, and political affairs. If the scales of human development do not finely balance the two sides, considerable human frustration may result.

According to this concept of human development, income is clearly only one option that people would like to have, albeit an important one. But it is not the sum total of their lives. Development must, therefore, be more than just the expansion of income and wealth. Its focus must be people.

Source: United Nations Development Programme, *Human Development Report 1990* (Oxford: Oxford University Press, 1990), p. 10.

the index, the HDI has evolved over time, including different variables and changing how the index is computed. In the *Human Development Report 2010*, the 20th anniversary of the HDI, significant changes were made to the variables used, the construction of the indices for each dimension, and the method of aggregation.

As a proxy for living a long and healthy life, the HDI employs a nation's life expectancy at birth and compares progress on this measure relative to other nations. The goalposts for assessing life expectancy are a minimum value of 20 years and a maximum of 83.2. Twenty years represents the minimum life expectancy that permits a society to sustain itself. Anything less than 20 years would be below the prime reproductive age range and a society would eventually die out. Historical evidence bears this out. The maximum value of 83.2 years is what Japan achieved in 2010, the highest level recorded in any nation over the 20 years the HDI has been calculated. A country's score on this dimension is a measure of its populations' life expectancy compared to the maximum and minimum scores. For example, in 2010, El Salvador

had life expectancy at birth of 72 years. El Salvador's HDI life expectancy index is calculated as (72 – 20) ÷ (83.2 – 20) = 0.82; in other words, El Salvador has attained 82 percent of the potential range in life expectancy.

As a proxy for acquiring knowledge, the HDI includes two variables. One is the mean years of schooling achieved by the adult population, those 25 years and older. The second variable is expected years of schooling for children of school-going ages. It is based on enrollment data. The goalposts of the adult schooling variable are 0 and 13.2, the observed maximum from the United States. For expected years of schooling, the goalposts are 0 and 20.6, the maximum referring to Australia. El Salvador's mean years of schooling among adults is 7.7 years (58 percent of the way in between the two goalposts); for expected years of schooling it is 12.1 years (59 percent). These values are then aggregated to form one composite index for education.

Access to resources is measured by transforming GNI per capita (PPP in US$).[12] The goalposts are $163 and $108,211; the minimum is the value attained in Zimbabwe in 2008, and the maximum is from the United Arab Emirates in 1980. The relative standing of a nation's GNI per capita is determined by taking the logarithms of all dollar values. The transformation into logarithms decreases the significance of income gains as income increases. This reflects the conclusions made by all the human development reports that there are diminishing returns to income as a means of securing a *decent* standard of living (or, alternatively, that the marginal utility of an extra dollar of income falls as income rises). El Salvador, with an estimated 2010 GNI per capita of $6,498, falls 57 percent between the logarithm-adjusted income goalposts (Box 2–3).

All three dimensions of the HDI are expressed in terms of a percentage, solving the problem of different units of measurement. The next challenge is how to aggregate the three dimensions. Up until 2010, this was done by giving the index of each dimension an equal weight of one third and computing the *arithmetic mean* of the three. In 2010, the three dimensions still have equal weight but now a *geometric mean* of the three percentages is computed.[13] The UNDP explains the reason for this change. Using an arithmetic mean implied "perfect substitutability" among the three components of HDI. If a nation lost, say, 10 percent on its schooling measure this could be compensated by a 10 percent improvement in income and the HDI would remain unchanged. The geometric mean does not have this property. Low achievement in one dimension is no longer linearly compensated by high achievement in another dimension. The level of each index matters, creating a situation of imperfect substitutability across

[12]Before 2010, the HDI was based on GDP values. The change to GNI was made to include remittances and foreign assistance income and to exclude income generated within a country but repatriated abroad. All three can be significant, especially for low-income countries. This recalls our earlier discussion of the differences between GDP and GNP/GNI.

[13]The calculation requires taking the cube root of the product of the three indices on life expectancy, schooling, and income.

BOX 2-3 WHY USE LOGARITHMS?

Logarithms have been used several times in this chapter and will be referred to throughout this text. It is worth reviewing some of their properties. It is easiest to understand logarithms if you remember that the answer to any question involving a logarithm is an exponent. If we want to know the logarithm of 100 in base 10, the answer would be 2, because 10 raised to the second power is 100. In mathematical notation,

$$\log_{10}(100) = 2 \text{ because } 10^2 = 100$$

The base of the logarithm merely determines what number will be raised to a given power to arrive at the value whose logarithm we are seeking. If our base were 2 instead of 10, then the logarithm of 16 would be 4, because 2 raised to the fourth power is 16:

$$\log_{2}(16) = 4 \text{ because } 2^4 = 16$$

The HDI uses the logarithm of income per capita rather than the level of income per capita in determining the importance of incomes in a country's human development. To see the implications, we build off our example. Assume that two nations have incomes of 10 and 100, respectively. Using these values implies a 10-fold importance to the higher income because 100 is 10 times as large as 10. What if we use logarithms instead? If we use base 10 logarithms, we can easily see how the relative importance of the higher income is reduced. As we already determined, the logarithm of 100 in base 10 is 2. What is the logarithm of 10 in base 10? It is 1, because 10 raised to the first power is 10. Using logarithms in this case reduced the relative importance of the higher income from a factor of 10 to a factor of only 2.

This reduced *relative importance* becomes more pronounced as the spread between incomes grows. If the two incomes were 10 and 1,000, using base 10 logarithms reduces the relative importance of the higher income from a factor of 100 to a factor of only 3! Logarithms are consistent with the UNDP's position that income exhibits diminishing returns in achieving human development.

Economists, including the authors of the HDI, often use the mathematical constant ***e*** as the base for their logarithms; these are called natural logarithms. Base ***e*** is closely related to the concept of continuous compounding and is a useful tool when examining economic variables that grow over time. Box 3–2 discusses this further.

One more property of logarithms is worth noting. If we track a variable in logs rather than in levels, differences in the log values imply equal *percentage* changes. For example, using logarithms in base 2,

$$\log_2(2) = 1 \text{ as } 2^1 = 2$$

$$\log_2(4) = 2 \text{ as } 2^2 = 4$$

$$\log_2(8) = 3 \text{ as } 2^3 = 8$$

$$\log_2(16) = 4 \text{ as } 2^4 = 16$$

As we move from the log values 1 to 2 to 3 to 4, the absolute change of the underlying value increases by 2 then 4 then 8, but the percentage change always remains the same at 100 percent.

This property of logarithms can be observed by considering the logarithmic scales used in Figures 2–1 and 2–2. In Figure 2–2, notice that the distance between $500 and $1,000 on the horizontal axis is the same as the distance between $20,000 and $40,000. We know that the absolute change is different along those two intervals (clearly $500 is smaller than $20,000), but both intervals represent a doubling (a 100 percent increase) of the variable. GNI per capita expressed in logarithms must behave this way.

the HDI's three dimensions. The UNDP argues that a geometric mean better reflects intrinsic differences across the indices than did the arithmetic mean.

Once the geometric mean is applied, El Salvador had a 2010 HDI of 0.659, placing it in the medium human development range, ranked 90th out of the 169 nations for which an HDI was computed. El Salvador's HDI improved from an estimated value of 0.456 in 1980, to 0.562 in 1995, to its most recent value of 0.659 in 2010. These trends suggest significant progress in El Salvador's human development over the past 30 years.

WHAT CAN WE LEARN FROM THE HUMAN DEVELOPMENT INDEX?

The basic concept behind human development is one with which many people would agree. But when we move from concept to measurement, problems arise. Many criticisms have been leveled against the HDI since it was first introduced. Some are concerned with limiting the index to only three dimensions of human development. In response, the human development reports now compute additional indices focusing on human poverty and gender-related development. An inequality-adjusted HDI was added in 2010.

Some commentators criticize the HDI for assuming diminishing returns only to income but not to either life expectancy or schooling. If the marginal utility of income declines with the more income one has, can the same be said for an extra year of life or schooling? Specific criticisms also have been raised about the introduction of a geometric mean to aggregate the three dimensions. Geometric means are

sensitive to low values; in the extreme, if one index equaled zero then the HDI would equal zero too. The same principle applies if one index is close to zero. As economist William Easterly put it, "The new HDI has a 'you're only as strong as your weakest link' property, and in practice the weakest link turns out to be very low income."[14] This is due to both the construction of the HDI and to the greater convergence across nations in life expectancy and schooling than in income. The implication for nations with the lowest HDI, including Zimbabwe, the Democratic Republic of the Congo, and Niger seems to be focus on economic growth rather than health or education. This seems the opposite of the message behind the entire human development approach. It may be the right message for these nations but it is also an unintended artifact of how the HDI is now constructed.[15]

Beyond these criticisms lies the central question of how much of an improvement the HDI is over national income per capita as an index of economic or human development. Figure 2–2 presents a comparison of HDI values and levels of GNI per capita. The scatter diagram of values for individual countries indicates, as might be predicted, that rising incomes raise the HDI. This is expected because incomes are a component of the HDI, and health and schooling also rise with incomes. According to the trend line, income alone explains 90 percent of the variation in the HDI. But the scatter diagram also indicates variance around this trend. Angola and Georgia have similar levels of per capita GNI (around US$4,900, PPP) but Angola has much lower life expectancy than does Georgia (48 versus 72 years). There is also an almost eight-year gap in schooling between the two nations. Despite identical per capita incomes, there is low human development in Angola, with an HDI of 0.403, and high human development in Georgia, with an HDI of 0.698. If one compares Mozambique and Togo the story is similar. Both have per capita incomes of close to $850, but life expectancy and schooling are much higher in Togo, raising its HDI value. In terms of HDI values, Angola and Mozambique lay well below the trend line in Figure 2–2, Georgia and Togo lay well above it.

We can conclude that alternative measures of economic development are significantly but not perfectly correlated with levels of income. This suggests that with economic growth, increasing levels of income can predict a lot about economic development. But the data also suggest that improved health and education depend on factors other than income. We elaborate on this point in the discussion of income and health in Chapter 9.

From an advocacy perspective, the HDI has been useful in calling attention to development issues. It is widely reported in the media and gets the attention of

[14]William Easterly, "The First Law of Development Stats: Whatever Our Bizarre Methodology, We Make Africa Look Worse," *AIDWATCH* (blog), December 2, 2010. http://aidwatchers.com/2010/12/the-first-law-of-development-stats-whatever-our-bizarre-methodology-we-make-africa-look-worse/.

[15]Allen Kelley, "The Human Development Index: 'Handle with Care,'" *Population and Development Review* 17, no. 2 (June 1991), 315–24 provides an early critique of the original HDI. Concerns about the 2010 version are discussed in Martin Ravallion, "Troubling Tradeoffs in the Human Development Index," Policy Research Working Paper 5484, World Bank, November 2010.

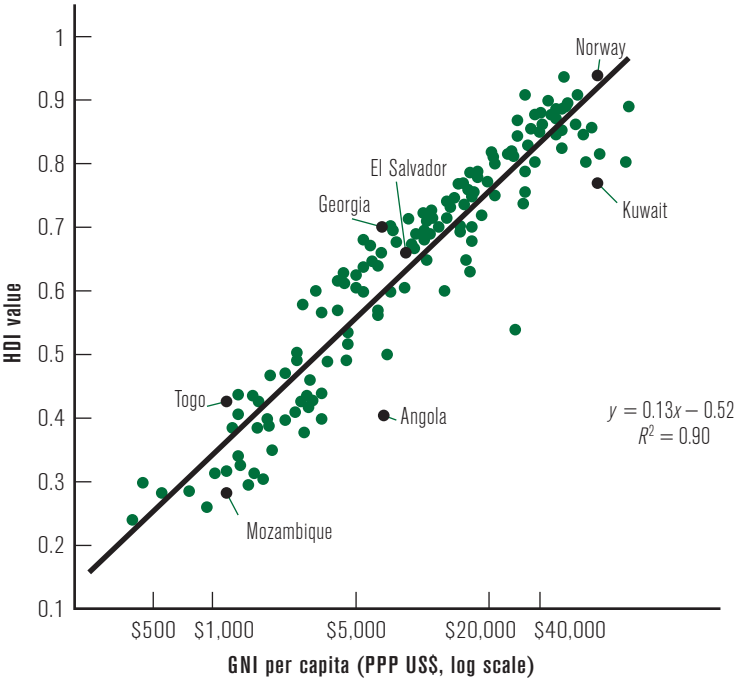


FIGURE 2–2 HDI Versus GNI per Capita by Country (2010)

Source: UN Development Programme, *Human Development Report 2010* (New York: UN Development Programme, 2010), http://hdr.undp.org/en/media/HDR_2010_EN_Contents_reprint.pdf.

political leaders. But the construction of the HDI is far from transparent and the aggregation of different dimensions of human development may not be better than a dashboard of several indicators. How much additional insight the HDI offers as a means of *measuring* economic development remains open to debate.

MILLENNIUM DEVELOPMENT GOALS

Defining economic development is inherently difficult. As with any normative concept, people have different opinions as to what should be included in the definition and on what weight to give to different goals. But even without a commonly agreed-on definition, policy makers need specific targets. One such set of targets is known as the **millennium development goals (MDGs)**.

In September 2000, 189 nations adopted the "United Nations Millennium Declaration," a broad-reaching document that states a commitment "to making the right to development a reality for everyone and to freeing the entire human race from want."[16] The declaration specifies a set of eight goals consistent with this commitment:

[16]UN General Assembly, "United Nations Millennium Declaration," section III, paragraph 11, September 18, 2000.

- Goal 1. Eradicate extreme poverty and hunger.
- Goal 2. Achieve universal primary education.
- Goal 3. Promote gender equality and empower women.
- Goal 4. Reduce child mortality.
- Goal 5. Improve maternal health.
- Goal 6. Combat HIV/AIDS, malaria, and other diseases.
- Goal 7. Ensure environmental sustainability.
- Goal 8. Develop a global partnership for development.

To more fully define these goals, a panel of experts developed a comprehensive set of targets and indicators for each of the MDGs. The eight MDGs contain 21 targets, which correspond to 60 indicators (Box 2–4). This combination of multiple goals, targets, and indicators is an articulation of what most of the world's governments believe should be achieved to make "development a reality for everyone." The

BOX 2–4 TARGETS OF THE MILLENNIUM DEVELOPMENT GOALS

Target 1A.	Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day.
Target 1B.	Achieve full and productive employment and decent work for all, including women and young people.
Target 1C.	Halve, between 1990 and 2015, the proportion of people who suffer from hunger.
Target 2A.	Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling.
Target 3A.	Eliminate gender disparity in primary and secondary education, preferably by 2005, and to all levels of education no later than 2025.
Target 4A.	Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate.
Target 5A.	Reduce by three-quarters, between 1990 and 2015, the maternal mortality ratio.
Target 5B.	Achieve, by 2015, universal access to reproductive health.
Target 6A.	Have halted by 2015 and begun to reverse the spread of HIV/AIDS.
Target 6B.	Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it.
Target 6C.	Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases.

Target 7A.	Integrate the principles of sustainable development into country policies and programs and reverse the loss of environmental resources.
Target 7B.	Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss.
Target 7C.	Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation.
Target 7D.	By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers.
Target 8A.	Develop further an open, rule-based, predictable, nondiscriminatory trading and financial system.
Target 8B.	Address the special needs of the least developed countries.
Target 8C.	Address the special needs of landlocked countries and small island developing states.
Target 8D.	Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long run.
Target 8E.	In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries.
Target 8F.	In cooperation with the private sector, make available the benefits of new technologies, especially information and communications.

Source: United Nations Statistics Division, "Official List of MDG Indicators," January 2008, available at http://unstats.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm, accessed February 2012.

millennium declaration even suggests ways in which this development agenda might be financed. For example, Target 8B, which focuses on the least developed countries, makes recommendations for debt relief and for more official development assistance from the rich nations. But none of these recommendations is binding.

The MDGs were drawn up in 2000 and their due dates are soon approaching. It is possible to assess progress thus far. Poverty reduction at the global level, for which the poverty indicator is living on less than US$1.25 (PPP) per day, is on track to meet or exceed the global target by 2015 (although the target will not be met in every country or region). Access to clean drinking water is proceeding well and should exceed the original target. More disappointing are trends in under-five mortality. In 1990, the developing regions experienced child mortality in the neighborhood of 100 deaths per 1,000 children. This fell to 66 per 1,000 by 2009. To reach the target of a two-thirds reduction in child mortality by 2015 will take, according to the United Nations, "substantial and accelerated action," especially in Sub-Saharan Africa and South Asia where child mortality rates remain high. Target 6B, universal treatment for HIV/AIDS, has met with some success. But because only 35 to 40 percent of those who

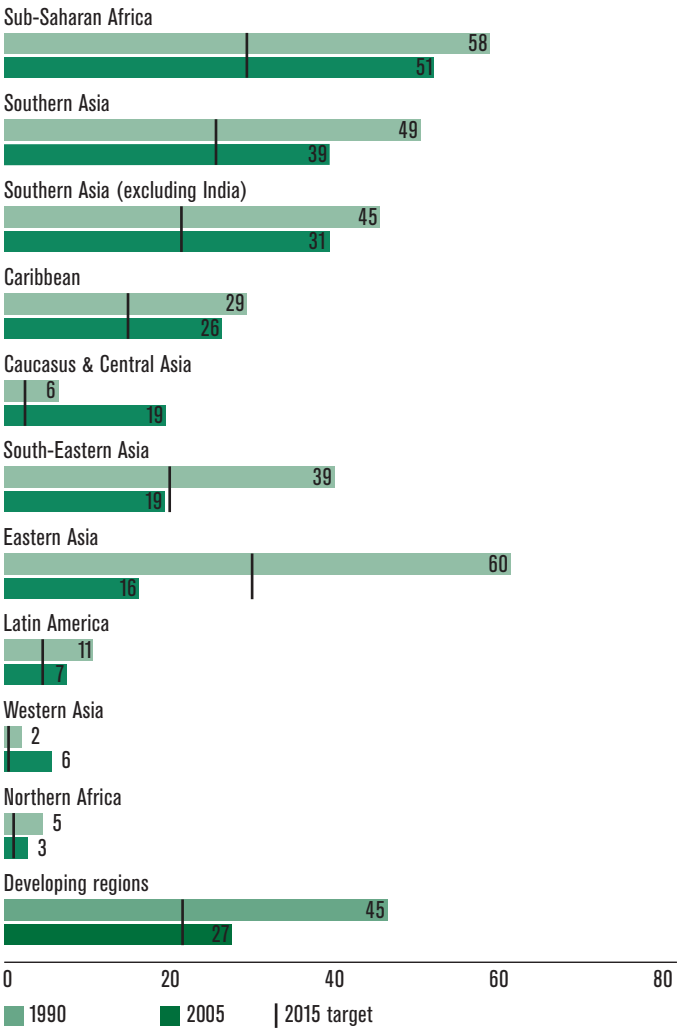


FIGURE 2–3 Proportion of People Living on Less Than $1.25 a Day, 1990 and 2005 (Percentage)

Source: Figure from the *Millennium Development Goals Report*, 2011. The World Bank. Reprinted by permission of United Nations Publications.

needed antiretroviral treatment received it in 2009, universal coverage will not be achieved by the target date. Universal treatment by 2010 was an ambitious and probably unrealistic goal.

Evaluating progress toward achieving the MDGs requires caution, especially regarding the level of aggregation across countries. Relative success and failure at meeting the MDGs varies not only among specific targets but also by region. Some of the explanation for differential performance is based on disparities in the rate of economic growth between countries. Success at cutting global poverty in half owes

a lot to China's performance alone. China reduced extreme poverty from 60 percent of its population in 1990 to only 16 percent by 2005. Yet Figure 2–3 illustrates that this global result does not describe progress in specific regions. Based on data for 2005, the most recent data available, Sub-Saharan Africa and South Asia appear far from being on track to halve poverty by 2015.

Rapid economic growth is widely seen as responsible for China's success at poverty alleviation and, by implication, much of global achievements. But *some* of the difficulty in achieving other MDGs is due to factors beyond economic growth. Better use and distribution of income and specific strategies aimed at some of the targets (such as reducing infant mortality through disease control), combined with economic growth, are essential to achieving the United Nations' goal of "freeing the entire human race from want." By setting targets, the MDGs have focused the attention of governments in poor nations and of donors in rich ones to achieve specific outcomes that promote economic development.

The MDGs have been challenged on the basis of including too much and setting targets that may be either too high or too low based on historical experience.[17] The MDGs also fail to address the fundamental economic problem of trade-offs and priorities. If one cannot fulfill all 21 targets simultaneously, which takes precedence: maternal mortality or access to safe drinking water, reducing hunger or promoting environmental sustainability? This is less of a problem for defining development: Economic development involves all these goals. But it is a practical problem for those charged with realizing such an ambitious development agenda.[18]

IS ECONOMIC GROWTH DESIRABLE?

After discussing the MDGs, which convey the absolute and relative deprivation of so many people around the world, and recognizing the positive correlation between economic growth and human development, it may seem odd to end this chapter by asking, Is economic growth desirable? The answer would seem to be an obvious and emphatic yes! But there are other perspectives. Some decry the spread of materialism, the Westernization of world cultures, and the destruction of traditional societies that seem to accompany economic growth. Others are troubled by environmental degradation, whether species loss or global warming, that has accompanied rising per capita incomes. Still others in the high-income world may wonder if we should be so quick to encourage people to follow the path we have taken: seemingly insatiable

[17]Michael Clemens, Charles Kenny, and Todd Moss, "The Trouble with the MDGs: Confronting Expectations of Aid and Development Success," *World Development* 35, no. 5 (2007), 735–51; and William Easterly, "How the Millennium Development Goals Are Unfair to Africa," *World Development* 37, no. 1 (2009), 26–35.

[18]A strategy for achieving the MDGs is laid out in a report by the UN Millennium Project, *Investing in Development: A Practical Plan to Achieve the UN Millennium Development Goals* (New York: United Nations Development Programme, 2005).

consumerism, the withering of extended and nuclear families, high levels of stress, and all the other ills associated with modern life.

Richard Easterlin, an economic historian, once observed that, although per capita incomes in the United States had risen dramatically over the preceding half century, people did not seem to be any happier. He based this conclusion, which came to be known as the **Easterlin paradox**, on survey data taken over time in which people were asked how happy they were with their lives. Easterlin found similar results when looking across a small number of high-income nations. The Easterlin paradox and, more generally, the analysis of subjective well-being, often referred to as the study of happiness, have received a great deal of attention over the past decade by economists and other social scientists. This research has been enabled by newly conducted surveys of happiness and life satisfaction covering countries at all income levels. The motivation behind this research has included many of the concerns raised throughout this chapter about whether GDP or GNI per capita is a useful measure of economic well-being.

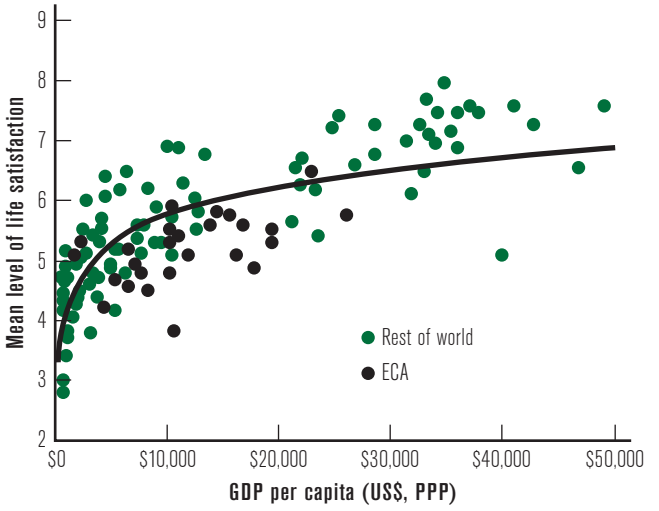
How can happiness be measured? One set of surveys asks individuals, "Taking all things together, would you say you are: very happy, quite happy, not very happy, or not at all happy?" Responses are coded from 1 to 4, with 4 signifying "very happy." Easterlin's work was based on questions like this. Another set of surveys directed at life satisfaction asks a related but different question:

> Please imagine a ladder with steps numbered from 0 at the bottom to 10 at the top. Suppose we say that the top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?

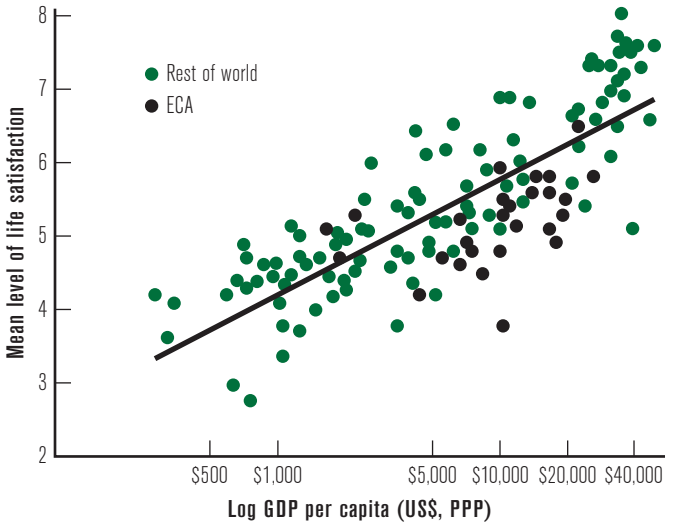
Surveys that ask both of these questions find that responses are well but not perfectly correlated. Happiness and life satisfaction are not identical.

Based on data from the life satisfaction question, which include a relatively large sample of countries, the plots in Figure 2–4a show the life satisfaction score against GDP per capita (US$, PPP). In Panel b, GDP per capita (US$, PPP) is entered in log values. Both graphs tend to reject the Easterlin paradox. Life satisfaction rises with per capita income. One reason Easterlin did not observe this relationship is the small sample of rich nations he had to work with. The graph using log values suggests that, in relative terms, life satisfaction increases with percentage changes in per capita incomes across the entire range of observed incomes. Although economic growth and happiness are correlated, within income categories (low, middle, and high) there appears to be much less of a trend than between them. Among any given income category the Easterlin paradox may hold.[19]

[19]Angus Deaton, "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll," *Journal of Economic Perspectives* 22, no. 2 (Spring 2008), 53-72; Betsey Stevenson and Justin Wolfers, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity* 39, no. 1 (Spring 2008), 1–102.



(a)



(b)

FIGURE 2–4 Life Satisfaction as a Measure of Well-Being (2008)

Life satisfaction measures the mean level of life satisfaction of a random group of people who were asked the following question: "Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?"

Source: Gallup World Poll.

There is also variance around the trend line because more than income levels determine life satisfaction. Costa Rica and the United States both have mean levels of life satisfaction of around 7.2, but Costa Rica has only one-quarter the income of the United States; Korea is more than three times richer than Thailand, but both share similar levels of life satisfaction (around 6.1). For reasons other than income, Costa Ricans and Thais are more satisfied than are Americans and Koreans. One group of nations stands out in terms of lower life satisfaction than expected based on income alone. Countries in Europe and Central Asia almost always lie below the trend line. After accounting for income level, mean life satisfaction is systematically lower for these nations than it is in other regions. Much of this can be explained by the wrenching effects of the transition from planned to market economies. People who benefited from the former system, including those who had decent jobs or relied on pensions, were made worse off. Public goods deteriorated, inequality rose, and growth rates turned negative. All contributed to declining measures of average life satisfaction. There is some evidence that with the resumption of growth in many of these economies during the 2000s, levels of happiness are rising and falling more in line with other regions.[20]

Recent research suggests that happiness and income levels are correlated, implying that economic growth improves happiness. But even in those situations in which the relationship is less certain, a case for economic growth still can be made. More than half a century ago, W. Arthur Lewis, one of the pioneers of the field of development economics and a Nobel laureate, provided one. Lewis was writing before there was survey evidence on happiness and at a time when the distinction between the terms *economic growth* and economic development were not as nuanced as they are today. The concluding chapter of his 1955 book, *The Theory of Economic Growth,* is titled, "Is Economic Growth Desirable?" Lewis's answer: "The case for economic growth is that it gives man greater control over his environment, and thereby increases his freedom."[21]

SUMMARY

- Understanding the process of economic development requires methods of measuring economic performance across countries and over time. Gross domestic product (GDP) per capita is a measure of the aggregate value of national output and, despite a number of limitations, is the most common standard for measuring economic growth.
- For cross-country comparisons, GDP per capita is best measured in terms of purchasing power parity (PPP). PPP estimates are superior to

[20]Sergei Guriev and Ekaterina Zhuravskaya, "(Un)Happiness in Transition," *Journal of Economic Perspectives* 23, no. 2 (Spring 2009), 143–68.

[21]W. Arthur Lewis, *The Theory of Economic Growth* (Homewood, IL: Richard D. Irwin, 1955), p. 420–21.

comparisons based solely on market exchange rates. Market exchange rates tend to underestimate the GDP levels of poorer nations. This is because market exchange rates are based on traded goods and capital flows and fail to account for the much lower prices of nontraded goods in poor nations. The PPP estimates obtained from the UN International Comparison Program correct this problem by expressing every nation's GDP per capita in terms of a common set of international prices.

- Angus Maddison provides a broad overview of economic growth over the past 2,000 years. For most of this period and in most regions, growth in output was just about sufficient to match growth in population, resulting in more or less stagnant per capita incomes. According to Maddison, modern economic growth began only around 1820, with the escalation of growth rates in western Europe and its offshoots. These regions had achieved higher per capita income levels by 1500 but accelerated these differences, especially, over the last 190 years. Rapid growth rates began to characterize other regions, especially in Asia, only in the decades after 1950. As a result of these divergent patterns of economic growth, there is much greater disparity in world incomes today than in the past.
- Unlike economic growth, which is a relatively objective measure of economic capacity, economic development is a normative concept. Various scholars and organizations offer specific indicators or goals for achieving development. Prominent among these is the human development index (HDI), a composite measure reflecting the goals of leading a long life, acquiring knowledge, and achieving material well-being. The millennium development goals (MDGs) rely on a multiplicity of goals and targets for advancing human well-being within the decade.
- Analysis of the HDI and the MDGs reveals the centrality of economic growth to achieving both. HDI ranks are well correlated with per capita income levels, and differential performance in achieving specific MDGs often is tied to differential growth performance. But economic growth is not a panacea. Achieving economic development also involves questions of distribution and strategies to reach specific development targets.
- Citizens of rich nations can engage in a debate over whether further economic growth will advance their well-being, but this is not a meaningful debate for poor nations. Even if rising per capita incomes are not perfectly correlated with perceptions of happiness, economic growth and development are essential for poor nations as a means of increasing choices and advancing human capabilities and freedoms.



Economic Growth: Concepts and Patterns

Why are some countries rich and others poor? Why do some economies grow quickly with their citizens enjoying rapid increases in their average incomes, while others grow slowly or not at all? How did some East Asian countries advance from poverty to relative prosperity in just 30 years, while many African countries remain mired in deep poverty, with few signs of sustained growth and development? These are some of the most important questions in the study of economics and indeed touch on some of the deepest problems facing human society.

As we saw in the last chapter, rapid economic growth and wide divergences of economic performance across countries are fairly recent phenomena in world history. Up until about 500 years ago—a relatively short period of time in human history—most people lived in conditions that we now would consider abject poverty. Housing was poor, food supplies were highly variable and dependent on the weather, nutrition was inadequate, disease was common, healthcare was rudimentary, and life spans rarely exceeded 40 years. Even as recently as 125 years ago, the vast majority of people living in the world's most modern cities, including New York, London, and Paris, lived in extremely difficult conditions on very meager incomes. One decade into the twenty-first century, however, income levels around the world generally are both much higher and far more diverse. A significant minority of the world's population has recorded relatively rapid and sustained income growth during the last several decades and now enjoys much longer and healthier lives, higher levels of education, and much-improved standards of living. Other countries have achieved important but more modest gains and now are middle-income countries.

The majority of the world's population, however, continues to live in poverty, in most cases better off than their ancestors but living at levels of income and welfare far below those of the world's richest countries.

Economic growth and economic development are not synonymous, as we discovered in Chapter 2. Yet economic growth is at the heart of the development process, and sustained development and poverty reduction cannot occur in the absence of economic growth. This chapter and the next explore in some detail the puzzles of economic growth and divergent levels of income across countries. Our objectives are to better understand the processes by which economies grow and develop and the characteristics that distinguish rapidly growing economies from those with slower growth. This chapter explores the basic empirical data on economic growth, the concepts underlying the leading ideas of what causes growth, and some of the patterns of changing economic structure that typically accompany growth. Chapter 4 expands the analysis by introducing some formal models of economic growth. Later chapters explore other dimensions of development introduced earlier, including the distribution of income, poverty, and improvements in health and education.

DIVERGENT PATTERNS OF ECONOMIC GROWTH SINCE 1960

As we begin to explore differences in growth rates, consider Thailand and Zambia. In 1960, the annual income of the average Zambian was almost twice as high as that of the average Thai, around $1,800 in Zambia and $960 in Thailand in constant 2005 purchasing power parity (PPP) dollars. Since that time, Thailand achieved very rapid economic growth of around 4.3 percent per person per year, so that the average income in Thailand is now around $7,800. In 2009, the income of the average Thai was over *eight times* higher in real terms than that of his or her grandparents 50 years ago. As a result, Thais can consume much more (and much higher-quality) food, housing, clothing, healthcare, education, and consumer goods. Thais are better off in other ways as well: life expectancy increased from 54 to 69 years, infant mortality dropped from 103 to 12 per thousand (meaning that an additional 91 children per thousand infants lived beyond their first birthday), and the percentage of adults that were literate grew from 80 (in 1970) to 94 percent. In Zambia, by contrast, average income actually fell slightly to $1,765, about 2 percent lower than in 1960. Life expectancy remained at around 45 years, pulled down to a large extent because of the dramatic spread of the HIV/AIDS pandemic in the 1990s. Infant mortality rates improved (from 126 to 86 per thousand), as did literacy rates (from 48 to 71 percent), but overall Zambians are arguably worse off than their parents and grandparents.

The different growth records of Thailand and Zambia are mirrored by the experiences of many other developing countries, as shown in Table 3–1. Madagascar joins

TABLE 3–1 Economic Growth across Countries, 1960–2009

COUNTRY	GDP PER CAPITA (2005 PPP) 1960	GDP PER CAPITA (2005 PPP) 2009	RATIO OF 2009 TO 1960 GDP/ CAPITA	AVERAGE ANNUAL GROWTH RATE (%)	INCOME AS A SHARE OF U.S. INCOME* 1960	INCOME AS A SHARE OF U.S. INCOME* 2009
Negative growth						
Madagascar	842	753	0.89	−0.23	0.07	0.02
Zambia	1,803	1,765	0.98	−0.04	0.14	0.04
Slow growth						
Senegal	1,421	1,492	1.05	0.01	0.11	0.04
Kenya	1,020	1,206	1.18	0.34	0.08	0.03
Rwanda	860	1,031	1.2	0.37	0.07	0.03
Nigeria	1,528	2,034	1.33	0.58	0.12	0.05
Venezuela	6,663	9,115	1.37	0.64	0.52	0.22
Chad	819	1,277	1.56	0.91	0.06	0.03
Jamaica	5,609	8,795	1.57	0.92	0.44	0.21
El Salvador	3,397	6,338	1.87	1.28	0.26	0.15
Argentina	6,244	11,961	1.92	1.33	0.49	0.29
Peru	3,759	7,280	1.94	1.36	0.29	0.18
South Africa	3,850	7,589	1.97	1.4	0.3	0.18
Ghana	603	1,239	2.05	1.48	0.05	0.03
Philippines	1,314	2,838	2.16	1.58	0.1	0.07
Moderate growth						
Turkey	3,243	9,909	3.06	2.31	0.25	0.24
Papua New Guinea	887	2,746	3.1	2.34	0.07	0.07
Brazil	2,581	8,160	3.16	2.38	0.2	0.2
Chile	3,780	11,999	3.17	2.39	0.29	0.29
Pakistan	728	2,353	3.23	2.43	0.06	0.06
Lesotho	401	1,311	3.27	2.45	0.03	0.03
Dominican Republic	2,355	9,911	4.21	3.06	0.18	0.24
Mauritius	2,208	9,484	4.29	3.01	0.17	0.23
Rapid growth						
India	711	3,239	4.55	3.14	0.06	0.08
Egypt	1,036	4,956	4.78	3.24	0.08	0.12
Sri Lanka	765	4,035	5.27	3.45	0.06	0.1
Indonesia	693	4,075	5.88	3.69	0.05	0.1
Malaysia	1,470	11,296	7.68	4.25	0.11	0.27
Thailand	961	7,794	8.11	4.36	0.07	0.19
Singapore	4,300	47,373	11.02	5.02	0.33	1.15
South Korea	1,782	25,034	14.05	5.54	0.14	0.61
Botswana	578	8,872	15.35	5.73	0.05	0.22
China	403	7,634	18.94	6.18	0.03	0.19
Industrialized countries						
United Kingdom	12,841	33,383	2.6	1.97	1	0.81
United States	15,438	41,099	2.66	2.02	1.2	1
Canada	12,988	36,209	2.79	2.11	1.01	0.88
France	10,101	30,822	3.05	2.31	0.79	0.75
Japan	5,850	30,008	5.13	3.4	0.46	0.73

*Growth rates are trend-growth calculated by ordinary least squares regressions and do not necessarily match endpoint-to-endpoint growth rates. Growth rates are listed in order of magnitude, but the ordering of magnitudes in the third and fourth columns sometimes differ.

PPP, purchasing power parity; GDP, gross domestic product.

Source: Penn World Tables 7.0, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed May 2011.

Zambia in the unfortunate experience of shrinking incomes. Because incomes were very low to begin with in many of these countries, negative growth has been a major tragedy. A second group of countries attained positive growth but at relatively slow rates. Average incomes in these countries increased but not by as much as in many other countries around the world. In Venezuela, for example, per capita growth averaged about 0.6 percent per year since 1960, enough for incomes to increase by around 37 percent, but less than what many Venezuelans may have hoped for.

A third group of countries has been more successful and has achieved moderate growth, shown in the table as per capita growth between 2 and 3 percent per year. By world historical standards (calculated by Angus Maddison and shown in Figure 2–1), these growth rates are relatively high and allow for solid increases in average income. For example, Egypt recorded a growth of 3.2 percent a year, enough for average income to double every 22 years and nearly quintuple between 1960 and 2009, a significant achievement. India, home to over 1 billion people, saw similar growth.

A fourth group of countries has done even better, recording rapid growth of more than 3 percent per capita per year. A few have achieved extraordinary growth exceeding 5 percent, including Singapore, South Korea, Botswana, and China. These are some of the fastest growth rates ever recorded over a 50-year period in the history of the world, and they have led to enormous changes. In South Korea, income expanded by a mind-boggling factor of 14, while in Botswana (Box 3–1), average income is *more than 15 times* what it was 50 years ago. (Unfortunately, Botswana's great success is now threatened by the HIV/AIDS pandemic, as we discuss later in the book.) China, where growth has averaged 6 percent per year (with almost all of it coming after 1980, meaning the growth rate since then has been much faster), has undergone perhaps the most remarkable transformation of all: For one-fifth of the world's population, including a huge number of people living in or near poverty, average incomes have increased by a factor of nearly 19 between 1960 and 2009. China's rapid growth is one of the most important events of the last century, and its continued growth will have profound implications for this century, as well.

There are some clear regional differences in growth rates, as the data in Table 3–1 suggest and as we saw in Table 2–3. Most of the rapidly growing countries are in East Asia, while most of the slowly growing countries are in Africa. But beware of taking these general statements too far because there are important exceptions. In East Asia, Myanmar (Burma) and Laos both recorded slow growth, and the Philippines' performance has been modest at best. In Africa, Botswana and Mauritius are among the fastest-growing countries in the world, and tiny Lesotho and Swaziland have also achieved steady growth.

Remember, apparently small differences in growth rates can make a huge difference over time. The difference between 1 percent growth and 2 percent growth is huge: It is not a 1 percent difference but a 100 percent difference. With growth of 1 percent per year, average income increases by about 65 percent over 50 years, and

BOX 3–1 BOTSWANA'S REMARKABLE ECONOMIC DEVELOPMENT[a]

Whereas most of sub-Saharan Africa has achieved little or no economic growth in per capita income on average over the past 50 years, Botswana stands out as a clear exception. Indeed, during the 20-year period between 1970 and 1990, Botswana was the fastest-growing economy in the world, with growth averaging an astonishing 7.9 percent per year, easily outpacing the more widely noted rapid growth in Singapore (6.3 percent), Korea (6.9 percent), and other countries around the world. Table 3–1 shows that over the longer period from 1960 to 2009, Botswana's growth rate was second only to China's. Over the 50-year period, average real income increased by a factor of more than 15 in just over two generations. A wide array of other development indicators improved dramatically as well. Life expectancy increased from 46 to 61 years in 1987 (before plunging again because of the HIV/AIDS pandemic), infant mortality fell from 118 to 43 per thousand by 2009, and literacy rates jumped from 46 percent (in 1970) to 83 percent by 2008.

Botswana did not seem to have strong prospects when it achieved independence from Great Britain in 1965. At that point, there were only 12 kilometers of paved roads in the entire country. Only 22 Batswana had graduated from university, and only 100 had graduated from secondary school. The country is landlocked, and more than 80 percent of the country is in the Kalahari Desert, leaving only a small amount of arable land. Yet, despite the long odds, Botswana prospered. What was behind this remarkable transformation?

Diamonds are part of the story, because Botswana sits atop some of the world's richest diamond deposits, and mining now accounts for about 40 percent of the country's output. But the answer is not that simple: Many other developing countries have rich natural resources, and in many cases, these have created more problems than benefits (see Chapter 17). More broadly, most observers point to Botswana's strong policies and institutions as key determinants of its development success.

Botswana clearly managed its resources much more prudently than other countries. Most of the receipts were invested productively: Botswana has built an impressive infrastructure, with paved roads connecting much of the country,

[a]This account draws heavily from Daron Acemoglu, Simon Johnson, and James Robinson, "An African Success Story," in Dani Rodrik, ed., *In Search of Prosperity: Analytic Narratives on Economic Growth* (Princeton, NJ: Princeton University Press, 2003); and Clark Leith, "Why Botswana Prospered," paper presented at the 34th Annual Meeting of the Canadian Economics Association, University of British Columbia, Vancouver, Canada, June 2000, available at www.ssc.uwo.ca/economics/faculty/Leith/Botswana.pdf, accessed February 2012.

a reasonably reliable electricity generation and distribution system, a substantial stock of housing, and many schools and clinics. Some of the diamond receipts were saved as reserves to help manage macroeconomic fluctuations. Corruption has been much less of a problem than in other countries. Overall macroeconomic policies were strong; inflation was relatively low, for the most part; and supporting fiscal, monetary, and exchange-rate policies were in place. Trade policies were relatively open, with the external tariff set through Botswana's membership in the Southern Africa Customs Union. The public sector has remained small, with a civil service based on merit rather than patronage and relatively few state-owned companies. Property rights and other legal protections have been generally well respected. Strong economic management, in turn, may have been due partly to the fact that Botswana is a democracy, and one of the few countries in Africa that has been so since independence.

However, not everything in Botswana is positive. Income inequality is high and has not declined over time. Unemployment remains high, particularly for migrants coming from rural to urban areas. Although democratic traditions are solid, with fair elections and a vibrant free press, one party has dominated politics since independence. The biggest concern, however, is HIV/AIDS. According to the World Health Organization, Botswana has the second-highest prevalence of HIV in the world; around 24 percent of the adult population is HIV-positive. The scourge of HIV/AIDS threatens to turn back much of the progress Botswana has realized in recent decades. With all that Botswana has achieved, its greatest challenges may lie ahead in fighting the disease and ensuring continued growth and development.

incomes double in about 70 years. With 2 percent growth, average income increases by 270 percent over 50 years, and it takes only 35 years for income to double. (Box 3–2 demonstrates the methods behind these calculations.)

FACTOR ACCUMULATION, PRODUCTIVITY, AND ECONOMIC GROWTH

Economists have been trying to understand the determinants of economic growth and the characteristics that distinguish fast-growing from slower-growing countries at least since Adam Smith wrote *An Inquiry into the Nature and Causes of the Wealth of Nations,* published in 1776. More than 200 years later, our knowledge about the growth process has expanded but is far from complete. A broad range of factors could

BOX 3-2 CALCULATING FUTURE VALUES, GROWTH RATES, AND DOUBLING TIMES

The illustrative growth calculations presented in the text are based on simple, yet quite useful mathematical techniques. Let us take a look at how to perform the calculations required to answer three basic questions.

QUESTION 1

If a country's current level of income per capita is X_0 and that country's income per capita grows for t years at rate r, what will its income per capita be at the end of that period (X_t)?

To answer this question, we will use the basic formula for compound growth over discrete periods (such as years),

$$X_t = X_0 \times (1+ r)^t$$

For example, we see from Table 3–1 that Senegal's income per capita in 2009 was $1,492, and its historical growth rate was 1.05 percent per year (which we will round down to 1 percent per year to simplify this calculation). Suppose we want to project Senegal's per capita income 10 years into the future (in 2019), assuming that its income continues to grow at 1 percent per year. Plugging in $1,492 for X_0, 0.01 for r, and 10 for t:

$$\$1{,}492 \times 1.01^{10} = \$1{,}648$$

The key here is to recognize that this future value is not simply 1 percent times 10 years (or 10 percent) greater than the initial value. Compound growth calculations take into account that after one year of 1 percent growth (starting at $1,492), Senegal's income is $1,507; the second year's growth of 1 percent thus starts from this new base. That is, for the first year, we'd calculate $\$1{,}492 \times 1.01 = \$1{,}507$; for the second year we'd calculate $(\$1{,}492 \times 1.01) \times 1.01$, or $\$1{,}492 \times 1.01^2 = \$1{,}522$, and so on.

QUESTION 2

If, instead, we know a country's initial level of per capita income (X_0) and its level of per capita income t years hence (X_t), what was its annually compounded average rate of growth?

To solve for the growth rate we begin with the basic compound growth formula from Question 1. With a bit of algebraic manipulation, we can solve that equation for r, resulting in the following formula:

$$\left(\frac{X_t}{X_0}\right)^{\frac{1}{t}} - 1 = r$$

If we knew that Senegal's per capita income in 2009 was $1,492 and that it would be $1,648 10 years later, we could calculate its average compound growth rate as:

$$\left(\frac{\$1{,}648}{\$1{,}492}\right)^{\frac{1}{10}} - 1 = 0.01 \text{ or } 1 \text{ percent}$$

Calculating growth rates using these endpoint data is potentially misleading if either the start or end year observation is unusually high or low relative to the underlying trend (that is, if there was a shock of some sort in either the start or end year, making it unrepresentative of the trend). It is possible to avoid this problem by using regression analysis to estimate the growth rate. This approach uses all of the observations in a given time series, rather than relying entirely on the first and last observations.

To estimate average growth rates by least-squares regression, we begin by transforming the basic formula for discrete periods of compound growth into a linear function by taking the natural logarithm of both sides of the formula:

$$\ln X_t = \ln X_0 + \ln(1+r) \times t$$

We can estimate this equation as the least-squares regression $\ln X_t = a + bt$, where $a = \ln X_0$, $b = \ln(1 + r)$, and t is time. That is, we regress the natural log of the series against a linear time trend. To recover the direct estimate of the compound growth rate t, we calculate $r = e^b - 1$.

QUESTION 3

If income (or, for that matter, population) were to grow at rate r, how long would it take to double?

To calculate doubling times (commonly used for illustrative purposes), we resort to the rule of 70. The rule of 70 is based on a slightly different concept of compounding from the one used for the other questions. In the previous examples, we compounded once each year. The rule of 70 is based on continuous compounding, in which the base is increased at each instant. The basic equation for continuous compounding uses the exponential function (e),

$$X_t = X_0 \times e^{rt}$$

In general, continuous compounding using this formula is more appropriate when calculating growth rates for populations, which growth continuously. (When the available data for a series consist of annual observations, such as is common for countries' gross domestic product [GDP], then it may be more appropriate to apply annual compounding for those discrete periods, as described earlier.) We

can use this formula to calculate the time it takes for income (or population) to double by setting $X_t = 2$ and $X_0 = 1$,

$$2 = 1 \times e^{rt}$$

If we transform this equation by taking natural logarithms of both sides, we have

$$\ln 2 = \ln 1 + (r \times t)$$

Conveniently, $\ln 1 = 0$ so that term drops out. This calculation is called the *Rule of 70* because $\ln 2 \approx 0.70$. So if we know r, we can solve for the doubling time:

$$\text{doubling time } (t) = \frac{0.70}{r}$$

If Senegal's economy grows at 1 percent per year, its level doubles every $(0.70/0.01) = 70$ years. If, instead, Senegal's economy grows at 2 percent per year, it would double in size every $(0.70/0.02) = 35$ years.

plausibly be important to growth, including the amount and type of investment, education and healthcare systems, natural resources and geographical endowments, the quality of government institutions, and the choice of public policy. All these play some role, as we see later in the chapter, but some are more central to the process of growth than others.

At the core of most theories of economic growth is a relationship between the basic factors of production—capital and labor—and total economic production. Some countries also are endowed with specific natural resources assets, such as petroleum deposits, gold, rubber, land, rich agricultural soil, forests, lakes, and oceans. These assets are often included as parts of a broad definition of the capital stock, but sometimes are treated separately. For simplicity, we focus our analysis on capital and labor. Depending on the products being produced, different combinations of these inputs are required. Growing rice requires significant amounts of labor (at least around planting and harvesting time), but not much machinery other than a plow. Garment production also requires lots of unskilled labor, but needs many sewing machines as well as decent infrastructure to get goods to overseas markets. A steel or chemical factory requires substantial amounts of machinery and other capital, including a reliable source of energy, and (relatively) less labor.

A country's total output—and thus its total income—is determined by how much capital and labor it has available and how productively it uses those assets. In turn, *increasing* the amount of production—that is, economic growth—depends on

increasing the amount of capital and labor available and increasing the productivity of those assets. In other words, economic growth depends on two basic processes:

- **Factor accumulation**, defined as increasing the size of the capital stock or the labor force. Producing more goods and services requires more machines, factories, buildings, roads, ports, electricity generators, computers, and tools along with more and better educated workers to put this capital equipment to work.
- **Productivity growth**, defined as increasing the amount of output produced by each machine or worker. Productivity can be increased in two broad ways. The first is to improve the **efficiency** with which current factors are being used. A small furniture maker might initially have each worker make a chair from start to finish. By reorganizing the workers so each specializes in one task (for example, cutting, assembling, finishing), total production might be increased. The second is **technological change**, through which new ideas, new machines, or new ways of organizing production can increase growth. Countries that can either invent new technologies or quickly adopt technologies invented elsewhere (a more relevant path for most developing countries) can achieve more rapid economic growth than other countries. Productivity growth often entails shifting resources from producing one good to another. The process of economic growth in low-income countries almost always corresponds to major structural shifts in the composition of output, generally from agriculture to industry, as we discuss in Chapter 15.

Understanding that factor accumulation and productivity gains are at the heart of the growth process is important but takes us only so far. To gain a deeper understanding of growth, we must understand what drives factor accumulation and productivity growth themselves. We begin with saving and investment.

SAVING, INVESTMENT, AND CAPITAL ACCUMULATION

Perhaps the most influential model of economic growth was developed by MIT economist Robert Solow in 1956.[1] At the heart of the **Solow growth model** (which we explore in depth in Chapter 4) and many other influential growth models is the process of capital accumulation. As mentioned previously, classical and neoclassical growth models typically give much less attention to the process of expanding the

[1]Robert Solow, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70 (February 1956), 65–94, and "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39 (August 1957), 312–20.

labor force (because labor is not seen as a binding constraint on growth), which usually is assumed to grow in line with the population. The key ideas in these kinds of models are relatively straightforward:

- *New investment increases the capital stock.* Investment in new factories or machines directly increases the capital stock, which facilitates greater production. For the capital stock to grow, the value of new investment must be greater than the amount of depreciation of existing capital. Factories and machines deteriorate over time, and a certain amount of new investment is needed simply to keep pace and maintain the current size of the capital stock. Investment greater than the amount of depreciation directly adds to the capital stock. Investment must be greater than depreciation and the growth of the labor force for there to be an increase in capital per worker.
- *Investment is financed by saving.* We know from standard national accounts identities that investment equals saving.[2] Thus the models postulate that the key to increasing investment (and the capital stock) is to increase saving.
- *Saving comes from current income.* Households save whatever income they do not consume. Corporations save in the form of retained earnings after distribution of dividends to stockholders. Governments either add to saving to the extent that they receive tax payments in excess of government current (noninvestment) spending (that is, a budget surplus, excluding investment spending) or detract from saving if they spend more than tax receipts (a budget deficit, excluding investment spending). We discuss these concepts in more depth in Chapters 10 and 11. Gross domestic saving, which combines all three sources of saving, provides the resources to finance investment.[3]

A key decision facing households, corporations, and governments is how much income to consume and how much to save. Individuals do not care about the level of capital or even the level of output, but they do care about the amount of goods and services they consume. However, there is a clear trade-off: The more that is consumed now, the less that is available for saving and therefore for investment, growth, and increased future consumption. On the one hand, people prefer to consume sooner rather than later. Given a choice between better housing now or in five years, everyone would choose now. On the other hand, people also recognize that consuming all income now is shortsighted. At a minimum, enough needs to be saved to compensate for depreciation of existing assets: to fix the roof on the house when it leaks, repair

[2]On the production side of the national accounts (for a closed economy with no trade), everything that is produced (total output, usually designated as Y) must be used for either consumption (C) or investment (I). On the income side, total income (also designated as Y because the value of total output must equal total income) is either used to purchase consumption goods (C) or saved (S). Putting these two identities together (because Y and C appear in both), saving must be equal to investment.

[3]In an open economy with trade and international capital flows, foreign saving (for example, borrowing from an overseas bank) can add to the total pool of available saving.

the motorbike, or replace a worn-out hoe. Additional saving can provide the basis for even higher income in the future. Deferring current consumption can lead to greater consumption later. For example, a farmer who wants to buy a water buffalo may have to reduce his or her consumption for several years to save enough to eventually make the purchase. The farmer's reward comes later, when (ideally) the water buffalo increases farm production and income by far more than the amount saved, allowing for even larger consumption in the future.

As the last sentence hints, although generating saving and investment is necessary for growth, it is not sufficient. The investments actually have to pay off with higher income in the future, and not all investments do so. The farmer who buys the water buffalo will not earn the reward if the field is too rocky to plow or the land does not have enough nutrients to support the crop. If the government forces down the price of the crop (for example, to try to keep food prices low for urban consumers), the farmer's income will fall and the investment will be much less profitable. If property rights are not secure, the farmer could lose the water buffalo or the land. Changes in world market prices (well out of the farmer's control) could also affect income. These issues highlight a key point: *Sustaining economic growth requires both generating new investment and ensuring that the investment is productive.* This idea is a recurrent theme in the next two chapters and throughout this book.

SOURCES OF GROWTH ANALYSIS

So far, we have identified factor accumulation and productivity gains as the two core determinants of growth. But how important are each of these in explaining growth? Robert Solow pioneered early efforts to quantify the contribution of each of the proximate causes of increased output—capital accumulation, labor accumulation, and productivity gains—to economic growth. This approach is more of an accounting framework based on actual data than an economic model. It seeks to answer the following question: What proportions of recorded economic growth can we attribute to growth in the capital stock, growth of the labor force, and changes in overall productivity?[4]

One way to explore how factor accumulation and productivity growth affect output and economic growth is by examining a **production function**, which characterizes how inputs (capital and labor) are combined to produce various levels of output. Figure 3–1 shows an example of a common production function. The horizontal axis shows one measure of factor inputs (the amount of capital per worker),

[4]Solow, "Technical Change and the Aggregate Production Function." A year before Solow's paper, Moses Abramovitz found similar estimates of the contribution of productivity gains to U.S. growth between 1870 and 1953 using a less formal methodology. See his "Resource and Output Trends in the United States since 1870," *American Economic Review* 46, no. 2 (May 1956), 5–23.

and the vertical axis shows the amount of output per worker (in this example, pairs of running shoes). Combining capital and labor into the single "capital per worker" term is a convenient way to simplify the analysis, but it also reflects the dominant role that capital has played in thinking on economic growth over the years. In most models, insufficient capital is seen as the binding constraint on growth. Labor is usually assumed to be in plentiful supply, based on the observation of unemployed or underemployed workers. Thus, in this formulation, increasing the amount of capital available for each worker is seen as key to growth.

Factor accumulation is shown in Figure 3–1a as a movement to the right along the *x*-axis. As the economy accumulates more capital for each of its workers, output expands, shown by the upward slope of the production function. Note that this particular production function begins to flatten out as capital per worker expands, a characteristic that we explore more fully later. In this case, as the value of capital

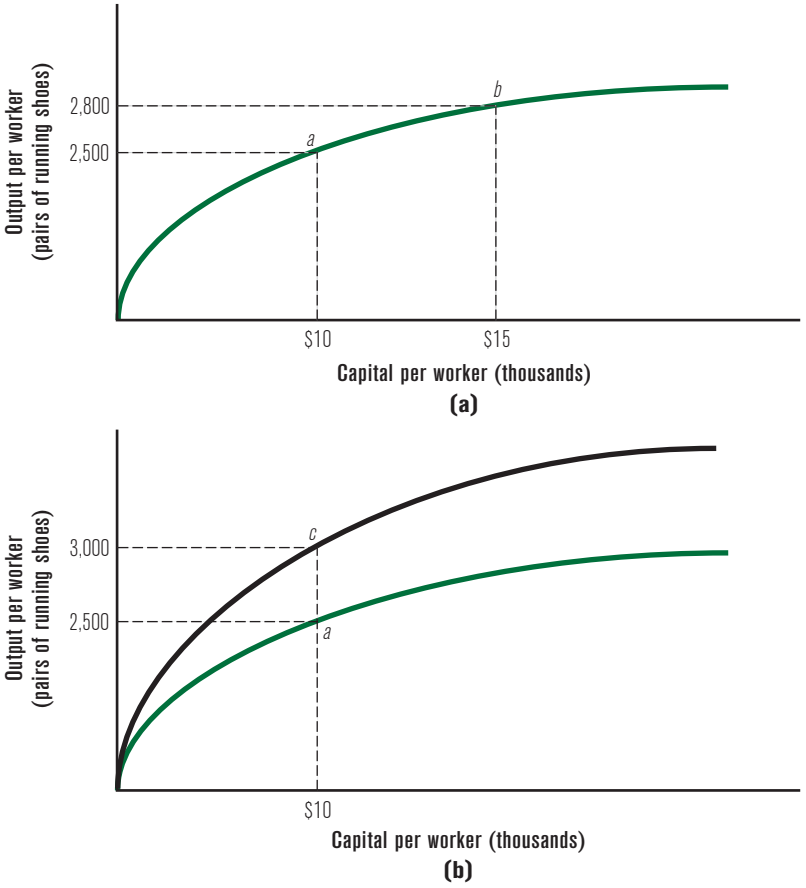


FIGURE 3–1 Basic Sources of Economic Growth

(a) Factor accumulation. As capital per worker expands, output per worker increases.
(b) Productivity gains. As the factors of production are used more efficiently or the economy acquires new technology, the same level of capital per worker can produce more output.

per worker increases from $10,000 to $15,000, output per worker grows from 2,500 to 2,800 pairs of shoes. The movement from point *a* to point *b* on the production function is the process of economic growth. Of course, growth of this magnitude (12 percent) does not happen instantaneously. It might take two years or so, in which case the annual rate of growth would be a brisk 5.8 percent.

Figure 3–1b shows the relationship between productivity change and economic growth. As factors of production are used more efficiently or as new technology is adopted, the production function shifts upward. With this productivity gain, any amount of capital per worker produces more output than it did before. The production function shifts upward so that $10,000 worth of capital per worker can now produce 3,000 pairs of running shoes, whereas it originally produced 2,500 pairs. This expansion of output of 20 percent might take four years, in which case the annual rate of growth would be 4.7 percent. In this case, the movement along the path from point *a* to point *c* depicts the process of economic growth through productivity gains.

Solow's procedure for measuring these relationships usually is referred to as **growth accounting** or **sources of growth analysis**. He starts with a standard production function relating the contribution of labor and capital to aggregate production, then adds a term to capture **total factor productivity (TFP)**. TFP is meant to measure the contribution to production of efficiency, technology, and other influences on productivity. This production function is then converted into a form that makes it possible to measure the contribution of changes in each term—expansion of the labor force, additions to the capital stock, and growth in TFP—to overall growth. The resulting equation is

$$g_Y = (W_K \times g_K) + (W_L \times g_L) + a \qquad [3\text{–}1]$$

where g_Y stands for the growth of total income, or gross domestic product (GDP); g_K and g_L are the growth rates of the capital stock (K) and the labor force (L), respectively; and W_L and W_K represent the shares in total income of wages and returns on capital, respectively. For example, if 60 percent of all income comes from wages and the remaining 40 percent comes from returns on capital (for example, interest payments and rent), then $W_L = 0.60$ and $W_K = 0.40$. These two shares must add up to 100 percent because all income must be allocated to either workers or the owners of capital. The final term, a, is the rate of change in TFP. The equation should make intuitive sense: It shows us how the growth in output depends on the growth in inputs (K and L) and the growth in the productivity of those inputs (a).

It is important to see the relationship between equation 3–1 and Figure 3–1. Equation 3–1 simply summarizes the two effects illustrated in Figure 3–1. Figure 3–1a shows the effect of factor accumulation on output, while Figure 3–1b illustrates productivity growth as a vertical shift of the production function (for example, getting more output per unit of input). Equation 3–1 combines these two dimensions of growth in a way that allows us to distinguish between them. The first two terms on the right-hand side of the equation, $(W_K \times g_K) + (W_L \times g_L)$, describe movements

along the production function (that is, using more inputs to increase output). TFP growth (indicated by *a* in equation 3–1) measures the upward shift in the production function in Figure 3–1b.

The basic procedure is to substitute actual data for all the variables in equation 3-1 except *a,* which cannot be measured directly, then calculate *a* as the residual. This is the famous **Solow residual**. In this way, the contribution of each of these variables to growth can be measured and identified. A straightforward numerical example illustrates the way in which this equation is used. From the statistical records of a developing country, we find the following values for the variables in the equation:

$g_Y = 0.05$ (GDP growth rate of 5 percent a year).
$g_K = 0.07$ (capital stock growth of 7 percent a year).
$g_L = 0.02$ (labor force growth of 2 percent a year).
$W_L = 0.6$ (the share of labor in national income is 60 percent).
$W_K = 0.4$ (the share of capital is 40 percent).

By substituting these figures into equation 3–1, we get

$$0.05 = (0.4 \times 0.07) + (0.6 \times 0.02) + a$$

Solving for *a,* we find that $a = 0.01$, meaning that TFP growth is 1 percent per year. These figures tell us the degree to which capital accumulation, labor accumulation, and TFP growth each contribute to the overall rate of output growth of 5 percent. TFP growth of 1 percent counts for one fifth (20 percent) of total growth. The growth in the capital stock accounts for slightly more than half (56 percent) the total growth: $(0.4 \times 0.07)/0.05$. Finally, growth in the labor force accounts for the remaining 24 percent of total growth: $(0.02 \times 0.6)/0.05$. In this particular example, capital accumulation is the main driver of growth, with labor accumulation and TFP growth each contributing similar amounts.

This type of accounting analysis has been used widely in many countries to examine the sources of growth, with particular attention paid to calculating TFP growth. Before examining some of these results, however, it is important to recognize the limits of this kind of study, arising from the fact that we can estimate productivity growth only as the residual output growth unexplained by having used more inputs. There are at least two kinds of problems:

- First, *a* represents a combination of influences that this analysis cannot entirely disentangle. Should improvements in *a* be attributed to efficiency gains stemming from improved trade policies, reduced corruption, or streamlined bureaucratic procedures? Or are they due to the introduction of faster computers, new seed varieties for agricultural crops, or other technologies? The limited growth accounting framework cannot definitively answer these questions without adding many more variables for which data do not exist (although that has not stopped analysts from assigning their own favorite explanation to the results).

- Second, *a* invariably is measured inaccurately because it is the residual in the equation. All economic data are measured with some inevitable errors, including all the data used in equation 3–1. As a result, in addition to TFP, *a* captures the net effect of all the errors and omissions in the other data.

What is labeled TFP actually, in practice, is a combination of errors in the data, omission of other factors that should be included in the growth equation, and efficiency gains and changes in technology. As a result, there is a danger in trying to read too much into these data when analysts interpret them as strictly efficiency gains or the effects of new technology. Rather than truly being TFP growth, *a* simply is the part of measured growth that cannot be explained by data on the traditional factors of production. For this reason, economist Moses Abramovitz famously referred to the residual *a* as a "measure of our ignorance" about the growth process.[5]

Sources of growth analyses have been carried out for many countries. Solow's initial study on the United States attributed a surprisingly large share of growth to the residual and a correspondingly small share to changes in the capital stock: 88 percent to TFP growth and only 12 percent to increases in capital per worker. Subsequent work by Abramovitz, Edward Denison, Zvi Griliches, Dale Jorgenson, and others attempted to measure, in a more precise way, the contribution of various inputs to the growth process. They divided labor into different skill categories, based on the amounts of formal education workers had received. A worker having a high school education and earning $50,000 a year is treated as the equivalent of two people having only primary school education and earning $25,000 a year each. Similar procedures are used to measure the increase in productivity that occurs when workers shift from low-productivity occupations in rural areas to higher-productivity occupations in urban areas. Other methods are used to measure improvements in the quality of capital and increasing economies of scale.

Many of these more detailed analyses of the U.S. economy found results similar to Solow's initial work: The bulk of the growth process could be attributed to the residual, with relatively small amounts apportioned to various categories of labor, capital, and other inputs. Over the years, many more studies have been completed for the industrialized countries. Increases in the capital stock frequently account for less than half the increase in output, particularly in rapidly growing countries. These results came as a bit of a surprise for most economists because most basic models put capital formation at the heart of the growth process.

Similar studies have now been carried out for a wide range of developing countries. Data problems and price distortions tend to be more severe for developing countries than for the industrialized countries, making the results even harder to interpret. Few developing countries, for example, have reliable measures for differences in the quality of alternative capital input and labor skill categories.

[5]Abramovitz, "Resource and Output Trends in the United States since 1870."

Generally speaking, sources of growth analyses in developing countries attribute a larger role to capital formation than in the industrialized country studies. This is consistent with the idea that developing countries have lower levels of capital per worker than the industrialized countries and can catch up (or converge incomes) through the investment process. Much of the capital equipment imported by developing countries (counted as investment) embodies advances in technology. Therefore, the mobilization of capital remains a major concern of policy makers in developing countries.

Economists Barry Bosworth and Susan Collins explored the relative contributions of physical capital, human capital (in the form of education), and TFP to economic growth in a large number of countries around the world between 1970 and 2000. Some of their results are shown in Table 3–2.[6] As with other studies, they found a fairly consistent pattern that capital accumulation was the main contributor to growth for developing countries, whereas for industrialized countries, the main contributions were more evenly split between capital accumulation and TFP growth. In East Asia, capital accumulation accounted for about two-thirds of total growth, with TFP growth accounting for a smaller share. In comparing TFP growth across countries, the rapidly growing East Asian economies generally (but not always) recorded faster TFP growth than did developing countries from other regions of the world. TFP growth in East Asia was generally faster than that of the industrialized countries during the 1970s and 1980s and about the same during the 1990s.

Average TFP growth was negative in all three decades in Africa, during the 1980s in Latin America, during the 1970s in South Asia, and during the 1970s and 1990s in the Middle East! What does this mean? Inputs actually became less productive over time. This might be the result of capital and labor lying idle, as often happens during wars, political unrest, or recessions. During Latin America's deep recession, induced by the debt crisis of the 1980s, growth was negative while new investment was essentially zero, implying less productive use of existing capital. Negative TFP growth also might reflect the accumulation of increasingly unproductive assets, like presidential palaces or so-called white elephant projects. For example, Ethiopia built one of the largest tanneries in the world, but it usually operates at a fraction of its capacity, and Nigeria's Ajaokuta Steel factory cost nearly $5 billion in construction costs over 25 years and has yet to produce any steel.

In a more recent study, Bosworth and Collins provide a detailed comparative analysis of the sources of growth in the world's two most populous countries—India and China.[7] Since 1980, GDP per capita has more than doubled in India and increased more than sevenfold in China. While China's growth has been driven by its

[6]Susan M. Collins and Barry Bosworth, "The Empirics of Growth: An Update," *Brookings Papers on Economic Activity* 2 (2003), 113–79.

[7]Barry Bosworth and Susan M. Collins, "Accounting for Growth: Comparing India and China," *Journal of Economic Perspectives*, 22, no. 1 (Winter 2008), 45–66.

TABLE 3–2 Sources of Growth in East Asia and Other Regions, 1960–2000 (Average Annual Growth Rate, Percent)

	CONTRIBUTION BY COMPONENT			
	GROWTH OF OUTPUT PER WORKER	PHYSICAL CAPITAL PER WORKER	EDUCATION PER WORKER	TOTAL FACTOR PRODUCTIVITY
Brazil				
1970s	4.86	2.02	0.12	2.72
1980s	−1.63	0.16	0.68	−2.47
1990s	0.71	0.07	0.38	0.25
Ecuador				
1970s	5.96	1.05	0.89	4.03
1980s	−1.42	−0.28	0.16	−1.30
1990s	−1.40	−0.46	0.31	−1.24
Egypt				
1970s	4.39	2.33	0.54	1.52
1980s	2.91	1.98	0.89	0.03
1990s	1.46	−0.12	0.64	0.94
Ethiopia				
1970s	0.55	0.22	0.13	0.20
1980s	−1.74	1.11	0.27	−3.12
1990s	1.84	0.81	0.29	0.74
Ghana				
1970s	−2.01	−0.24	0.24	−2.00
1980s	−1.14	−1.23	0.15	−0.07
1990s	1.62	0.80	0.16	0.65
India				
1970s	0.70	0.61	0.36	−0.27
1980s	3.91	1.06	0.36	2.48
1990s	3.13	1.35	0.49	1.29
Singapore				
1970s	4.41	3.53	0.11	0.78
1980s	3.79	2.01	0.39	1.38
1990s	5.08	1.96	0.91	2.22
Taiwan				
1970s	5.93	3.69	1.11	1.14
1980s	5.36	2.19	0.24	2.94
1990s	4.84	2.66	0.41	1.77
United States				
1970s	0.83	0.11	0.71	0.01
1980s	1.82	0.55	0.12	1.15
1990s	1.84	0.74	0.11	0.98
Africa				
1970s	1.03	1.28	0.08	−0.32
1980s	−1.06	−0.07	0.42	−1.41
1990s	−0.16	−0.09	0.40	−0.48
East Asia				
1970s	4.27	2.74	0.67	0.86
1980s	4.36	2.45	0.66	1.25
1990s	3.36	2.35	0.50	0.52

Industrial countries				
1970s	1.75	0.95	0.52	0.28
1980s	1.82	0.69	0.24	0.90
1990s	1.52	0.75	0.22	0.54
Latin America				
1970s	2.69	1.25	0.34	1.10
1980s	−1.77	0.04	0.47	−2.28
1990s	0.91	0.16	0.34	0.41
Middle East				
1970s	1.92	2.08	0.45	−0.61
1980s	1.15	0.55	0.53	0.07
1990s	0.84	0.34	0.52	−0.01
South Asia				
1970s	0.68	0.56	0.34	−0.23
1980s	3.67	1.02	0.40	2.25
1990s	2.78	1.19	0.42	1.17

Source: Susan M. Collins and Barry Bosworth, "The Empirics of Growth: An Update," *Brookings Papers on Economic Activity* 2 (2003), 113–79.

industrial sector, India's less traditional growth pattern has been driven by its service sector. Comparing these two economies in the aggregate for the years 1978–2004, Bosworth and Collins estimate that TFP growth accounted for just under half of the growth in output per worker in both countries. Yet, the absolute rate of TFP growth in China over that period was 3.6 percent per year compared with 1.6 percent per year in India. The contribution of TFP to growth in output per worker in India and China, they find, differs sharply from the broader pattern in East Asia (excluding China), where average TFP growth was slower (0.9 percent per year) and accounted for less than 25 percent of growth in output per worker.

The aggregate figures for the contribution of TFP to output growth in Sub-Saharan Africa presented in Table 3–2 are bleak. Yet, a more recent and more detailed analysis of that region by economist Steven Radelet demonstrates that for the 17 countries of the region that succeeded in stabilizing their economic and political circumstances by the mid-1990s, average TFP growth rates soared from negative and unstable rates to positive TFP growth since 1995 on the order of 1.5 percent per year.[8] In summary, sources of growth analyses suggest that capital accumulation is the main source of growth for developing countries, consistent with the Solow growth model. TFP can play an important part in the growth process in the appropriate

[8]Steven Radelet, *Emerging Africa: How 17 Countries Are Leading the Way*. Washington, DC: Center for Global Development, 2010. Radelet's 17 emerging African countries are Botswana, Burkina Faso, Cape Verde, Ethiopia, Ghana, Lesotho, Mali, Mauritius, Mozambique, Namibia, Rwanda, São Tomé and Príncipe, Seychelles, South Africa, Tanzania, Uganda, and Zambia.

policy and structural context. In rapidly growing economies, both factor accumulation and TFP growth appear to play an important role. TFP growth tends to become more important as income rises and is a major contributor to growth in the high-income industrialized countries.

CHARACTERISTICS OF RAPIDLY GROWING COUNTRIES

We identified the key proximate causes of economic growth: factor accumulation (accumulating additional productive assets) and productivity growth. Productivity growth, in turn, comes either from efficiency gains or new technology. Sustaining economic growth requires both generating new investment and ensuring that the new investment is productive. These basic points, however, raise a new set of questions. What are the more fundamental characteristics that explain a country's ability to attract investment and accumulate capital, increase efficiency, and obtain new technologies? More broadly, what are the deep characteristics that distinguish more rapidly growing economies from slowly growing ones?

A large body of research over the last few decades tried to answer this question by searching for broad characteristics common to rapidly growing economies. Until relatively recently, it was difficult for researchers to systematically examine these issues due to severe data limitations. Many researchers examined trends in individual countries, but it was difficult to draw general conclusions from these case studies. A few pioneering efforts, such as Irma Adelman and Cynthia Taft Morris's *Society, Politics, and Economic Development—A Quantitative Approach*,[9] paved the way for today's cross-country empirical growth research. In recent years, this type of research has grown very rapidly, in line with the emergence of many new and large data sets on income in PPP terms, education levels, health characteristics, quality of governance, and a host of related items.[10]

Most of the recent studies are modeled on research conducted by economist Robert Barro in the early 1990s. These studies try to explain the variance in growth rates across countries. With country growth rates as the dependent variable, this approach examines several variables that might affect growth through one of the channels identified earlier (controlling for the initial level of income in each country). These variables include levels of education and health, policy choices, resource endowments, geographic characteristics (latitude, whether the country is landlocked, etc.), and political systems.

[9]Irma Adelman and Cynthia Taft Morris, *Society, Politics, and Economic Development—A Quantitative Approach* (Baltimore, MD: Johns Hopkins University Press, 1967).

[10]The vast literature on cross-country growth empirics began with Robert Barro, "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics* 106, no. 2 (May 1991), 407–43. For the most comprehensive recent survey of this type of research, see S. Durlauf, P. Johnson, and J. Temple, "Growth Econometrics," in P. Aghion and S. Durlauf eds., *Handbook of Economic Growth* (Amsterdam: North Holland, 2005).

This type of research has been controversial, and there is far from a consensus on the exact group of variables that affects growth.[11] For one thing, although this research starts with the Solow model, for many of the variables tested, there is no well-developed theoretical link between the variable and either economic growth or the proximate causes of growth (factor accumulation or productivity growth). The existing theories on economic growth simply are not explicit enough about exactly what variables determine the shape of the production function, the rate of investment, the profitability of investment, efficiency, and the rate of change in technology. Some characteristics may appear statistically important in one research study that includes a certain group of variables but unimportant in another study with a different group of variables.[12]

A second issue has to do with interpretation of the statistical results. For example, most economists would agree that high saving rates are associated with rapid economic growth. But which causes which? Higher saving can lead to more rapid growth, as suggested by the Solow model, while faster economic growth might provide more disposable income and a higher saving rate (we discuss this issue more in Chapter 10). It is a major statistical challenge to precisely estimate the magnitude of these two effects.

Despite these issues, such empirical growth research has helped analysts better understand some of the broad characteristics associated with rapid growth, albeit very imperfectly. The broad thrust of the conclusions from research across countries is consistent with many studies of individual countries. Although the debate is far from over about which variables influence long-run growth, how they do so, and the magnitude of the effect, this research has helped provide broad clues about why some economies grow faster than others. The most rapidly growing developing countries tend to share six broad characteristics.[13]

1. MACROECONOMIC AND POLITICAL STABILITY

Stability is good for growth. Economic and political instability undermine investment and growth and are especially hard on the poor, who are least able to protect themselves against volatility. Consider the Democratic Republic of the Congo (formerly Zaire), which suffered through inflation rates averaging an astonishing 2,800 percent

[11]For critiques of the cross-country approach, see Durlauf et al., "Growth Econometrics;" and David Lindauer and Lant Pritchett, "What's the Big Idea? The Third Generation of Policies for Economic Growth," *Economia* 3, no. 1 (Fall 2002).

[12]A seminal study on the robustness of explanatory variables across different specifications is Ross Levine and David Renelt, "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review* 82, no. 4 (September 1992), 942–63.

[13]Economist Xavier Sala-i-Martin has examined a wide range of studies and identified a list of variables that are most consistently and robustly found to be closely associated with economic growth. See Xavier X. Sala-i-Martin, "I Just Ran Two Million Regressions," *American Economic Review* 87, no. 2 (May 1997), 178–83. The specific list of key areas used here is similar to that suggested by Lawrence Summers and Vinod Thomas in "Recent Lessons of Development," *World Bank Research Observer* 8, no. 2 (July 1993), 241–54.

per year between 1990 and 2002 and civil and cross-border wars involving troops from at least five other countries. It is not surprising that its growth and development performance was about the worst in the world: "growth" of −7.2 percent per year (meaning that average incomes fell by 60 percent over 12 years), life expectancy fell from 52 to 45 years, and infant mortality rose from 128 to 139 per thousand.

Relatively low budget deficits over time (with corresponding high rates of government saving), prudent monetary policy (which keeps inflation in check), appropriate exchange rates, suitable financial markets (depending on the stage of development), and judicious foreign borrowing at sustainable levels are the key elements to macroeconomic stability. Stability reduces risk for investors, whether they are multinational conglomerates or coffee farmers considering planting more trees. A high rate of inflation, for example, makes prices and profits much less predictable, undermining growth.[14] Volatile short-term capital flows can lead to wide swings in the exchange rate, affecting prices throughout the economy and undermining investment. In extreme situations, volatile capital flows can lead to full-blown financial crises, as we explore more deeply in Chapter 13.

Political stability is, of course, also good for growth and development. Civil and cross-border wars, military coups, and other incidences of political instability undermine investment and growth. Once again, the poor are the most vulnerable and least able to protect themselves from the consequences of political unrest. In the late 1990s, nearly one-third of the 42 countries in sub-Saharan Africa were embroiled in cross-border or civil wars, which took a huge toll on human lives, infrastructure, institutions, and economic activity and commerce. Figure 3–2 shows sharp declines in income, averaging 28 percent, after civil war in seven developing countries. By contrast, most of the relatively successful developing countries during the last several decades were politically stable for long periods of time. Although some of the successful countries experienced periods of instability, for the most part they were short-lived. Economist Paul Collier and others have pointed out the insidious negative cycle of civil war in low-income countries: poverty increases the risk of conflict, and conflict undermines growth and entrenches poverty.[15] Of course, the absence of war is no guarantee of economic growth. Cuba, Jamaica, and Kenya were politically stable for decades but still experienced low growth.

[14]The negative association between inflation and growth has been demonstrated in numerous studies, prominent among them Robert Barro, "Inflation and Growth," *Review* [Federal Reserve Bank of St. Louis] 78, no. 3, 153–69 (May–June 1996).

[15]See Paul Collier, "On the Economic Consequences of Civil War," *Oxford Economic Papers* 51 (1999), 168–83; and Paul Collier, V. L. Elliot, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis, *Breaking the Conflict Trap: Civil War and Development Policy* (Washington, DC: World Bank and Oxford University Press, 2003). Collier summarizes this and related work in *The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It,* (Oxford: Oxford University Press, 2007).

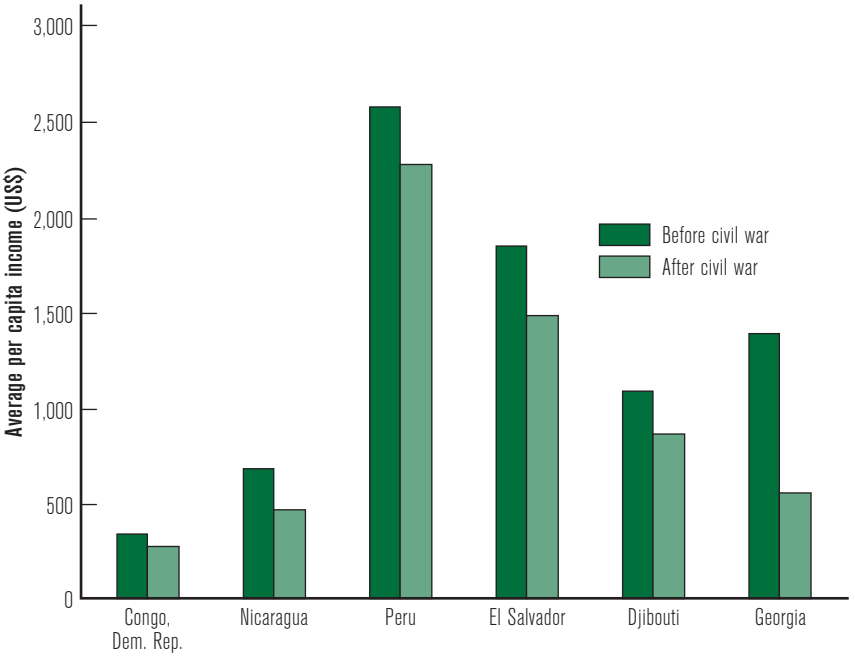


FIGURE 3–2 GDP per Capita before and after a Civil War

In these six countries, average income was 28 percent lower after the civil war than before.

Sources: Civil war data from Paul Collier, V. L. Elliot, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis, *Breaking the Conflict Trap: Civil War and Development Policy* (Washington, DC: World Bank and Oxford University Press, 2003). Gross domestic product per capita data from *World Development Indicators 2004*, (Washington, DC: World Bank), http://data.worldbank.org/data-catalog/world-development-indicators.

2. INVESTMENT IN HEALTH AND EDUCATION

Countries with *longer life expectancy* (and therefore better health) tend to grow faster, after accounting for other factors affecting growth, as shown in Figure 3–3.[16] A longer life expectancy indicates general improvements in the health of a population, which in turn means a healthier and more productive labor force. Thus, one way that life expectancy affects growth is by influencing productivity. In addition, a higher life expectancy might also boost saving and capital accumulation because businesses may be more likely to invest where workers are healthier and more productive. Moreover, people are more likely to invest in education to deepen their skills if they expect to live longer and reap greater benefits. Accessible basic healthcare facilities, clean water and sanitation, disease control programs, and strong reproductive and

[16]See, for example, World Health Organization, *Macroeconomics and Health: Investing in Health for Economic Development, Report of the Commission on Macroeconomics and Health* (Geneva: World Health Organization, 2001). For a brief nontechnical summary, see David E. Bloom, David Canning, and Dean Jamison, "Health, Wealth, and Welfare," *Finance and Development* 41, no. 1 (March 2004), 10–15.

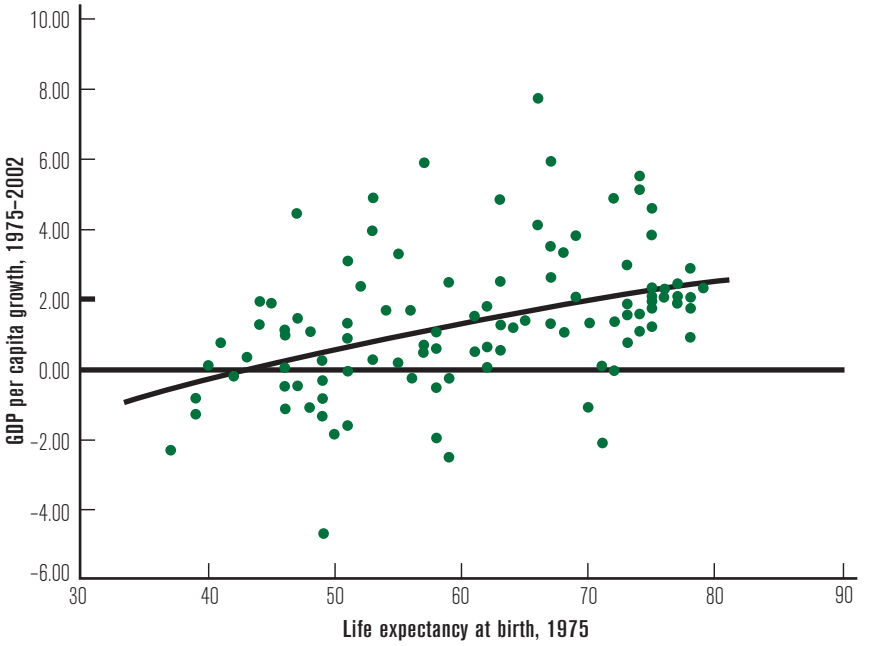


FIGURE 3–3 **Growth and Life Expectancy**

maternal and child health programs help countries lengthen life expectancy and improve worker productivity.

Malaysia's efforts to reduce malaria and improve health are a good example. When settlers first began to arrive in what is now Kuala Lumpur in the 1850s in search of tin, nearly half died of malaria. A century later, things had improved, but malaria and other diseases were still a problem, and life expectancy in 1960 was just 54 years. Growth during the 1960s was still a respectable 3.4 percent per year. An active government malaria control program began to make significant inroads, and by 1975, the number of malaria cases had been cut by two-thirds relative to 1960. Partly as a result, life expectancy increased to 64 years. Improved health (along with several other factors) contributed to a surge in growth, which accelerated to about 5 percent per year between 1976 and 1996, slowing only in the wake of the 1998 East Asian financial crisis. The incidence of malaria has continued to decline, falling to just above 7,000 confirmed cases in 2009.[17]

Note that the relationship between life expectancy and growth works both ways: Better health helps spur faster growth, and faster growth (and higher income) helps improve life expectancy, as we discuss in more detail in Chapter 9. That is, better

[17]The data on the incidence of malaria come from the World Health Organization, World Malaria Report 2011," available at www.who.int/malaria/world_malaria_report_2011/en, accessed February 2012. The country profile for Malaysia is available at www.who.int/entity/malaria/publications/country-profiles/profile_mys_en.pdf, accessed February 2012.

health is both an *input* to and *outcome* of the growth process. In Chapter 2, we discussed increased life expectancy as an important development goal as part of the human development index (HDI). Here we emphasize the role of good health as an input to the growth process. These two positions are not contradictory—far from it. Rather, they imply a positive reinforcing cycle in which better health supports faster economic growth, and the higher income from growth facilitates even better health.

Similarly, both *increased levels and improved quality of education* should translate into a more highly skilled workforce and increased productivity. A skilled workforce should be able to work more quickly with fewer errors, use existing machinery more effectively, and invent or adapt new technologies more easily. As with better health, a more highly educated workforce may also help attract new investment, thereby contributing to capital accumulation as well. Also, like better health, education has a pro-cyclical relationship with growth, in which better education helps support growth and growth generates the resources to finance stronger educational systems. Education of girls has a particularly strong effect on growth, both the direct impact on their skills and the indirect impact in the next generation on their children's health and education. The impact of education on growth can take a long time because investments in primary school education today may not show up as improved worker productivity for many years. The quality of service delivery is just as important as quantity. It is not enough to build schools and increase enrollment rates; teachers have to show up, be motivated, and have adequate basic supplies (such as textbooks) to do their job.[18]

As discussed in Chapter 8, most micro-level studies in individual developing countries show very high rates of return to education, especially girl's education. However, in macro-level cross-country studies, the statistical strength of the relationship often is relatively modest. This may be due to difficulties in accurately measuring the quantity and quality of education across a large number of countries in a consistent way. It also suggests that a better-educated workforce is no guarantee of more rapid economic growth. Human capital, just like physical capital, can be squandered in an environment that is not otherwise supportive of economic growth.

3. EFFECTIVE GOVERNANCE AND INSTITUTIONS

The role of governance and institutions in economic growth in development began to receive serious attention only beginning in the early 1990s (Box 3–3). This work was heavily influenced by the research and writing of Nobel Prize–winning economist Douglass C. North of Washington University.[19] Since that time, many studies have found a positive relationship between economic growth and the strength of the rule

[18]These issues are explored more deeply in Chapter 8 and in the World Bank's world development report, *Making Service Work for Poor People* (Washington, DC: World Bank and Oxford University Press, 2004).

[19]See, for example, Douglass C. North, *Institutions, Institutional Change, and Economic Performance* (New York: Cambridge University Press, 1990).

BOX 3-3 INSTITUTIONS, GOVERNANCE, AND GROWTH[a]

The role of institutions and governance in supporting and sustaining economic growth began to receive strong attention in the 1990s, following the path-breaking work by Nobel Prize–winning economist Douglass C. North.[b] In its broadest definition, institutions include a society's formal rules (for example, constitutions, laws, and regulations), informal constraints (such as conventions, norms, traditions, and self-enforced codes of conduct), and the organizations that operate within these rules and constraints.[c] There are many different kinds of institutions that influence growth and development. Dani Rodrik and Arvind Subramanian suggest four broad types of economic institutions, to which we add a fifth for political institutions:

- *Market-creating institutions* protect property rights, ensure that contracts are enforced, minimize corruption, and generally support the rule of law. Without these institutions in place, markets are likely not to exist or to perform poorly; by contrast, strengthening them can help boost investment and entrepreneurship. Examples include an independent judiciary, an effective police force, and enforceable contracts.
- *Market-regulating institutions* deal with market failures, such as imperfect information and economies of scale. These institutions limit monopoly power and help provide the basis for building and managing public goods, such as roads and fisheries. Examples include regulatory agencies in telecommunications, transportation, water and forestry resources, and financial services.
- *Market-stabilizing institutions* ensure low inflation, minimize macroeconomic volatility, ensure fiscal stability, and avert financial crises. Central banks, exchange rate systems, ministries of finance, and fiscal and budgetary rules are all market-stabilizing institutions.
- *Market-legitimizing institutions* provide social protection and insurance, focus on redistribution, and manage conflict. These institutions help

[a]This text draws heavily on the discussion of institutions found in *Finance and Development* 40, no. 2 (June 2003), particularly Dani Rodrik and Arvind Subramanian, "The Primacy of Institutions (and What This Does and Does Not Mean)"; and Jeffrey D. Sachs, "Institutions Matter, but Not for Everything."

[b]Douglass C. North, *Institutions, Institutional Change, and Economic Performance* (New York: Cambridge University Press, 1990).

[c]North distinguishes between the formal rules and informal constraints (which he includes in his definition of institutions) and organizations (which he excludes). He refers to institutions as the "rules of the game" and organizations as "the players." In common usage, however, many people use the term *institutions* to cover both the rules and the organizations.

protect individuals and corporations from shocks or disasters or from adverse market outcomes. Examples include pension systems, unemployment insurance schemes, welfare programs, and other social funds.

- *Political institutions* determine how society is governed and the extent of political participation. In many countries, there is a strong focus on the key institutions that support democracy, including a free press, elections, competitive political parties, and participatory politics.

A large body of evidence now shows a robust relationship between stronger institutions, rapid economic growth, and improved development outcomes. The evidence is based partly on major advancements in the ability to better measure governance and institutions, such as the data compiled by Dani Kaufmann and Aart Kraay at the World Bank Institute.[d] Strong institutions are central to managing financial systems, building public education and health systems, ensuring efficient trade and commerce, and governing legal systems. Much of neoclassical economic theory on well-functioning markets is based on the assumption that fundamental institutions are in place (such as contract enforcement, perfect information, and the rule of law). But in many low-income countries, these key institutions are weak or nonexistent.

Understanding the importance of institutions for growth brings us only so far, however. Economic analysis tells us very little about the specific forms of institutions that are best suited for a particular environment (for instance, common law versus civil law). There is significant debate about which institutions are most important for low-income countries, their specific form, and the relative importance of institutions versus other factors. Although some analysts claim that institutions dominate all other factors in the growth process, the bulk of the evidence suggests that other factors play an important role, such as policies, geography, and resource endowments. Much of the research indicates that institutions themselves are heavily influenced by geography, history, resource endowments, and the extent of integration with the global economy.

Perhaps even more important, theory and research do not tell us much about how institutions change and how a country with weak institutions can best strengthen them. Institutions change only slowly, but fortunately they do change. Deepening our understanding of how institutions affect growth and development, the appropriate form for institutions in different circumstances and how institutions change over time are major challenges for economists and development specialists in the future.

[d]Available at http://info.worldbank.org/governance/wgi/index.asp.

of law, the extent of corruption, property rights, the quality of government bureaucracies, and other measures of governance and institutional quality.[20]

Stronger governance and institutions help improve the environment for investment by reducing risk and increasing profitability. Investors are more likely to make long-term investments where they feel property rights are secure and their factory, machines, or land will not be confiscated. Strong legal systems can help settle commercial disputes in a predictable, rational manner. Low levels of corruption help reduce the costs of investment, reduce risks, and increase productivity, as managers focus their attention on production rather than influencing politicians and government officials. Strong government economic institutions, such as the central bank, ministry of finance, ports authority, and ministry of trade can help establish effective government policies that influence both factor accumulation and productivity.

The most effective governments established institutions that helped facilitate (rather than hinder) strong economic management, effective social programs, and a robust private sector. Governance in the most rapidly growing countries varied widely from very effective (Singapore and Botswana) to more mixed (Indonesia and Thailand) but generally was better than in slower-growing countries.

4. FAVORABLE ENVIRONMENT FOR PRIVATE ENTERPRISE

Sustained economic growth requires millions of private individuals to make decisions every day regarding saving, investment, education, and job opportunities. Small-scale farmers, business owners, factory workers, and market stall vendors all strive daily to increase their incomes, and the regulatory and policy environment has a significant effect on their success or failure. For many countries, *agricultural policies* are central to the growth process. Where governments have pushed farm-gate prices low to keep food prices cheap or forced farmers to sell their products to government-owned marketing boards, agricultural production (and farmer income) has suffered. The most dramatic example of reducing restrictions of farmers is China's moves to decollectivize agricultural production in the early 1980s and to allow farmers to sell their produce on markets. China's agricultural output soared in the decade that followed. Farmers need reasonable access to fertilizers, seeds, and pesticides, and the construction of rural roads has had a dramatic impact on rural incomes in many countries, such as Indonesia. Absolute free markets are not necessarily the solution—some countries have subsidized fertilizer or other inputs to encourage their use,

[20]See, for example, Stephen Knack and Philip Keefer, "Institutions and Economic Performance: Cross Country Tests Using Alternative Institutional Measures," *Economics and Politics* 7, no. 3 (1995), 207–27; Daniel Kaufmann, Aart Kraay, and Pablo Zoido-Lobatón, "Governance Matters," World Bank Policy Research Paper No. 2196, October 1999; World Bank, Washington, DC; and Daron Acemoglu, Simon Johnson, and James Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91, no. 5 (December 2001), 1369–401.

whereas others have used buffer stocks to counter large swings in prices—but policies that consistently push against markets (rather than helping strengthen markets) almost always fail in the long run.

The climate for small-scale businesses and manufacturing is also important for long-term growth. While some regulation is crucial for well-functioning markets, most governments in developing countries impose unnecessarily high costs on businesses through licensing, permits, and other restrictions. Hernando de Soto's *The Mystery of Capital* demonstrated the damaging effects of heavy business regulation and weak property rights.[21] When the regulatory burden to start a business is high, fewer entrepreneurs bother to start businesses, and when they do, they tend to operate on a smaller scale and in the informal sector. Moreover, government investments in infrastructure, at the core of capital formation, are central. No matter how favorable the policy environment, businesses cannot operate if the electricity shuts off every day, the water is brown, and the phones do not work.

5. TRADE, OPENNESS, AND GROWTH

Most economists agree that international trade plays an important role in economic growth. In discussing this role, however, it is important to distinguish between the volume of trade and trade policy. Economists Jeffrey Frankel and David Romer surmounted significant statistical challenges to demonstrate that trade (as a proportion of GDP) causes growth.[22] They find that a 1 percentage point increase in the ratio of trade to a country's GDP increases income per capita by at least 0.5 percent. Trade, they conclude, plays this role by encouraging countries to accumulate both physical and human capital and by increasing productivity. This finding is consistent with earlier studies that hypothesized trade could contribute to growth by facilitating transfers of both capital and technology and by creating incentives for competitiveness. Yet the conclusion that trade as a share of GDP contributes to growth says nothing directly about the relationship between trade *policy* and growth.

Casual observation suggests that outward orientation (a set of policies designed to encourage exports) contributes to growth. For instance, it's widely accepted that outward-oriented trade policies contributed to the rapid growth of many economies in East Asia, whereas inward-oriented trade policies (that had the effect of discouraging exports) contributed to slow growth in much of Latin America and sub-Saharan Africa. Yet, economists continue to debate the evidence in support of such

[21]Hernando de Soto, *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else* (New York: Basic Books, 2000). The World Bank and International Finance Corporation maintain an extensive program to monitor the costs of doing business in 183 economies. Data and publications from the Doing Business project are available at www.doingbusiness.org, accessed February 2012.

[22]Jeffrey Frankel and David Romer, "Does Trade Cause Growth?" *American Economic Review* 89, no. 3 (June 1999), 379–99.

conclusions. Part of the controversy lies in constructing measurable indicators of trade policy orientation. For instance, economist David Dollar constructed an indicator of outward orientation based on real exchange rate distortion and volatility.[23] He found that, for the period 1976–85, the developing countries that were most open to trade had an average annual growth rate of 2.9 percent per capita, whereas the most closed portion of his sample countries experienced negative growth of −1.3 percent.

In a highly influential study, economists Jeffrey Sachs and Andrew Warner constructed a composite indicator of openness to trade in which countries were deemed either open or closed based on several trade policy indicators.[24] They found openness to have a substantial influence on growth, concluding that open countries (between 1970 and 1989) grew at least 2 percentage points faster on average than did closed countries. Applying the Sachs-Warner indicator to a longer time period, economists Romain Wacziarg and Karen Horn Welch found that open economies over the period 1950–1998 had an average growth rate of income per capita of 2.7 percent as compared with an average growth of of 1.18 percent in closed economies, a margin of nearly 1.5 percentage points.[25] In a prominent response to these studies, however, economists Francísco Rodriguez and Dani Rodrik question the validity, interpretation, and robustness of these policy indicators, concluding that the evidence linking pro-trade policies to growth remains inconclusive.[26] We address these issues in depth in Chapters 17 and 18.

6. FAVORABLE GEOGRAPHY

A striking fact is that there are no rich economies located between the Tropic of Cancer and the Tropic of Capricorn other than Singapore and a few small, oil-rich countries. Figure 3–4 shows that the poorest countries in the world are almost all in the tropics, while the richest countries tend to be in more temperate zones. Even within the temperate zones, the regions closer to the tropics tend to be less well off: Northern Europe is richer than southern Europe, the northern part of the United States is wealthier than the southern parts, and southern Brazil is better off than the north part of the country. In Latin America and Africa, the wealthiest countries are located in the temperate south: Chile, Argentina, and South Africa.

[23]David Dollar, "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976–1985," *Economic Development and Cultural Change*, 40, no. 3 (April 1992), 523–44.

[24]Jeffrey Sachs and Andrew Warner, "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity* 26 (1995), 1–118.

[25]Romain Wacziarg and Karen Horn Welch, "Trade Liberalization and Growth: New Evidence," *World Bank Economic Review* 22, no. 2 (June 2008), 187–231.

[26]Francísco Rodriguez and Dani Rodrik, "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence" (pp. 261–338), in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2000* (Cambridge, MIT Press: National Bureau of Economic Research, 2001).

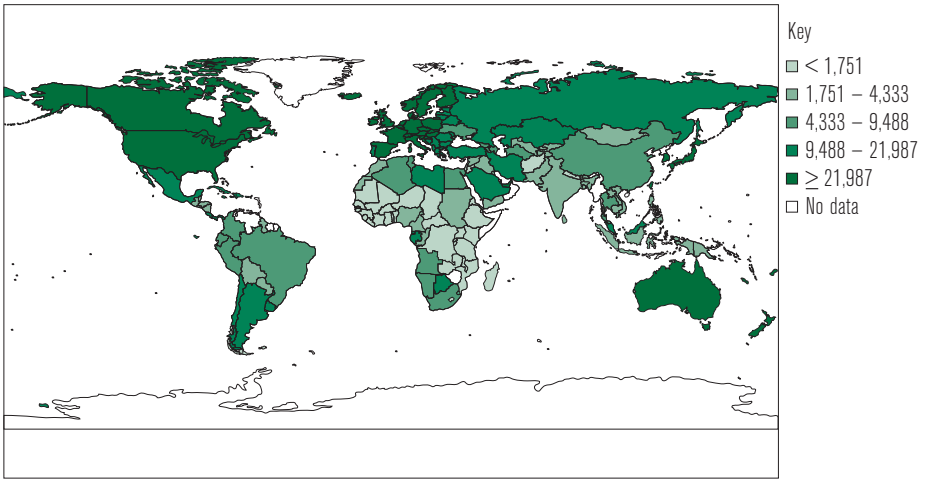


FIGURE 3–4 Income Levels and Geography

Data are gross domestic product per capita in 2009 (constant 2005 international dollars, PPP).

Source: World Bank, "World Bank Indicators," http://databank.worldbank.org.

Several studies have shown a strong relationship between location in the tropics, other geographical characteristics, and growth.[27] Figure 3–5 shows that the average growth rate per capita between 1975 and 2009 for countries located in the tropics was significantly lower than for countries outside the tropics. Tropical countries have to deal with a greater burden from virulent diseases, erratic climate, and at least in some areas, very poor-quality soil for agriculture. Most of the world's most virulent diseases are centered in the tropics, including malaria and HIV/AIDS. These diseases seriously undermine worker productivity and add to healthcare costs. Similarly, although erratic climate can occur anywhere around the world, floods, droughts, and violent storms tend to be more concentrated in the tropics. Hurricanes and typhoons, of course, are by definition tropical phenomena. Hotter climates make a long, hard day of outdoor work much more difficult, reducing labor productivity (one way Singapore has compensated for being in the tropics is by air-conditioning the vast majority of buildings in the country, a step that is much easier for a small city-state like Singapore than for most other countries). And while some tropical regions have very fertile soils (as in the rich lands in Java, one of the main islands of Indonesia), most of the great Sahara desert is in the tropics, as are the arid lands of northern Brazil. These characteristics work to reduce both factor productivity and

[27]Studies that explore the impact of geographical factors on levels of income and growth rates include Robert Hall and Charles Jones, "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics* 114 (February 1999), 83–116; and John Gallup and Jeffrey Sachs, "Geography and Economic Development" (127–78), in Boris Pleskovic and Joseph Stiglitz, eds., *World Bank Annual Conference on Development Economics 1998* (Washington, DC: World Bank, 1998).

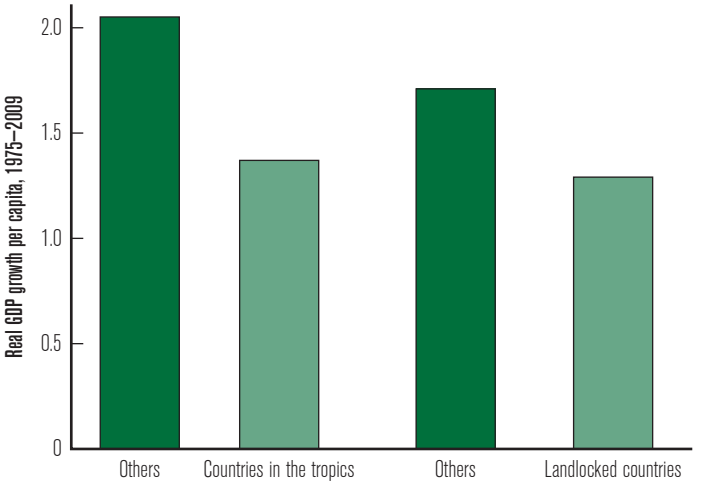


FIGURE 3–5 Geography and Growth

GDP, gross domestic product.

Source: Penn World Tables 7.0, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed May 2011.

the incentives for investment and factor accumulation. Location in the tropics does not necessarily preclude growth and development because some of the burdens can be alleviated at least partially through policies and institutional development, but it clearly creates difficulties and obstacles that make growth and development more difficult.

Another geographical feature that can affect growth is isolation from major markets, such as for landlocked countries and for small island nations (for example, located in the Pacific Ocean). These isolated countries face higher transport costs and fewer economic opportunities than do coastal economies and countries located nearer to major markets. Landlocked African countries face overland shipping costs that can be three times higher than their coastal neighbors'. Higher transport costs make imports more expensive, which both reduces income left for consumption and raises production costs. They also make it more expensive to export products to other countries, reducing profits.

Not all landlocked countries have had poor economic performance. Switzerland and Austria are in some ways the exceptions that prove the rule. Although they are landlocked, they are far from being isolated, as they are located in the heart of Europe. Perhaps the clearest exception is Botswana, which has deftly managed its vast diamond mines to generate sustained growth for the last four decades. Figure 3–5 shows that economic growth in landlocked countries has averaged 1.29 percent since 1975 (excluding Bhutan and Botswana, this figure falls to 0.99 percent), while in coastal economies, growth averaged 1.71 percent, about one-third higher (and nearly twice as high excluding Botswana). Being landlocked does not mean growth is impossible,

but it does limit options and add to production costs. Geographical isolation can be overcome by investments that reduce overland transport costs (for example, better roads or trucks) or by producing goods that rely more on air rather than sea transport. Landlocked Uganda, for example, grows flowers near its international airport for export to Europe. Advancements in satellite communications open up new possibilities for isolated countries, such as data entry or accounting services provided for firms located in other countries.

Note that some countries face multiple geographical obstacles that significantly limit their development options. Perhaps the most challenging are the landlocked countries in the midst of the Sahara desert; Burkina Faso, Chad, Ethiopia, Mali, Niger, and Sudan are among the very poorest countries in the world. These countries have far fewer options and face much more difficult challenges than the average developing country.

These six broad areas are not a complete list of the characteristics that influence factor accumulation, productivity, and economic growth, but they are among the most prominent attributes identified by research and experience. It is important to recognize that this list is not absolute: There is significant variation across countries, and these characteristics are neither a guarantee of success nor a set of rigid requirements for growth. Some countries have done relatively well in many of these areas and still have not experienced rapid economic growth. At the same time, while almost all of the fastest-growing countries score well in most of these areas, some do not. Our understanding of the precise pathways through which each of these factors influence growth is far from complete. However, the evidence does show that these characteristics are among the most important factors supporting factor accumulation, productivity, and growth.

SUMMARY

- Countries have varied widely in their experiences of economic growth since 1960. While some economies, such as Madagascar and Zambia shrank slightly in per capita terms, others such as Botswana and China increased their per capita incomes by factors of 15 to 19 between 1960 and 2009.
- Theories of economic growth have identified several key dimensions of the growth process. Factor accumulation—the expansion of a country's capital stock along with the growth of its labor force—is critical. These are the principal inputs for production of national output.
- In the long run, factor accumulation alone cannot sustain rapid growth in per capita income. Long-run growth depends on being able to produce more output per unit of input. This is the definition of productivity growth. Productivity growth is ultimately a reflection of technical change (which

we can think of as the invention of better recipes for producing national output).

- Productivity growth is measured as a residual. It is the increase in output that is not explained by increased use of inputs. Any errors in the measurement of either inputs or output spill over into our estimate of productivity growth. The goal of sources of growth analysis is to decompose a change in output into the portion explained by the changing use of each input and the unexplained portion that we designate as productivity growth.
- Sources of growth analyses indicate that capital accumulation tends to contribute substantially to growth in low-income countries, while increases in the size and quality of the labor force and productivity growth also make important contributions. Productivity growth tends to account for a larger share of growth in high-income countries.
- Some of the key country characteristics most closely associated with rapid growth include economic and political stability, investments in health and education, strong governance and institutions, a favorable environment for private enterprise (including agricultural, regulatory policies), and more-favorable geography. Participation in international trade is also strongly associated with economic growth. Our understanding of precisely how these and other factors affect the growth process is far from complete.



Theories of Economic Growth

Economists have long puzzled over the question of economic growth. What is it that makes some countries rich while others remain poor? Formal studies of this question date back at least to the eighteenth-century writings of Scottish social philosopher Adam Smith. The search for answers continues to dominate economic thinking. In a lecture presented in 1985, Nobel Prize–winning economist Robert Lucas noted the then-rapid economic growth of Indonesia and Egypt and slow growth of India, famously asking,

> Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so,what exactly? If not, what is it about the "nature of India" that makes it so? The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them it is hard to think of anything else.[1]

More than a quarter century later, economists Dani Rodrik, Arvind Subramanian, and Francesco Trebbi noted,

> Average income levels in the world's richest and poorest nations differ by a factor of more than 100. Sierra Leone, the poorest economy for which we have national income statistics, has a per-capita GDP of $490, compared to Luxembourg's $50,061. What accounts for these differences, and what (if anything) can we do

[1]Robert E. Lucas, "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, no. 1 (July 1988), 3–42.

> to reduce them? It is hard to think of any question in economics that is of greater intellectual significance, or of greater relevance to the vast majority of the world's population.[2]

Two observations emerge from the juxtaposition of these strikingly similar quotations: (1) Long-run economic growth may be the single most fundamental determinant of human welfare around the world, and (2) despite substantial efforts and significant progress toward solving the puzzle of economic growth during the 27 years that separate these comments, we remain far from a complete and policy-relevant understanding of the deep determinants of growth.

This chapter explores key contributions to the theory of economic growth. We began to explore these issues in the last chapter by examining some of the basic processes and patterns that characterize economic growth in low-income countries. We emphasized that growth depends on two processes: the *accumulation of assets* (such as capital, labor, and land), and *making those assets more productive*. Saving and investment are central, but investments must be productive for growth to proceed. Our approach was largely empirical, as we examined much of the data on growth and some of the key findings from research on the determinants of growth across countries. We saw that government policy, institutions, political and economic stability, geography, natural resource endowments, and levels of health and education all play some role in influencing economic growth. We emphasized that growth is not the same as development, but it remains absolutely central to the development process.

This chapter develops these ideas more formally by introducing the underlying theory and the most important basic models of economic growth that influence development thinking today. These models provide consistent frameworks for understanding the growth process and provide a theoretical foundation for the empirical approach we took in the last chapter. Here, we identify specific mathematical relationships between the quantity of capital and labor, their productivity, and the resulting aggregate output. It is important that these models also explore the process of accumulating *additional* capital and labor and *increasing* their productivity, which shifts the model from determining the *level* of output to the *rate of change* of output, which of course is the rate of economic growth.

As we begin to examine the models, it is useful to consider the words of Robert Solow, the father of modern growth theory, who once wrote: "All theory depends on assumptions that are not quite true. That is what makes it theory. The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that

[2]Dani Rodrik, Arvind Subramanian, and Francesco Trebbi, "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9, no. 2 (2004), 131–65.

the final results are not very sensitive."[3] The best models are simple, yet still manage to communicate powerful insights into how the real world operates. In this spirit, the models presented here make assumptions that clearly are not true but allow us to simplify the framework and make it easier to grasp key concepts and insights. For example, we begin by assuming that our prototype economy has one type of homogeneous worker and one type of capital good that combine to produce one standard product. No economy in the world has characteristics even closely resembling these assumptions, but making these assumptions allows us to cut through many details and get to the core concepts of the theory of economic growth.

THE BASIC GROWTH MODEL

The most fundamental models of economic output and economic growth are based on a small number of equations that relate saving, investment, and population growth to the size of the workforce and capital stock and, in turn, to aggregate production of a single good. These models initially focus on the *levels* of investment, labor, productivity, and output. It then becomes straightforward to examine the *changes* in these variables. Our ultimate focus is to explore the key determinants of the *change in output*—that is, on the rate of economic growth. The version of the basic model that we examine here has five equations: an aggregate production function, an equation determining the level of saving, the saving-investment identity, a statement relating new investment to changes in the capital stock, and an expression for the growth rate of the labor force.[4] We examine each of these in turn.

Standard growth models have at their core a production function. At the individual firm or microeconomic level, a production function relates the number of employees and machines to the size of the firm's output. For example, the production function for a textile factory would reveal how much more output the factory could produce if it hired (say) 50 additional workers and purchased five more looms. Production functions often are derived from engineering specifications that relate given amounts of physical input to the amount of physical output that can be produced with that input. At the national or economywide level, production functions describe the relationship of the size of a country's total labor force and the value of its capital stock with the level of that country's gross domestic product (GDP; its total output). These economywide relationships are called **aggregate production functions**.

[3]Robert Solow, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70 (February 1956), 65–94.

[4]This five-equation presentation is based on teaching notes compiled by World Bank economist Shantayanan Devarajan, to whom we are indebted.

Our first equation is an aggregate production function. If Y represents total output (and therefore total income), K is the capital stock, and L is the labor supply, at the most general level, the aggregate production function can be expressed as

$$Y = F(K, L) \qquad [4\text{–}1]$$

This expression indicates that output is a function (denoted by F) of the capital stock and the labor supply. As the capital stock and labor supply grow, output expands. Economic growth occurs by increasing either the capital stock (through new investment in factories, machinery, equipment, roads, and other infrastructure), the size of the labor force, or both. The exact form of the function F (stating precisely *how much* output expands in response to changes in K and L) is what distinguishes many different models of growth, as we will see later in the chapter. The other four equations of the model describe how these increases in K and L come about.

Equations 4–2 through 4–4 are closely linked and together describe how the capital stock (K) changes over time. These three equations first calculate total saving, then relate saving to new investment, and finally describe how new investment changes the size of the capital stock. To calculate saving, we take the most straightforward approach and assume that saving is a fixed share of income:

$$S = sY \qquad [4\text{–}2]$$

In this equation, S represents the total value of saving, and s represents the average saving rate. For example, if the average saving rate is 20 percent and total income is $10 billion, then the value of saving in any year is $2 billion. We assume that the saving rate s is a constant, which for most countries is between 10 and 40 percent (typically averaging between 20 and 25 percent), although for some countries it can be higher or lower. China's savings rate in 2008 (along with those of several large oil exporters) exceeded 50 percent, while several countries (including Mozambique, Guinea, the Seychelles, and Georgia) reported savings rates less than 5 percent of GDP. Actual saving behavior is more complex than this simple model suggests (as we discuss in Chapter 10), but this formulation is sufficient for us to explore the basic relationships between saving, investment, and growth.

The next equation relates total saving (S) to investment (I). In our model, with only one good, there is no international trade (because everyone makes the same product, there is no reason to trade). In a closed economy (one without trade or foreign borrowing), saving must be equal to investment. All output of goods and services produced by the economy must be used for either current consumption or investment, while all income earned by households must be either consumed or saved. Because output is equal to income, it follows that saving must equal investment. This relationship is expressed as

$$S = I \qquad [4\text{–}3]$$

We are now in a position to show how the capital stock changes over time. Two main forces determine changes in the capital stock: new investment (which adds to the capital stock) and depreciation (which slowly erodes the value of the existing capital stock over time). Using the Greek letter delta (Δ) to represent the change in the value of a variable, we express the change in the *capital* stock as ΔK, which is determined as follows:

$$\Delta K = I - (dK) \qquad [4\text{–}4]$$

In this expression d is the rate of depreciation. The first term (I) indicates that the capital stock *increases* each year by the amount of new investment. The second term $-(d \times K)$ shows that the capital stock *decreases* every year because of the depreciation of existing capital. We assume here that the depreciation rate is a constant, usually in the range of 2 to 10 percent.

To see how this works, let us continue our earlier example, in which total income is \$10 billion and saving (and therefore investment) is \$2 billion. Say that the value of the existing capital stock is \$30 billion and the annual rate of depreciation is 3 percent. In this example, the capital stock increases by \$2 billion because of new investment but also decreases by \$0.9 billion (3 percent $\times$ \$30 billion) because of depreciation. Equation 4–4 puts together these two effects, calculating the change in the capital stock as $\Delta K = I - (d \times K) = \$2\,\text{billion} - (0.03 \times \$30\,\text{billion}) = \$1.1\,\text{billion}$. Thus the capital stock increases from \$30 billion to \$31.1 billion. This new value of the capital stock then is inserted into the production function in equation 4–1, allowing for the calculation of a new level of output, Y.

The fifth and final equation of the model focuses on the supply of labor. To keep things simple, we assume that the labor force grows exactly as fast as the total population. Over long periods of time, this assumption is fairly accurate. If n is equal to the growth rate of both the population and the labor force, then the change in the labor force (ΔL) is represented by

$$\Delta L = nL \qquad [4\text{–}5]$$

If the labor force consists of 1 million people and the population (and labor force) is growing by 2 percent, the labor force increases annually by 20,000 (1 million $\times$ 0.02) workers. The labor force now consists of 1.02 million people, a figure that can be inserted into the production function for L to calculate the new level of output. (If we divide both sides of equation 4–5 by L, we can see directly the rate of growth of the labor force, $\Delta L/L = n$.)

These five equations represent the complete model.[5] Collectively, they can be used to examine how changes in population, saving, and investment initially affect

[5]Note that because the model has five equations and five variables (Y, K, L, I, and S) it always can be solved. In addition, there are three fixed parameters (d, s, and n), the values of which are assumed to be fixed exogenously, or outside the system.

the capital stock and labor supply and ultimately determine economic output. New saving generates additional investment, which adds to the capital stock and allows for increased output. New workers add further to the economy's capacity to increase production.

One way these five equations can be simplified slightly is to combine equations 4–2, 4–3, and 4–4. The aggregate level of saving (in equation 4–2) determines the level of investment in equation 4–3, which (together with depreciation) determines changes in the capital stock in equation 4–4. Combining these three equations gives us

$$\Delta K = sY - dK \qquad [4\text{–}6]$$

This equation states that the change in the capital stock (ΔK) is equal to saving (sY) minus depreciation (dK). This expression allows us to calculate the change in the capital stock and enter the new value directly into the aggregate production function in equation 4–1.

THE HARROD-DOMAR GROWTH MODEL

As we have stressed, the aggregate production function (shown earlier as equation 4–1) is at the heart of every model of economic growth. This function can take many different forms, depending on what we believe is the true relationship between the factors of production (K and L) and aggregate output. This relationship depends on (among other things) the mix of economic activities (for example, agriculture, heavy industry, light labor-intensive manufacturing, high-technology processes, services), the level of technology, and other factors. Indeed, much of the theoretical debate in the academic literature on economic growth is about how to best represent the aggregate production process.

THE FIXED-COEFFICIENT PRODUCTION FUNCTION

One special type of a simple production function is shown in Figure 4–1. Output in this figure is represented by **isoquants**, which are combinations of the inputs (labor and capital in this case) that produce equal amounts of output. For example, on the first (innermost) isoquant, it takes capital (plant and equipment) of $10 million and 100 workers to produce 100,000 keyboards per year (point *a*). Alternatively, on the second isoquant, $20 million of capital and 200 workers can produce 200,000 keyboards (point *b*). Only two isoquants are shown in this diagram, but a nearly infinite number of isoquants are possible, each for a different level of output.

The L-shape of the isoquants is characteristic of a particular type of production function known as **fixed-coefficient production functions**. These production

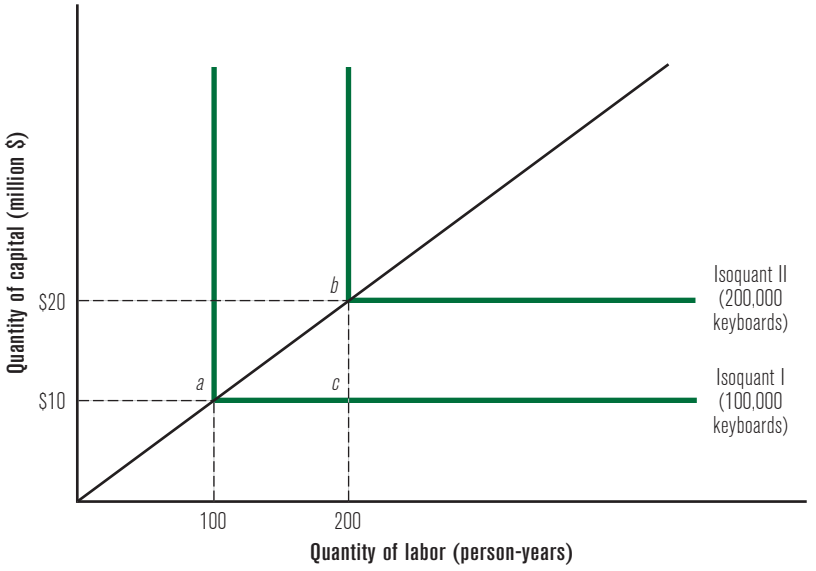


FIGURE 4–1 Isoquants for a Fixed-Coefficient Production Technology

With constant returns to scale, the isoquants will be L-shaped and the production function will be the straight line through their minimum-combination points.

functions are based on the assumption that capital and labor need to be used in a fixed proportion to each other to produce different levels of output. In Figure 4–1, for the first isoquant, the **capital–labor ratio** is 10 million:100, or 100,000:1. In other words, $100,000 in capital must be matched with one worker to produce the given output. For the second isoquant, the ratio is the same: $20 million:200, or 100,000:1. This constant capital–labor ratio is represented in Figure 4–1 by the slope of the ray from the origin through the vertices (*a* and *b*) of the isoquants. These vertices represent the least cost and hence most efficient mix of capital and labor to produce a given quantity of output. In the case of fixed coefficients, this mix is the same for every quantity of output.

With this kind of production function, if more workers are added *without* investing in more capital, output does *not* rise. Look again at the first isoquant, starting at the elbow (with 100 workers and $10 million in capital). If the firm adds more workers (say, increasing to 200 workers) without adding new machines, it moves horizontally to the right along the first isoquant to point *c*. But at this point, or at any other point on this isoquant, the firm still produces just 100,000 keyboards. In this kind of production function, new workers need more machines to increase output. Adding new workers without machines results in idle workers, with no increase in output. Similarly, more machinery without additional workers results in underused machines. On each isoquant, the most efficient production point is at the elbow,

where the minimum amounts of capital and labor are used. To use any more of either factor without increasing the other is a waste.

The production technology depicted in Figure 4–1 also is drawn with **constant returns to scale**, so if capital is doubled to $20 million and labor is doubled to 200 workers, output also exactly doubles to 200,000 keyboards per year.[6] With this further assumption, two more ratios remain constant at any level of output: capital to output and labor to output. If keyboards are valued at $50 each, then 100,000 keyboards are worth $5 million. In this case, in the first isoquant, $10 million in capital is needed to produce $5 million worth of keyboards, so the **capital–output ratio** is $10 million: $5 million, or 2:1. In the second isoquant the ratio is the same ($20 million:$10 million, or 2:1). Similarly, for each isoquant the **labor–output ratio** is also a constant, in this case equal to 1:50,000, meaning that each worker produces $50,000 worth of keyboards, or 1,000 keyboards each.

THE CAPITAL–OUTPUT RATIO AND THE HARROD-DOMAR FRAMEWORK

The fixed-coefficient, constant-returns-to-scale production function is the centerpiece of a well-known early model of economic growth that was developed independently during the 1940s by economists Roy Harrod of England and Evsey Domar of MIT, primarily to explain the relationship between growth and unemployment in advanced capitalist societies.[7] It ultimately focuses attention on the role of capital accumulation in the growth process. The **Harrod-Domar model** has been used extensively (perhaps even overused) in developing countries to examine the relationship between growth and capital requirements. The model is based on the real-world observation that some labor is unemployed and proceeds on the basis that capital is the binding constraint on production and growth. In the model, the production function has a very precise form, in which output is assumed to be a *linear* function of capital (and only capital). As usual, the model begins by specifying the level of output, which we later modify to explore changes in output, or economic growth. The production function is specified as follows:

$$Y = 1/v \times K \quad \textit{or} \quad Y = K/v \qquad [4\text{–}7]$$

where v is a constant. In this equation, the capital stock is multiplied by the fixed number $1/v$ to calculate aggregate production. If $v = 3$ and a firm has $30 million in capital, its annual output would be $10 million. It is difficult to imagine a simpler

[6] In a constant-returns-to-scale production function, if we multiply both capital and labor by any number, w, output multiplies by the same number. In other words, the production function has the following property: $wY = F(wK, wL)$.

[7] Roy F. Harrod, "An Essay in Dynamic Theory," *Economic Journal* (1939), 14–33; Evsey Domar, "Capital Expansion, Rate of Growth, and Employment," *Econometrica* (1946), 137–47; and Domar, "Expansion and Employment," *American Economic Review* 37 (1947), 34–55.

production function. The constant v turns out to be the capital–output ratio because, by rearranging the terms in equation 4–7, we find

$$v = K/Y \qquad [4\text{–}8]$$

The capital–output ratio is a very important parameter in this model, so it is worth dwelling for a moment on its meaning. This ratio essentially is a measure of the productivity of capital or investment. In the earlier example in Figure 4–1, it took $10 million in investment in a new plant and new equipment to produce $5 million worth of keyboards, implying a capital-output ratio of 2:1 (or just 2). A larger v implies that more capital is needed to produce the same amount of output. So, if v were 4 instead, then $20 million in investment would be needed to produce $5 million worth of keyboards.

The capital–output ratio provides an indication of the capital intensity of the production process. In the basic growth model, this ratio varies across countries for two reasons: either the countries use different technologies to produce the same goods or they produce a different mix of goods. Where farmers produce maize using tractors, the capital–output ratio will be much higher than in countries where farmers rely on a large number of workers using hoes and other hand tools. In countries that produce a larger share of **capital-intensive products** (that is, those that require relatively more machinery, such as automobiles, petrochemicals, and steel), v is higher than in countries producing more **labor-intensive products** (such as textiles, basic agriculture, and foot wear). In practice, as economists move from the v of the model to actually measuring it in the real world, the observed capital–output ratio can also vary for a third reason: differences in efficiency. A larger measured v can indicate less-efficient production when capital is not being used as productively as possible. A factory with lots of idle machinery and poorly organized production processes has a higher capital–output ratio than a more-efficiently managed factory.

Economists often calculate the **incremental capital–output ratio (ICOR)** to determine the impact on output of additional (or incremental) capital. The ICOR measures the productivity of additional capital, whereas the (average) capital–output ratio refers to the relationship between a country's total stock of capital and its total national product. In the Harrod-Domar model, because the capital–output ratio is assumed to remain constant, the average capital–output ratio is equal to the incremental capital–output ratio, so the ICOR $= v$.

So far, we have been discussing total output, not growth in output. The production function in equation 4–7 easily can be converted to relate *changes* in output to *changes* in the capital stock:

$$\Delta Y = \Delta K/v \qquad [4\text{–}9]$$

The growth rate of output, g, is simply the increment in output divided by the total amount of output, $\Delta Y/Y$. If we divide both sides of equation 4–9 by Y, then

$$g = \Delta Y/Y = \Delta K/Yv \qquad [4\text{–}10]$$

Finally, from equation 4–6, we know that the change in the capital stock ΔK is equal to saving minus the depreciation of capital ($\Delta K = sY - dK$). Substituting the right-hand side of equation 4–6 into the term for ΔK in equation 4–10 and simplifying[8] leads to the basic Harrod-Domar relationship for an economy:

$$g = (s/v) - d \qquad [4\text{–}11]$$

Underlying this equation is the view that capital created by investment is the main determinant of growth in output and that saving makes investment possible.[9] It rivets attention on two keys to the growth process: saving (s) and the productivity of capital (v). The message from this model is clear: Save more and make productive investments, and your economy will grow.

Economic analysts can use this framework either to predict growth or to calculate the amount of saving required to achieve a target growth rate. The first step is to try to estimate the incremental capital–output ratio (v) and depreciation rate (d). With a given saving rate, predicting the growth rate is straightforward. If the saving (or investment) rate is 24 percent, the ICOR is 3, and the depreciation rate is 5 percent, then the economy can be expected to grow by 3 percent (because $0.24/3 - 0.05 = 0.03$).

How does this model work in practice? Consider Indonesia, which from 2002 to 2007 had an investment rate of about 30 percent and recorded a GDP growth just under 5.5 percent per year. Assuming a depreciation rate of 5 percent, the implied ICOR was approximately $v = 2.86$.[10] Would these figures have helped the Indonesian government predict the 2007–08 growth rate? In 2008, the investment rate was 29 percent, so the Harrod-Domar model would have predicted growth of 5.1 percent ($g = 0.29/2.86 - 0.05$). The actual growth rate in 2007–08 was 6.0 percent, within sight of the prediction but not highly accurate.

STRENGTHS AND WEAKNESSES OF THE HARROD-DOMAR FRAMEWORK

The basic strength of the Harrod-Domar model is its simplicity. The data requirements are small, and the equation is easy to use and to estimate. And, as we saw with the example of Indonesia, the model can be somewhat accurate from one year to the next. Generally speaking, in the absence of severe economic shocks (such as a drought, a financial crisis, or large changes in export or import prices), the model can do a reasonable job of estimating expected growth rates in most countries over very short periods of time (a few years). Another strength is its focus on the key role of saving. As discussed

[8]Substituting equation 4–6 into 4–10 leads to $g = (sY - dK)/Y \times 1/v$, which can be simplified to $g = (s - d \times K/Y) \times 1/v$. Since $K/Y = v$, we have $g = (s - dv) \times 1/v$, which leads to $g = s/v - d$.

[9]For an important early contribution to the discussion of the importance of capital accumulation to the growth process, see Joan Robinson, *The Accumulation of Capital* (London: Macmillan, 1956).

[10]Because $g = s/v - d$, then $v = s/(g + d)$. For Indonesia between 2002 and 2007, $v = 0.30/(0.055 + 0.05) = 2.86$.

in Chapter 3, individual decisions about how much income to save and consume are central to the growth process. People prefer to consume sooner rather than later, but the more that is consumed, the less that can be saved to finance investment. The Harrod-Domar model makes it clear that saving is crucial for income to grow over time.

The model, however, has some major weaknesses. One follows directly from the strong focus on saving. Although saving is necessary for growth, the simple form of the model implies that it is also sufficient, which it is not. As pointed out in Chapter 3, the investments financed by saving actually have to pay off with higher income in the future, and not all investments do so. Poor investment decisions, changing government policies, volatile world prices, or simply bad luck can alter the impact of new investment on output and growth. Sustained growth depends both on generating new investment and ensuring that investments are productive over time. In this vein, the allocation of resources across different sectors and firms can be an important determinant of output and growth. Because (for simplicity) the Harrod-Domar assumes only one sector, it leaves out these important allocation issues.

Perhaps the most important limitations in the model stem from the rigid assumptions of fixed capital-to-labor, capital-to-output, and labor-to-output ratios, which imply very little flexibility in the economy over time. To keep these ratios constant, capital, labor, and output must all grow at exactly the same rate, which is highly unlikely to happen in real economies. To see why these growth rates must all be the same, consider the growth rate of capital. If the capital stock grew any faster or slower than output at rate g, the capital–output ratio would change. Thus the capital stock must grow at g to keep the capital–output ratio constant over time. With respect to labor, in our original five-equation model, we stipulated (in equation 4–5) that the labor force would grow at exactly the same pace as the population at rate n. Therefore, the only way that the capital stock and the labor force can grow at the same rate is if n happens to be equal to g. This happens only when $n = g = s/v - d$, and there is no particular reason to believe the population will grow at that rate.

In this model, the economy remains in equilibrium with full employment of the labor force and the capital stock *only* under the very special circumstances that labor, capital, and output all *grow* at the rate g. On the one hand, if n is larger than g, the labor force grows faster than the capital stock. In essence, the saving rate is not high enough to support investment in new machinery sufficient to employ all new workers. A growing number of workers do not have jobs and unemployment rises indefinitely. On the other hand, if g (or $s/v - d$) is larger than n, the capital stock grows faster than the workforce. There are not enough workers for all the available machines, and capital becomes idle. The actual growth rate of the economy no longer is g, as the model stipulates, but slows to n, with output constrained by the number of available workers.

So, unless $s/v - d$ (or g) is exactly equal to n, either labor or capital is not fully employed and the economy is not in a stable equilibrium. This characteristic of the Harrod-Domar model has come to be known as the *knife-edge* problem. As long as

$g = n$, the economy remains in equilibrium, but as soon as either the capital stock or the labor force grows faster than the other, the economy falls off the edge with continuously growing unemployment of either capital or labor.

The rigid assumptions of fixed capital–output, labor–output, and capital–labor ratios may be reasonably accurate for short periods of time or in very special circumstances but almost always are inaccurate over time as an economy evolves and develops. Each of these varies among countries and, for a single country, over time. Consider the incremental capital–output ratio. The productivity of capital can change in response to policy changes, which in turn affects v. Moreover, the capital intensity of the production process can and usually does change over time. A poor country with a low saving rate and surplus labor (unemployed and underemployed workers) can achieve higher growth rates by utilizing as much labor as possible and thus relatively less capital. For example, a country relying heavily on labor-intensive agricultural production will record a low v. As economies grow and per capita income rises, the labor surplus diminishes and economies shift gradually toward more capital-intensive production. As a result, the ICOR shifts upward. Thus a higher v may not necessarily imply inefficiency or slower growth. ICORs can also shift through market mechanisms, as prices of labor and capital change in response to changes in supplies. As growth takes place, saving tends to become relatively more abundant and hence the price of capital falls while employment and wages rise. Therefore, all producers increasingly economize on labor and use more capital and the ICOR tends to rise.

Consider again the example of Indonesia. The ICOR changed from approximately 2.4 during the 1980s, to 4.1 during the 1990s, to 3.6 during 2000–09, reflecting a trend toward more capital-intensive production processes. To continue to use the 1980s ICOR in 2009 would have been very misleading and betrayed a significant misunderstanding of the growth process. The structure of the economy had changed substantially during that time period and, with it, the ICOR. Thailand provides a similar example, as described in Box 4–1.

As a result of these rigidities, the Harrod-Domar framework tends to become increasingly inaccurate over longer periods of time as the actual ICOR changes and, with it, the capital–labor ratio. In a world with fixed-coefficient production functions, little room is left for a factory manager to increase output by hiring one more worker without buying a machine to go with the worker or to purchase more machines for the current workforce to use. The fixed-proportion production function does not allow for any substitution between capital and labor in the production process. In the real world, of course, at least some substitution between labor and capital is possible in most production processes. As we see in the next section, adding this feature to the model allows for a much richer exploration of the growth process.

A final weakness of the Harrod-Domar model is the absence of any role for productivity growth—the ability to produce increasing quantities of output per unit of input. In Figure 4–1, increased factor productivity can be represented by an inward shift of each isoquant toward the origin, implying that less labor and capital would

BOX 4-1 ECONOMIC GROWTH IN THAILAND

In the 1960s, Thailand's agrarian economy depended heavily on rice, maize, rubber, and other agricultural products. About three-quarters of the Thai population derived its income from agricultural activities. Gross domestic product (GDP) per capita in 1960 (measured in 2005 international dollars) was around $1,200—less than one-tenth the average income in the United States. Life expectancy was 53 years and the infant mortality rate was 103 per 1,000 births. Few observers expected Thailand to develop rapidly.

However, since the mid-1960s, the Thai economy has grown rapidly (if not always steadily), benefiting from relatively sound economic management and a favorable external environment. The government regularly achieved surpluses on the current account of its budget and used these funds (plus modest inflows of foreign assistance) to finance investments in rural roads, irrigation, power, telecommunications, and other basic infrastructure. At least until the mid-1990s, the government's fiscal, monetary, and exchange rate policies kept the macroeconomy relatively stable with fairly low inflation, despite the turbulent period of world oil price shocks in the 1970s and 1980s. Beginning in the 1970s, the government began to remove trade restrictions and promote the production of labor-intensive manufactured exports. These products found a ready market in the booming Japanese economy of the 1980s and provided a growing number of jobs for Thai workers.

Thailand's ability to make investments and deepen its capital stock depended on its capacity to save. The country's saving rate averaged about 20 percent in the 1960s, already high for developing countries, and increased steadily over time to an average of 35 percent in the 1990s, falling to about 32 percent during the 2000s. These high saving rates, combined with relatively prudent economic policies, supported very rapid economic growth and development.

Thailand's development experience has been far from completely smooth, however. In mid-1997, a major financial crisis erupted. Huge short-term offshore borrowing combined with a fixed exchange rate and weak financial institutions led to a collapse of a real estate bubble, rapid capital flight, a substantial depreciation of the Thai baht, and a deep recession (see Chapter 13). In some ways, Thailand had become the victim of its own success, with its rapid growth attracting significant numbers of investors looking to gain quick profits, who rapidly fled once the bubble began to collapse. After two years of negative growth (with GDP falling 10 percent in 1998), the economy began to recover and growth rebounded to 3.9 percent between 1999 and 2007.

Over the longer period between 1960 and 2007, per capita growth averaged 4.5 percent, so that the average income in Thailand is now more than eight times higher than it was in 1960. By 2009, life expectancy grew to 69 years, infant mortality fell to 12 per thousand, and adult literacy reached 93 percent. During this period, the structure of the economy changed significantly. By 2009, manufacturing accounted for 34 percent of GDP, up from just 14 percent in 1965, while the share of agricultural production dropped commensurately. The composition of exports shifted away from rice, maize, and other agricultural commodities toward labor-intensive manufactured products, which now account for a large majority of all exports. As the Harrod-Domar model predicts, Thailand's high saving rate and resulting capital accumulation was accompanied by a dramatic increase in output (and income) per capita. Contrary to the Harrod-Domar model, however, the ICOR did not remain constant. As the stock of capital grew and the economy shifted toward more capital-intensive production techniques, the ICOR increased from 2.6 in the 1970s to nearly 5 by the early 2000s. The rising ICOR indicated that, as the Thai economy expanded and the level of capital per worker increased, an ever-larger increment of new capital was required to bring about a given increase in total output.

be needed to produce the same amount of output. The simplest way to capture this in the Harrod-Domar framework is to introduce a smaller ICOR, but of course, this would contradict the idea of a constant ICOR.

Despite these weaknesses, the Harrod-Domar model is still used to a surprisingly wide extent. Economist William Easterly documented how the World Bank and other institutions use the model to calculate "financing gaps" between the amount of available saving and the amount of investment supposedly needed to achieve a target growth rate.[11] He shows how simplistic and sometimes careless use of the model can lead to weak analysis and faulty conclusions. In essence, analysts enamored by the simplicity of the model tend to overlook its shortcomings when applying it to the real world.

The Harrod-Domar model provides some useful insights but does not take us very far. The fixed-coefficient assumption provides the model with very little flexibility and does not capture the ability of real world firms to change the mix of inputs in the production process. The model can be reasonably accurate from one year to the next (in the absence of shocks), and it rightly focuses attention on the importance

[11]See William Easterly, "Aid for Investment," *The Elusive Quest for Growth* (Cambridge: MIT Press, 2001), chap. 2; and Easterly, "The Ghost of the Financing Gap: Testing the Growth Model of the International Financial Institutions," *Journal of Development Economics* 60, no. 2 (December 1999), 423–38.

of saving. But it is quite inaccurate for most countries over longer periods of time and implies that saving is sufficient for growth, although it is not. Indeed, in the late 1950s, Domar expressed strong doubts about his own model, pointing out that it was originally designed to explore employment issues in advanced economies rather than growth per se and was too rigid to be useful for explaining long-term growth.[12] Instead, he endorsed the new growth model of Robert Solow, to which we now turn our attention.

THE SOLOW (NEOCLASSICAL) GROWTH MODEL

THE NEOCLASSICAL PRODUCTION FUNCTION

In 1956, MIT-economist Robert Solow introduced a new model of economic growth that was a big step forward from the Harrod-Domar framework.[13] Solow recognized the problems that arose from the rigid production function in the Harrod-Domar model. Solow's answer was to drop the fixed-coefficients production function and replace it with a **neoclassical production function** that allows for more flexibility and substitution between the factors of production. In the Solow model, the capital–output and capital–labor ratios no longer are fixed but vary, depending on the relative endowments of capital and labor in the economy and the production process. Like the Harrod-Domar model, the Solow model was developed to analyze industrialized economies, but it has been used extensively to explore economic growth in all countries around the world, including developing countries. The Solow model has been enormously influential and remains at the core of most theories of economic growth in developing countries.

The isoquants that underlie the neoclassical production function are shown in Figure 4–2. Note that the isoquants are curved rather than L-shaped as in the fixed-coefficient model. In this figure, at point *a*, $10 million of capital and 100 workers combine to produce 100,000 keyboards, which would be valued at $5 million (because, as stated earlier, keyboards are priced at $50 each). Starting from this point, output could be expanded in any of three ways. If the firm's managers decided to expand at constant factor proportions and move to point *b* on isoquant II to produce 200,000 keyboards, the situation would be identical to the fixed propor-

[12]Evsey Domar, *Essays in the Theory of Economic Growth* (Oxford: Oxford University Press, 1957).

[13]The two classic references of Solow's work are his "A Contribution to the Theory of Economic Growth" and "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39 (August 1957), 312–20. For an excellent and very thorough undergraduate exposition of the Solow and other models of economic growth, see Charles I. Jones, *Introduction to Economic Growth* (New York: W. W. Norton and Company, 2001). In 1987, Solow was awarded the Nobel Prize in economics, primarily for his work on growth theory.

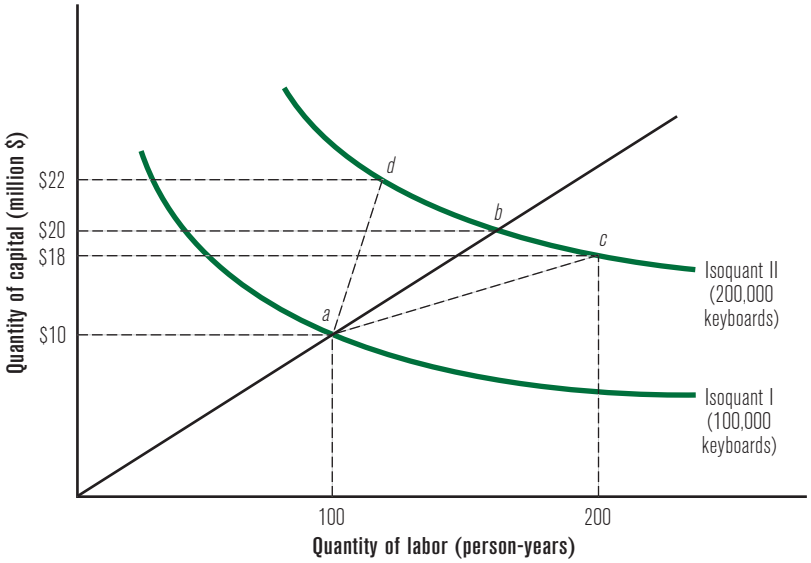


FIGURE 4–2 Isoquants for a Neoclassical Production Technology

Instead of requiring fixed factor proportions, as in Figure 4–1, output can be achieved with varying combinations of labor and capital. This is called a *neoclassical* production function. Note that the isoquants are curved rather than L-shaped.

tions case of Figure 4–1. The capital–output ratio at both points *a* and *b* would be 2:1, as it was before ($10 million of capital produces $5 million of keyboards at point *a*, and $20 million of capital produces $10 million of keyboards at point *b*). Note that the Solow model retains from the Harrod-Domar model the assumption of constant returns to scale, so that a doubling of labor and capital leads to a doubling of output. But by dropping the fixed-coefficients assumption, production of 200,000 keyboards could be achieved by using different combinations of capital and labor. For example, the firm could use more labor and less capital (a more labor-intensive method), such as at point *c* on isoquant II. In that case, the capital–output ratio falls to 1.8:1 ($18 million in capital to produce $10 million in keyboards).

Alternatively, the firm could choose a more capital-intensive method, such as at point *d* on isoquant II, where the capital–output ratio would rise to 2.2:1 ($22 million in capital to produce $10 million in keyboards). The kinds of tools that policy makers use to try to decrease or increase the capital–output ratio are discussed in depth in several chapters later in this text.

THE BASIC EQUATIONS OF THE SOLOW MODEL

The Solow model is understood most easily by expressing all the key variables in per-worker terms (for example, output per worker and capital per worker). To do so, we divide both sides of the production function in equation 4–1 by L, so that it takes the form

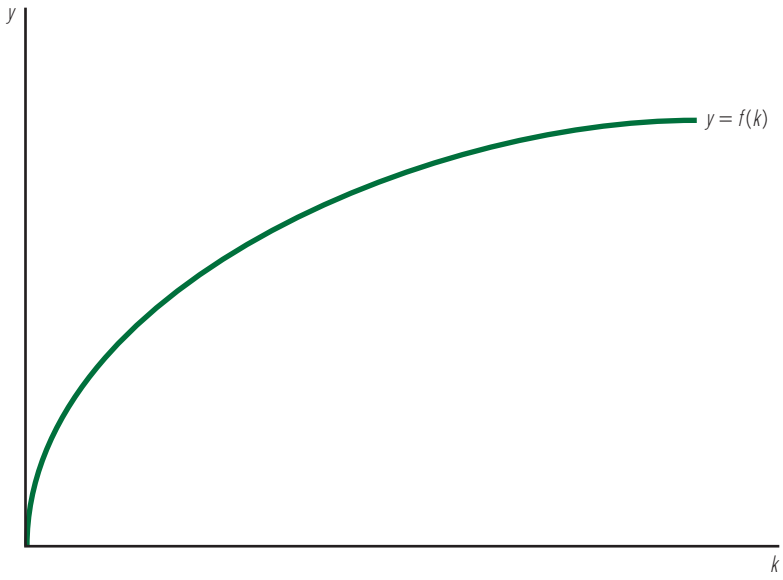


FIGURE 4–3 The Production Function in the Solow Growth Model

The neoclassical production function in the Solow model displays diminishing returns to capital so that each additional increment in capital per worker (k) is associated with smaller increases in output per worker (y).

$$Y/L = F(K/L, 1) \qquad [4–12]$$

The equation shows that output per worker is a function of capital per worker.[14] If we use lower-case letters to represent quantities in per-worker terms, then y is output per worker (that is, $y = Y/L$) and k is capital per worker ($k = K/L$). This gives us the first equation of the Solow model, in which the production function can be written simply as

$$y = f(k) \qquad [4–13]$$

Solow's model assumes a production function with the familiar property of **diminishing returns to capital**. With a fixed labor supply, giving workers an initial amount of machinery to work with results in large gains in output. But as these workers are given more and more machinery, the addition to output from each new machine gets smaller and smaller. An aggregate production function with this property is shown in Figure 4–3. The horizontal axis represents capital per worker (k), and the vertical axis shows output per worker (y). The slope of the curve declines as the capital stock increases, reflecting the assumption of the diminishing marginal product of capital.

[14]We can divide both sides by L because the Solow model (like the Harrod-Domar model) assumes the production function exhibits constant returns to scale and has the property that $wY = F(wK, wL)$. To express the Solow model in per-worker terms, we let $w = 1/L$.

Each movement to the right on the horizontal axis yields a smaller and smaller increase in output per worker. (By comparison, the fixed-coefficient production function assumed in the Harrod-Domar model would be a straight line drawn through the origin.)

The first equation of the Solow model tells us that capital per worker is fundamental to the growth process. In turn, the second equation focuses on the determinants of changes in capital per worker. This second equation can be derived from equation 4–6[15] and shows that capital accumulation depends on saving, the growth rate of the labor force, and depreciation:

$$\Delta k = sy - (n + d)k \qquad [4\text{–}14]$$

This is a very important equation, so we should understand exactly what it means. It states that the change in capital per worker (Δk) is determined by three things:

- *The Δk is positively related to saving per worker*. Because s is the saving rate and y is income (or output) per worker, the term sy is equal to saving per worker. As saving per worker increases, so does investment per worker, and the capital stock per worker (k) grows.
- *The Δk is negatively related to population growth*. This is shown by the term $-nk$. Each year, because of growth in the population and labor force, there are nL new workers. If there were no new investment, the increase in the labor force would mean that capital *per worker* (k) falls. Equation 4–14 states that capital per worker falls by exactly nk.
- *Depreciation erodes the capital stock*. Each year, the amount of capital per worker falls by the amount $-dk$ simply because of depreciation.

Therefore, saving (and investment) adds to capital per worker, whereas labor force growth and depreciation reduce capital per worker. When saving per capita, sy,

[15]To derive equation 4–14, we begin by dividing both sides of equation 4–6 by K so that

$$\Delta K/K = sY/K - d$$

We then focus on the capital per worker ratio, $k = K/L$. The growth rate of k is equal to the growth rate of K minus the growth rate of L:

$$\Delta k/k = \Delta K/K - \Delta L/L$$

With a little rearranging of terms, this equation can be written as $\Delta K/K = \Delta k/k + \Delta L/L$. We earlier assumed that both the population and the labor force were growing at rate n, so $\Delta L/L = n$. By substitution we obtain

$$\Delta K/K = \Delta k/k + n$$

Note that in both the first equation of this footnote and the equation just given, the left-hand side is equal to $\Delta K/K$. This implies that the right-hand sides of these two equations are equal to each other, as follows:

$$\Delta k/k + n = sY/K - d$$

By subtracting n from both sides and multiplying through by k, we find that

$$\Delta k = sy - nk - dk \quad \text{or} \quad \Delta k = sy - (n + d)k$$

is larger than the amount of new capital needed to compensate for labor force growth and depreciation, $(n + d)k$, then Δk is a positive number. This implies that capital per worker k increases.

The process through which the economy increases the amount of capital per worker, k, is called **capital deepening**. Economies in which workers have access to more machines, computers, trucks, and other equipment have a deeper capital base than economies with less machinery, and these economies are able to produce more output per worker.

In some economies, however, the amount of saving is just enough to provide the same amount of capital to new workers and compensate for depreciation. An increase in the capital stock that just keeps pace with the expanding labor force and depreciation is called **capital widening** (referring to a widening of both the total amount of capital and the size of the workforce). Capital widening occurs when sy is exactly equal to $(n + d)k$, implying no change in k. Using this terminology, equation 4–14 can be restated as saying that *capital deepening* (Δk) *is equal to saving per worker* (sy) *minus the amount needed for capital widening* $[(n + d)k]$.

A country with a high saving rate can easily deepen its capital base and rapidly expand the amount of capital per worker, thus providing the basis for growth in output. In Singapore, for example, where the saving rate has averaged more than 40 percent since the early 1980s, it is not difficult to provide capital to the growing labor force and make up for depreciation and still have plenty left over to supply existing workers with additional capital. By contrast, Kenya, with a saving rate of about 15 percent, has much less saving to spare for capital deepening after providing machines to new workers and making up for depreciation. As a result, capital per worker does not grow as quickly, and neither does output (or income) per worker. Partly because of this large difference in saving rates, output per person in Singapore grew by an average of 4.9 percent per year between 1960 and 2009, while Kenya's growth averaged about 0.34 percent.

We can summarize the two basic equations of the Solow model as follows. The first $[y = f(k)]$ simply states that output per worker (or income per capita) depends on the amount of capital per worker. The second equation, $\Delta k = sy - (n + d)k$, says that change in capital per worker depends on saving, the population growth rate, and depreciation. Thus, as in the Harrod-Domar model, saving plays a central role in the Solow model. However, the relationship between saving and growth is not linear because of diminishing returns to capital in the production function. In addition, the Solow model introduces a role for the population growth rate and allows for substitution between capital and labor in the growth process.

Now that we are equipped with the basic model, we can proceed to analyze the effects of changes in the saving rate, population growth, and depreciation on economic output and economic growth. This is accomplished most easily by examining the model in graphical form.

THE SOLOW DIAGRAM

The diagram of the Solow model consists of three curves, shown in Figure 4–4. The first is the production function $y = f(k)$, given by equation 4–13. The second is a saving function, which is derived directly from the production function. The new curve shows saving per capita, *sy*, calculated by multiplying both sides of equation 4–13 by the saving rate, so that $sy = s \times f(k)$. Because saving is assumed to be a fixed fraction of income (with s between 0 and 1), the saving function has the same shape as the production function but is shifted downward by the factor s. The third curve is the line $(n + d)k$, which is a straight line through the origin with the slope $(n + d)$. This line represents the amount of new capital needed as a result of growth in the labor force and depreciation just to keep capital per worker (k) constant. Note that the second and third curves are representations of the two right-hand terms of equation 4–14.

The second and third curves intersect at point A, where $k = k_0$. (Note that, on the production function above the *sy* curve, $k = k_0$ corresponds to a point directly above A where $y = y_0$ on the vertical axis.) At point A, *sy* is exactly equal to $(n + d)k$, so capital per worker does not change and k remains constant. At other points along the horizontal axis, the *vertical difference* between the *sy* curve and the $(n + d)k$ line determines the *change* in capital per worker. To the left of point A (say, where $k = k_1$

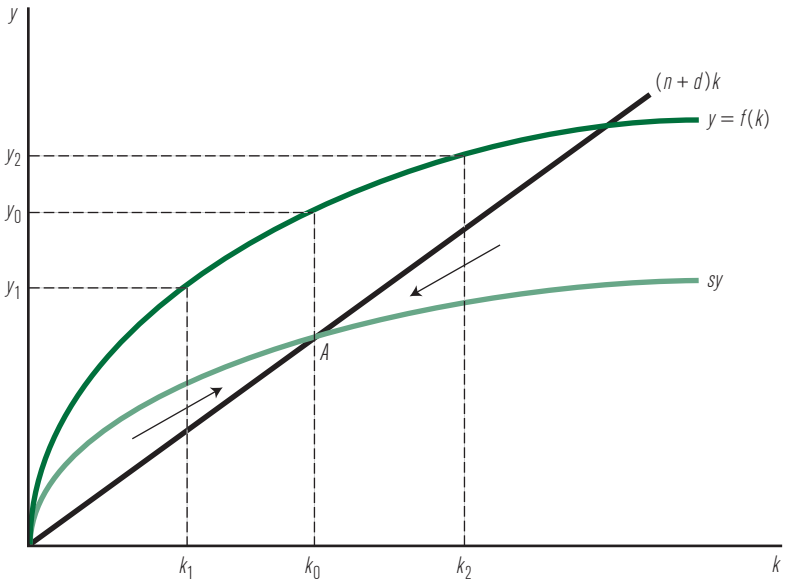


FIGURE 4–4 The Basic Solow Growth Model Diagram

In the basic Solow diagram, point A is the only place where the amount of new saving, *sy*, is exactly equal to the amount of new capital needed to compensate for growth in the workforce and depreciation $(n + d)k$. Point A is the steady state level of capital per worker and output per worker.

and on the production function $y = y_1$), the amount of saving in the economy per person (sy) is larger than the amount of saving needed to compensate for new workers and depreciation [$(n + d)k$]. As a result, the amount of capital per person (k) grows (capital deepening) and the economy shifts to the right along the horizontal axis. The economy continues to shift to the right as long as the sy curve is *above* the $(n + d)k$ curve, until eventually the economy reaches an equilibrium at point A. In terms of the production function, the shift to the right implies an increase in output per worker, y (or income per capita), from y_1 to y_0. To the right of point A (say, where $k = k_2$ and $y = y_2$), saving per capita is smaller than the amount needed for new workers and depreciation, so capital per worker falls and the economy shifts to the left along the horizontal axis. Once again, this shift continues until the economy reaches point A. The shift to the left corresponds to a decline in output per worker from y_2 to y_0.

Point A is the only place where the amount of new saving, sy, is exactly equal to the amount of new capital needed for growth in the workforce and depreciation. Therefore, at this point, the amount of capital per worker, k, remains constant. Saving per worker (on the vertical axis of the saving function) also remains constant, as does output per worker (or income per capita) on the production function, with $y = y_0$. As a result, point A is called the **steady state** of the Solow model. Output per capita at the steady state (y_0) is alternatively referred to as the **steady state**, **long run**, or **potential level of output per worker**.

It is very important to note, however, that all the values that remain constant are expressed as *per worker*. Although output per worker is constant, *total* output continues to grow at rate n, the same rate the population and workforce grow. In other words, *at the steady state GDP (*Y*) grows at the rate n, but GDP per capita (*y*) is constant (average income remains unchanged)*. Similarly, although capital per worker and saving per worker are constant at point A, total capital and total saving grow.

CHANGES IN THE SAVING RATE AND POPULATION GROWTH RATE IN THE SOLOW MODEL

Both the Solow and Harrod–Domar models put saving (and investment) at the core of the growth process. In the Harrod-Domar model, an increase in the saving rate translates directly (and linearly) into an increase in aggregate output. What is the impact of a higher saving rate in the Solow model?

As shown in Figure 4–5, increasing the saving rate from s to s' shifts the saving function sy up to $s'y$, without shifting either the production function or the capital widening line $(n + d)k$. The increase in the saving rate means that saving per worker (and investment per worker) now is greater than $(n + d)k$, so k gradually increases. The economy shifts to a new long-run equilibrium at point B. In the process, capital per worker increases from k_0 to k_3 and output per worker increases from y_0 to y_3. The aggregate economy initially grows at a rate faster than its steady-state growth

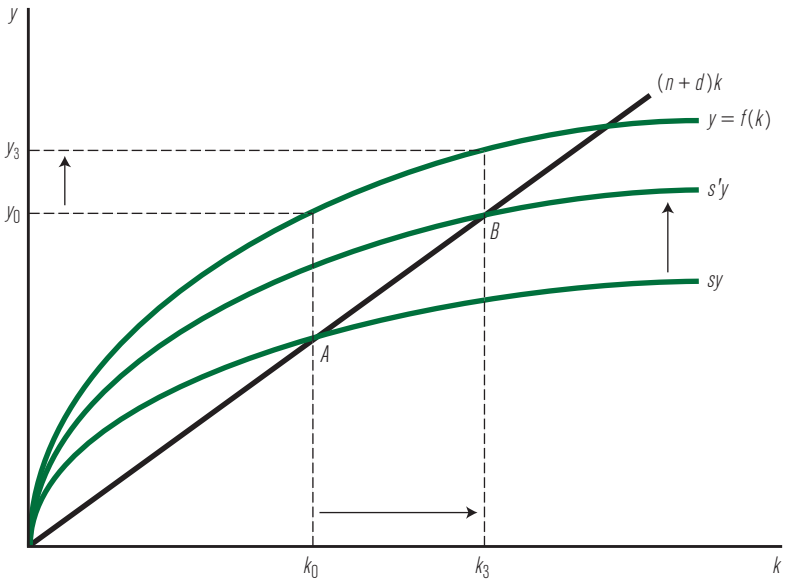


FIGURE 4–5 An Increase in the Saving Rate in the Solow Model

An increase in the saving rate from s to s' results in an upward shift in the capital deepening curve, so that capital per worker increases from k_0 to k_3.

rate of n until it reaches point B, where the long-run growth rate reverts to n. Thus, the higher saving rate leads to more investment, a permanently higher stock of capital per worker, and a permanently higher level of income (or output) per worker. In other words, the Solow model predicts that economies that save more have higher standards of living than those that save less. (The increase in per capita income, however, is smaller than for a similar increase in s in the Harrod-Domar model because the Solow model has diminishing returns in production.) Higher saving also leads to a *temporary* increase in the economic growth rate as the steady state shifts from A to B. However, the increase in the saving rate does *not* result in a permanent increase in the long-run rate of output growth, which remains at n.

The Solow diagram also can be used to evaluate the impact of a change in the population (or labor force) growth rate. An increase in the population growth rate from n to n' rotates the capital widening line to the left from $(n + d)k$ to $(n' + d)k$, as shown in Figure 4–6. The production and saving functions do not change. Because there are more workers, savings per worker (sy) becomes smaller and no longer is large enough to keep capital per worker constant. Therefore, k begins to decline, and the economy moves to a new steady state, C. More workers also means that capital per worker declines from k_0 to k_4 and saving per worker falls from sy_0 to sy_4. Output per worker (or income per capita) also declines, from y_0 to y_4. Thus, an increase in the population growth rate leads to lower average income in the Solow model. Note, however, that the new steady-state growth rate of the entire economy has increased

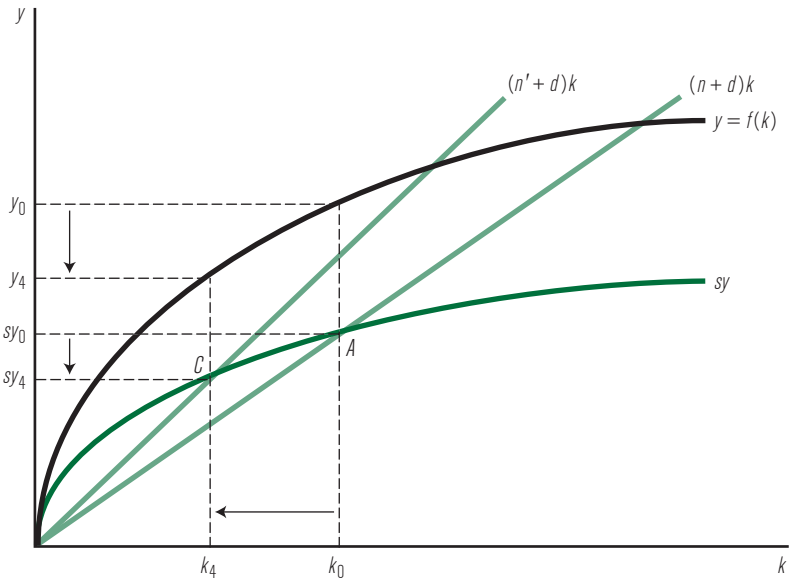


FIGURE 4–6 Changes in the Population Growth Rate in the Solow Model

An increase in the rate of population growth from n to n' causes the capital widening curve to rotate to the left. Equilibrium capital per worker drops from k_0 to k_4.

from n to n' at point C. In other words, with a higher population growth rate, Y needs to grow faster to keep y constant.[16] By contrast, a *reduction* in the population growth rate rotates the $(n + d)k$ line to the right and leads to a process of capital deepening, with an increase in both k and in the steady-state level of income per worker, y. However, the relationship between population growth and economic growth is not quite so simple, as described in Box 4–2.

The Solow growth model (as described to this point) suggests that growth rates differ across countries for two main reasons:

- Two countries with the same current level of income may experience different growth rates *if one has a higher steady-state level of income than the other*. To the extent that two countries with the same current level of income have different aggregate production functions, saving rates, population growth rates, or rates of change in productivity (described later), their steady-state income levels will differ and so will their growth rates during the transition to their respective steady states.
- Two countries with the same long-run steady-state level of income may have different growth rates *if they are in different points in the transition to*

[16]A similar exercise can be used to determine the impact of an increase in the depreciation rate, d. Such an increase results in a reduction in k and y to a lower steady-state income per capita. The subtle difference between an increase in n and an increase in d is that the latter case does not lead to a change in the long-run growth rate of Y, which remains equal to n.

BOX 4-2 POPULATION GROWTH AND ECONOMIC GROWTH

The inverse correlation between population growth and economic growth suggested by the Solow model has a lot of intuitive appeal. Countries such as Belize, Burkina Faso, Liberia, Niger, and Yemen have some of the world's fastest rates of population growth. They are also among some of the poorest nations in the world. But closer inspection of the Solow model reveals other predictions about how population growth may affect economic growth, and a further examination of empirical trends suggests a much more complex relationship.

Figure 4–6 illustrates that, in the Solow model, an increase in the rate of population growth lowers the steady-state level of income, y. However, y refers to output per *worker*, whereas the more common measure of aggregate economic welfare is output per person, y^*. These two measures of output, of course, are related to one another, as follows:

$$y^* = y \times (N/Pop)$$

where N equals the number of workers, and Pop is the total population. Differences in the level of output per capita, therefore, depend on both the amount of output per worker and the ratio of workers to total population. Growth in per capita output, similarly, depends on growth in output per worker and the growth in the worker-to-population ratio.

The Solow model suggests that more rapid population growth reduces capital deepening and hence reduces growth in output per worker. But the effect of population growth on the ratio of workers to total population is more complex. It depends on the age structure of the population. Because of rapid population growth, most developing countries have a young age structure, with a larger share of younger people than is the case in developed nations that have growing populations of more elderly people. As a result of previously higher population growth rates, many developing nations today are experiencing an increase in their ratio of workers to total population. This positive effect of a changing age structure on per capita incomes, sometimes referred to as a *demographic gift*, can play a positive and large role in determining economic growth rates.

The impact of population growth on economic growth goes beyond its effects on capital widening and a nation's age structure. In the Solow model, saving and technological change are considered exogenous. But population growth can affect these parameters as well. The net effect of population growth on economic growth is therefore an empirical matter. Econometric investigations of the impact of population growth on economic growth generally show no systematic relationship (see Chapter 7). Population growth influences many aspects of economic growth and development, not only those described by the Solow model.

the steady state. For example, consider two countries that are identical in every way except that one has a higher saving rate than the other and, so, initially has a higher steady-state level of income. At the steady states, the country with the higher saving rate has a higher level of output per worker, but both are growing at the rate n. If the country with the lower saving rate suddenly increases its saving to match the other country, its growth rate will be higher than the other country until it catches up at the new steady state. Thus, even though everything is identical in the two countries, their growth rates may differ during the transition to the steady state, which may take many years.

TECHNOLOGICAL CHANGE IN THE SOLOW MODEL

The Solow model, as described to this point, is a powerful tool for analyzing the interrelationships between saving, investment, population growth, output, and economic growth. However, the unsettling conclusion of the basic model is that, once the economy reaches its long-run potential level of income, economic growth simply matches population growth, with no chance for sustained increases in average income. How can the model explain the historical fact reported in Chapter 2 that many of the world's countries have seen steady growth in average incomes since 1820? Solow's answer was technological change.[17] According to this idea, a key reason why France, Germany, the United Kingdom, the United States, and other high-income countries have been able to sustain growth in per capita income over very long periods of time is that technological progress has allowed output per worker to continue to grow.

Technological progress is a key driver, but not the only driver, of productivity growth. Historically, most technological innovation has originated in today's developed countries, where technological progress has played a central role in explaining productivity growth. Technological progress is also important for today's developing countries; yet, most such progress in developing countries is adopted and adapted from developed countries. In developing countries, the productivity-enhancing effects of technological innovation have played a smaller role in driving productivity growth than in developed countries. In the former, productivity growth has also resulted from improvements in physical infrastructure, increased education of the labor force, and improvements in regulatory environments and incentives. Absent these (and related) factors, technology adoption in developing countries tends to be more limited and less effective in enhancing factor productivity. With these important caveats, we proceed to introduce productivity growth into the Solow model using technological progress as a shorthand.

[17]See Solow, "Technical Change and the Aggregate Production Function." For an early discussion about the relationship between capital accumulation and technological progress, see Joan Robinson, *Essays in the Theory of Economic Growth* (London: Macmillan, 1962).

To incorporate an economy's ability to produce more output with the same amount of capital and labor, we slightly modify the original production function and introduce a variable, T, to represent technological progress, as follows:

$$Y = F(K, T \times L) \qquad [4\text{–}15]$$

In this specification, technology is introduced in such a way that it directly enhances the input of labor, as shown by the specification in which L is multiplied by T. This type of technological change is referred to as *labor augmenting*.[18] As technology improves (T rises), the efficiency and productivity of labor increases because the same amount of labor can now produce more output. Increases in T can result from improvements in technology in the scientific sense (new inventions and processes) or in terms of **human capital**, such as improvements in the health, education, or skills of the workforce.[19]

The combined term $T \times L$ is sometimes referred to as the amount of **effective units of labor**. The expression $T \times L$ measures both the amount of labor and its efficiency in the production process. An increase in either T or L increases the amount of effective labor and therefore increases aggregate production. For example, an insurance sales office can increase its effective workforce by either adding new workers or giving each worker a faster computer or better cell phone. An increase in T differs from an increase in L, however, because the rise in aggregate income from new technology does not need to be shared with additional workers. Therefore, *technological change and productivity growth more broadly allow output (and income) per worker to increase*.

The usual assumption is that technology improves at a constant rate, which we denote by the Greek letter theta (θ), so that $\Delta T/T = \theta$. If technology grows at 1 percent per year, then each worker becomes 1 percent more productive each year. With the workforce growing at n, growth in the effective supply of labor is equal to $n + \theta$. If the workforce (and population) grows by 2 percent per year and technology grows by 1 percent per year, the effective supply of labor increases by 3 percent per year.

To show technological change in the Solow diagram, we need to modify our notation. Whereas earlier we expressed y and k in terms of output and capital *per worker*, we now need to express these variables in terms of output and capital *per effective worker*. The change is straightforward. Instead of dividing Y and K by L as

[18]Two other possibilities are *capital-augmenting technological change* [$Y = F(T \times K, L)$], which enhances capital inputs, and *Hicks-neutral technological change* [$Y = F(T \times K, T \times L)$], which enhances both capital and labor input. For our purposes, the specific way in which technology is introduced does not affect the basic conclusions of the model.

[19]Keep in mind, however, that, while these two broad categories of improvements in technology have similar general effects in this aggregate model, their true effects are somewhat different in the real world. Technological change in the mechanical sense or from the spread of a new idea can be shared widely across the workforce and considered a public good. Improvements in human capital, by contrast, are specific to individual workers and are not necessarily widely shared. However, both have the effect of augmenting the supply of labor and increasing total output.

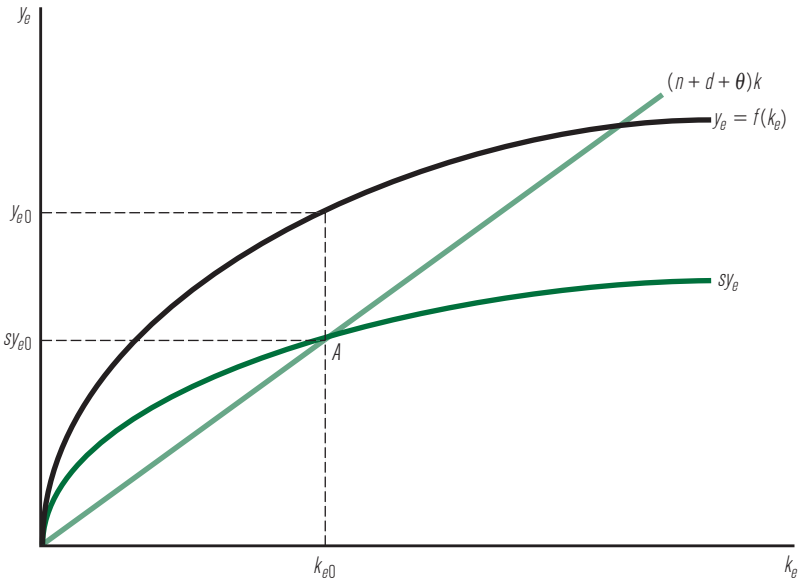


FIGURE 4–7 The Solow Model with Technical Change

In the Solow model with technical change, the equilibrium level of effective capital per worker (K_{e0}) is determined by point *A*, the intersection of the effective capital widening curve ($n + d + \theta$) and the effective saving curve (sy_e).

previously (to obtain y and k), we now divide each by $(T \times L)$. Thus **output per effective worker** (y_e) is defined as $y_e = Y/(T \times L)$. Similarly, **capital per effective worker** (k_e) is defined as $k_e = K/(T \times L)$.[20]

With these changes, the production function can be written as $y_e = f(k_e)$ and saving per effective worker expressed as sy_e. With effective labor now growing at the rate $n + \theta$, the capital accumulation equation (4–14) changes to

$$\Delta k_e = sy_e - (n + d + \theta)k_e \qquad [4\text{–}16]$$

The new term $(n + d + \theta)k_e$ is larger than the original $(n + d)k$, indicating that more capital is needed to keep capital *per effective worker* constant.

These changes are shown in Figure 4–7, which looks very similar to the basic Solow diagram, with only a slight change in notation. There still is one steady-state point, at which saving per effective worker is just equal to the amount of new capital needed to compensate for changes in the size of the workforce, depreciation, and technological change in order to keep capital per effective worker constant.

One change, however, is very important. At the steady state, output per *effective* worker is constant, rather than output per worker. *Total output now grows at the rate $n + \theta$, so that output per actual worker (or income per person) increases at rate θ.*

[20]Note that this is consistent with the earlier notation. If there is no technological change (our earlier assumption), so that $T = 1$ (and remains unchanged), then $y_e = y$ and $k_e = k$.

With the introduction of technology, the model now incorporates the possibility of an economy experiencing sustained growth in per capita income at rate θ. This mechanism provides a plausible explanation for why the industrialized countries never seem to reach a steady state with constant output per worker but instead historically have recorded growth in output per worker of between 1 and 2 percent per year.

STRENGTHS AND WEAKNESSES OF THE SOLOW FRAMEWORK

Although the Solow model is more complex than the Harrod-Domar framework, it is a more powerful tool for understanding the growth process. By replacing the fixed coefficients production function with a neoclassical one, the model provides more reasonable flexibility of factor proportions in the production process. Like the Harrod-Domar framework, it emphasizes the important role of factor accumulation and saving, but its assumption of diminishing marginal product of capital provides more realism and accuracy over time. It departs significantly from the Harrod-Domar framework in distinguishing the current level of income per worker from the long-run steady-state level and focuses attention on the transition path to that steady state. The model provides powerful insights into the relationship between saving, investment, population growth, and technological change on the steady-state level of output per worker. The Solow model does a much better job, albeit far from perfect, of describing real world outcomes than the Harrod-Domar model.

A particularly important contribution of the model is the simple yet powerful insights it provides into the role of technological change and productivity growth in the growth process. For policy makers, key questions then become how to best acquire new technologies and how to magnify their potential contributions to productivity growth through complementary investments and policies. For most low-income countries, while some domestic innovation is possible, for many industries it is probably most cost-effective for entrepreneurs to acquire the bulk of their new technologies from other countries (one of the benefits of "globalization") and adapt them to local circumstances. The willingness of entrepreneurs to make such investments, however, may depend on a number of factors that government can influence. Investments in education (for example, improvements in the *quality* of the labor force) may enable firms in developing countries to make better use of imported technologies. Investments in physical infrastructure might motivate technology adoption by better connecting firms to markets. Improved public institutions to protect property rights and extend the rule of law may also provide powerful incentives for firms to invest in improving their own productivity. However, the model's focus on the role of factor accumulation and productivity (including technology) as the proximate determinants of the steady state raises a new set of questions that the model does not answer.

The model's most troubling limitation is that Solow specified productivity growth as exogenous (that is, determined independently of all the variables and

parameters specified in the model). He did not spell out exactly how it takes place or how the growth process itself might affect it. In this sense, productivity growth has been called "manna from heaven" in the Solow model. This is an appropriate abstraction for the purpose of explaining the theoretical role of technological change. In practice, policy makers in developing countries need to know the sources of this manna.

What are the more fundamental determinants of factor accumulation and productivity that affect the steady state and the rate of economic growth? The empirical evidence in Chapter 3 suggests that the most rapidly growing developing countries share certain common characteristics: greater economic and political stability, relatively better health and education, stronger governance and institutions, more export oriented trade policies, and more favorable geography. Box 4–3 provides an estimate of the quantitative importance of these factors in East Asia's rapid growth relative to other countries. In the language of the Solow model, these characteristics operate through factor accumulation and productivity to help determine the precise shape of the production function and the steady-state level of output per worker. Changing any of these factors—say, encouraging more open trade—changes the steady-state level of output per worker and therefore the current rate of economic growth as the economy adjusts to the new steady state. Thus, the model helps us focus attention on these more fundamental influences on the steady state and the growth rate, but it does not provide a full understanding of the precise pathways through which these factors influence output and growth.

Certain characteristics of the neoclassical production function embedded in the Solow model also lead to one of the model's broadest and most problematic empirical predictions—that initially poorer countries will grow more rapidly than initially wealthier countries and eventually catch up. To understand this theoretical prediction, we begin by reviewing the neoclassical production function introduced in Chapter 3.

DIMINISHING RETURNS AND THE PRODUCTION FUNCTION

For output and income to continue to grow over time, a country must continue to attract investment and achieve productivity gains. But as the capital stock grows, the magnitude of the impact of new investment on growth may change. Most growth models are based on the assumption that the return on investment declines as the capital stock grows. We illustrate this aspect of the aggregate neoclassical production in Figure 4–8. The shape of this production function reflects the important but common assumption of diminishing returns to capital or, more precisely, a **diminishing marginal product of capital (MPK)**. This property is indicated by the gradual

BOX 4-3 EXPLAINING DIFFERENCES IN GROWTH RATES

Many recent studies have shown that the initial levels of income, openness to trade, healthy populations, effective governance, favorable geography, and high saving rates all contribute to rapid economic growth. But which are most important? One study sought to explain differences in growth during the period 1965–90 among three groups of countries: 10 East and Southeast Asian countries (in which per capita growth averaged 4.6 percent), 17 sub-Saharan African countries (in which growth averaged 0.6 percent), and 21 Latin American countries (in which growth averaged 0.7 percent).[a]

Policy variables explained much of the differences in growth rates. The East and Southeast Asian countries recorded higher government saving rates, were more open to trade, and had higher-quality government institutions. Together, the differences in these policies accounted for 1.7 percentage points of the 4.0 percentage point difference between the East and Southeast Asian and sub-Saharan African growth rates, and 1.8 percentage points of the difference between East and Southeast Asia and Latin America. Openness to trade stood out as the single most important policy choice affecting these growth rates.

Initial levels of income also were important, as the Solow model predicts. Because the Latin American countries had higher average income (and therefore greater output per worker) than the East Asian countries in 1965, the Solow model would predict somewhat slower growth in Latin America. Sure enough, this study estimates that Latin America's higher initial income slowed its growth rate by 1.2 percentage points relative to East and Southeast Asia, after controlling for other factors. By contrast, the sub-Saharan African countries had lower average initial income, indicating that (all else being equal) these countries could have grown 1.0 percentage point faster than the East and Southeast Asian countries, rather than the actual outcome of 4.0 percentage points slower. This suggests that the other factors had to account for a full 5.0 percentage point difference in growth rates between East and Southeast Asia and sub-Saharan Africa.

Initial levels of health, as indicated by life expectancy at birth, were a major factor contributing to sub-Saharan Africa's slow growth. Life expectancy at birth averaged 41 years in sub-Saharan Africa in 1965, compared to 55 years in East

[a]Steven Radelet, Jeffrey Sachs, and Jong-Wha Lee, "The Determinants and Prospects for Economic Growth in Asia," *International Economic Journal* 15, no. 3 (Fall 2001), 1–30. These results are summarized in the Asian Development Bank's study *Emerging Asia: Changes and Challenges* (Manila: Asian Development Bank, 1997), 79–82.

and Southeast Asia. The study estimates that this reduced sub-Saharan Africa's growth rate by 1.3 percentage points relative to East and Southeast Asia. By contrast, because average life expectancy in Latin America in 1965 was almost the same as in East and Southeast Asia, health explains little of the difference in growth between these regions.

Favorable geography helped East and Southeast Asia grow faster. The combination of fewer landlocked countries, longer average coastline, fewer countries located in the deep tropics, and less dependence on natural resource exports all favored Asia. Taken together, these factors accounted for 1.0 percentage point of East and Southeast Asia's rapid growth compared to sub-Saharan Africa, and 0.6 percentage points relative to Latin America. Differences in initial levels of education and the changing demographic structure of the population accounted for the remaining differences in growth rates across these regions.

Of course, this simple accounting framework does not fully explain the complex relationships that underlie economic growth. Each of the variables in the study captures a range of other factors that affect growth rates. For example, differences in government saving rates probably reflect differences in fiscal policy, inflation rates, political stability, and many other factors. Because of lack of sufficient data, the analysis omits several factors (such as environmental degradation) that may be important. And it certainly does not begin to explain why different policy choices were made in different countries. As a result, studies like these should be seen as a first step to understanding growth, rather than as a precise explanation for the many complex differences across countries.

flattening (or declining slope) of the curve as capital per worker grows. It is important, however, to bear in mind the distinction between diminishing returns to individual factors of production holding constant the use of all other factors and the **returns to scale** of the entire production function. The production functions illustrated in Figure 4–8 exhibit diminishing returns to capital per worker but (by assumption) still have constant returns to scale. This latter characteristic means that if producers doubled all inputs (rather than just one input) then total output would also double.

Looking at the production function in Figure 4–8, we can see that at low levels of capital per worker (such as point *a*), new investment leads to relatively large increases in output per worker. But at higher levels of capital per worker (such as point *b*), the same amount of new investment leads to a smaller increment in output. Each addition of a unit of capital per worker (moving to the right along the *x*-axis) yields smaller and smaller increases in output per worker. More generally, giving the same number of workers more and more machinery yields smaller and smaller additions to output.

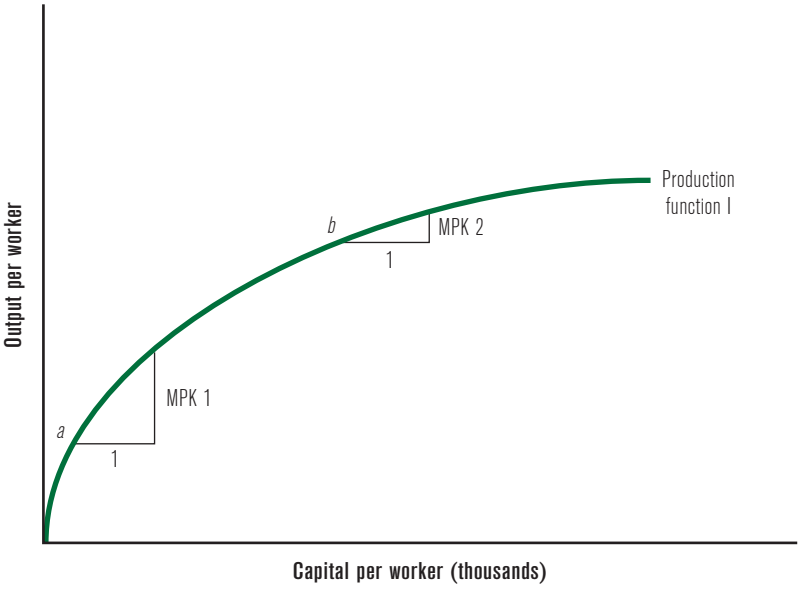


FIGURE 4–8 Diminishing Marginal Product of Capital

Along production function I, an addition of 1 unit of capital per worker at point *a* yields a much larger increase in output per worker than the same investment at point *b*. The incremental output produced by adding an additional unit of capital is called the marginal product of capital (MPK).

The assumption of a diminishing marginal product of capital has many implications, but three are particularly important for developing countries. Consider countries located toward the left on the *x*-axis of Figure 4–8, such as at point *a*. These countries have both relatively small amounts of capital per worker and low levels of output per worker. *The latter means, by definition, that these countries are relatively poor.* By contrast, countries toward the right have both higher levels of capital per worker and more output per worker, the latter implying that they are relatively rich. In general, low-income countries tend to have much less capital per worker than do richer countries. Therefore, if all else is equal between the two countries—a crucial qualifier—new investment in a poor country will tend to have a much larger impact on output than the same investment in a rich country. The three key implications are as follows:

- If all else is equal, poor countries have the *potential* to grow more rapidly than do rich countries. In Figure 4–8, a country located at point *a* has the potential to grow more rapidly than does a country at point *b* because the same investment will lead to a larger increase in output.
- As countries become richer (and capital stocks become larger), growth rates tend to slow. In other words, as a country moves along the production

function from point *a* to point *b* over a long period of time, its growth rate tends to decline.

- Because poor countries have the *potential* to grow faster than do rich countries, they can catch up and close the gap in relative income. To the extent this happens, income levels between rich and poor countries would converge over time.

These are very powerful implications. It is important to recognize that they rest on the assumption that all else is equal between the two countries, in particular that countries have the same rates of savings, population growth, and depreciation, and hence the same steady-state level of capital and income per capita. As this interpretation of the Solow model is not conditional on countries differing from one another in these key parameters, it is known as **unconditional convergence**. For all else to be equal, both countries also have to be operating along the same production function, have access to the same technology. If they are not, the predictions for rich and poor countries do not necessarily hold. For instance, a given poor country might actually be operating along different, and perhaps much flatter, production function than the one shown in Figure 4–8 if it does not have access to the same technology as reflected in the production function in Figure 4–8. In that case, each new investment (at a given level of the capital stock) would produce less output on the margin than that same additional investment would produce on the steeper production function. In that case, the poor country might not be expected to grow faster than the rich country and may never catch up. The phrase *ceteris paribus* (all else being equal), which is much used and often overlooked in economics, is of great importance in the convergence debate.

THE CONVERGENCE DEBATE

If it were true that poorer countries could grow fast while richer countries experience slower growth, poorer countries (at least those in which all else is equal) could begin to catch up and see their income levels begin to converge with the rich countries. Has this actually happened?

The short answer is that it has for some countries but not for most. Consider the example of Japan. In the 1960s, Japan's income per capita was only about 35 percent of average U.S. income, and it had a much smaller capital stock, giving it the potential for very rapid growth. (We use the United States as the benchmark for convergence because it has among the highest per capita incomes in the world and is usually considered the global technological leader.) Indeed, Japan's GDP growth rate exceeded 9 percent during the 1960s. By the time Japan had reached 70 percent of U.S. per capita income in the late 1970s, its GDP growth rate had slowed to about 4 percent. As its income continued to grow, its growth rate fell further, and growth was very slow after Japan reached about 85 percent of U.S. income in the early 1990s. Japan's experience illustrates the preceding three points very well: (1) When it was relatively poor,

it could grow fast; (2) as its income increased, its growth rate declined; and (3) as a result, its income converged significantly toward U.S. income. People who boldly predicted in the 1960s and 1970s that Japan could grow at 7 to 9 percent per year indefinitely—and many people did—ignored the impact of diminishing returns of capital on long-term growth rates.

Japan is not the only country whose income has converged with the world leaders since 1960. Look again at the group of rapidly growing countries shown in Table 3–1. All these countries were relatively poor in 1960, and all grew by an average of between 3 and 6 percent per capita for nearly 50 years. Rich countries cannot grow that fast over a period of many years (in the absence of a continuous infusion of new technology), but poor countries can because they start with low levels of capital.

However, being poor and having low levels of capital per worker by no means guarantees rapid growth. As the upper sections of Table 3–1 show, many low-income countries recorded low growth. Not only did these countries not catch up but they fell further behind and their incomes diverged even more from the world leaders. The point is that low-income countries have the potential for rapid growth, *if* they can attract new investment and *if* that new investment actually pays off with a large increment in output.

Looking beyond the experience of a few individual countries, is there a general tendency for poor countries to grow faster and catch up with the richer countries? Broadly across all countries, the short answer is no. Figure 4–9 shows the initial level of per capita income in 1960 and subsequent rates of growth from 1960 to 2009 for

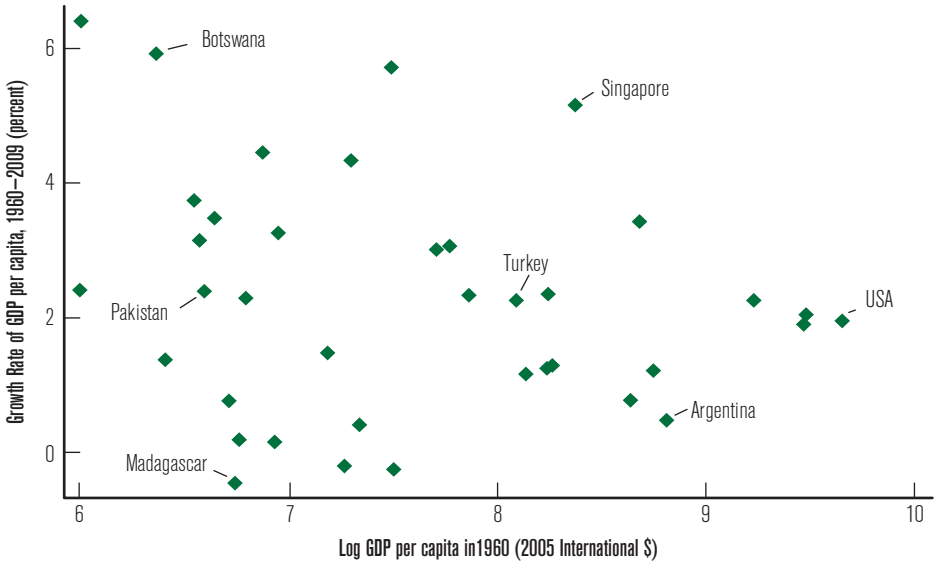


FIGURE 4–9 Gross Domestic Product (GDP) Growth, Unconditional

See Table 3–1 for data.

Source: Penn World Tables 7, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed June 2011.

the 38 countries listed in Table 3–1. If it were true that poor countries were growing faster than rich countries, the graph would show a clear downward slope from left to right. Poor countries would record a high rate of growth (and appear in the upper left part of the figure) and rich countries would display a slower growth rate (and be in the bottom right). But there is no clear pattern evident in the figure. For example, income per capita in 1960 was approximately the same in Botswana, Pakistan, and Madagascar; yet while Botswana subsequently grew at nearly 6 percent per year, Pakistan's subsequent growth was only 2.3 percent per year and Madagascar's income per capita actually declined. Similarly, Pakistan, Turkey, and the United States all grew at roughly the same rate from 1960 to 2009, yet they began the period with quite different levels of income. The only part that seems accurate is that almost all the rich countries display relatively slow growth rates, as expected. Such results have been documented in many studies using larger samples and more sophisticated statistical techniques. The empirical fact is clear: There has been no *general* tendency for poor countries to catch up to the world leaders. If anything, the opposite has been true. As we saw in Figure 2–1, for the last two centuries, the gap between the richest and the poorest regions of the world has grown, implying a divergence of incomes for these countries.

However, the simple graph in Figure 4–9 does not really do justice to the predictions of convergence, which are based on the critical assumption that all else is equal across countries. This assumption clearly is not true for all countries in the world. Instead, if convergence were to occur, we should expect to find it among countries that share some broad key characteristics, such as a similar underlying production function and similar rates of saving, population growth, depreciation, and technology growth. Some of the poor countries in Figure 4–9, for example, have low saving rates or very little growth in technology compared with other countries and, therefore, have much less potential for rapid growth. To see if the convergence predictions of the Solow model hold under these stricter conditions, we have to dig a little deeper.

Figure 4–9 demonstrates the lack of unconditional convergence, *unconditional* in the sense that it imposes the assumption that all countries share the same key parameters (depreciation, savings, and population growth rates in particular). Digging deeper into the Solow model's predictions, however, we consider the implications for convergence of allowing each country to take on its own individual combination of these key parameters. This allows each country to have its own individual steady-state level of capital and income per capita. Once we allow that, the question is whether initially poorer countries grow faster as they approach their own steady state. Figure 4–10 applies statistical techniques to relax the previous assumption that all countries have the same rates of population growth. Looking at the data from the same countries as in Figure 4–9, but conditional on the countries' population growth rates, we begin to see the predicted negative association between initial income levels and subsequent growth rates. This is known as **conditional convergence**.



FIGURE 4–10 Gross Domestic Product (GDP) Growth, Conditional on Population Growth

See Table 3–1 for data.

Sources: Penn World Tables 7, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed June 2011; World Bank, "World Development Indicators," http://databank.worldbank.org.

Another approach is to pick a group of countries from all income levels that are similar in their policy choices, geographic characteristics, or some other variable. Economists Jeffrey Sachs and Andrew Warner, for example, examined the evidence for convergence among all countries that had been consistently open to world trade since 1965.[21] Open economies, as discussed in Chapter 3, are similar in that they have similar (global) markets for their products, purchase their inputs on world markets, and can acquire new technology relatively quickly from other open economies through imports of new machinery and their connections to global production networks. What if we include countries' openness to trade among the conditioning variables? Figure 4–11 conditions GDP growth on countries' rates of savings and population growth in addition to the proportion of time between 1960 and 2009 that they were open to trade. Clearly, these additional conditioning variables sharpen the inverse correlation between initial income and subsequent growth as indicated by the tighter fit around the trend line.[22] Initially wealthier countries may still grow faster in absolute terms than initially poorer countries if the former are converging toward

[21]Jeffrey D. Sachs and Andrew Warner, "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity*, 26, no. 1 (1995), 1–118.

[22]Tracing the countries highlighted in Figure 4–9 through Figures 4–10 and 4–11, we note that, despite the additional conditioning variables, Botswana and Singapore consistently grew faster than predicted for their initial incomes (indicated by their positions above the trend lines), whereas Madagascar and Argentina consistently grew more slowly than predicted.

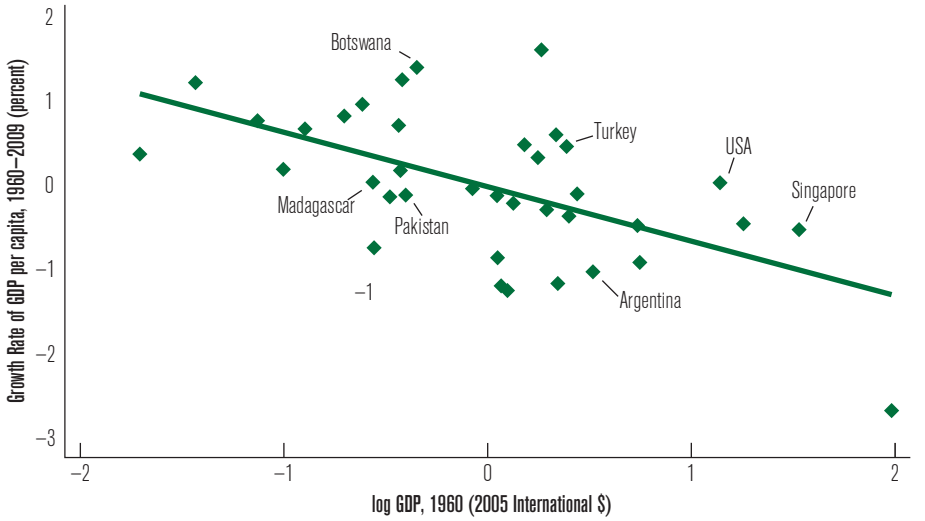


FIGURE 4–11 Gross Domestic Product (GDP) Growth, Conditional on Openness, Savings, and Population Growth

See Table 3–1 for data.

Sources: Penn World Tables 7, http://pwt.econ.upenn.edu/php_site/pwt_index.php, accessed June 2011; World Bank, "World Development Indicators," http://databank.worldbank.org.

much higher steady-state levels of capital per capita. That is why we do not observe unconditional convergence across rich and poor countries' levels of income. On the other hand, if each country is allowed to have its own steady-state level, we do see faster growth among initially poorer countries, as predicted by the Solow model and its implication of conditional convergence.

BEYOND SOLOW: NEW APPROACHES TO GROWTH

A new generation of models takes off where Solow left off, by moving beyond the assumptions of an exogenously fixed saving rate, growth rate of the labor supply, workforce skill level, and pace of technological change. In reality, the values of these parameters are not just given but are determined partially by government policies, economic structure, and the pace of growth itself. Economists have begun to develop more-sophisticated models in which one or more of these variables is determined within the model (that is, these variables become endogenous to the model).[23]

[23]The seminal contributions to the new growth theory are Paul Romer, "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94 (October 1986), 1002–37; Robert Lucas, "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22 (January 1988), 3–42; and Paul Romer, "Endogenous Technological Change," *Journal of Political Economy* 98 (October 1990), S71–S102.

These models depart from the Solow framework by assuming that the national economy is subject to **increasing returns to scale**, rather than constant returns to scale. A doubling of capital, labor, and other factors of production leads to *more* than a doubling of output. To the extent this occurs, the impact of investment on both physical capital and human capital would be larger than suggested by Solow.

How can a doubling of capital and labor lead to more than a doubling of output? Consider investments in research or education that not only have a positive effect on the firm or the individual making the investment but also have a positive "spillover" effect on others in the economy. This beneficial effect on others, called a **positive externality**, results in a larger impact from the investment on the entire economy. The benefits from Henry Ford's development of the production line system, for example, were certainly large for the Ford Motor Company, but they were even larger for the economy as a whole because knowledge of this new technique soon spilled over to other firms that could benefit from Ford's new approach.

Similarly, investments in research and development (R&D) lead to new knowledge that accrues not only to those that make the investment but to others who eventually gain access to the knowledge. The gain from education is determined not just by how much a scientist's or manager's productivity is raised by investment in his or her own education. If many scientists and managers invest in their own education, there then will be many educated people who will learn from each other, increasing the benefits from education. An isolated scientist working alone is not as productive as one who can interact with dozens of well-educated colleagues. This interaction constitutes the externality. In the context of the "sources of growth" analysis introduced in Chapter 3, such externalities suggest that the measured contribution of physical and human capital to growth may be larger than that captured by the Solow framework. Among other implications, this outcome could account for a significant portion of the residual in the Solow accounting framework, meaning that actual TFP growth is smaller than many studies have suggested.

Another important implication is that economies with increasing returns to scale do not necessarily reach a steady-state level of income as in the Solow framework. When the externalities from new investment are large, diminishing returns to capital do not necessarily set in, so growth rates do not slow, and the economy does not necessarily reach a steady state. An increase in the saving rate can lead to a *permanent* increase in the rate of economic growth. These models can explain continued per capita growth in many countries without relying on exogenous technological change. Moreover, they do not necessarily lead to the conclusion that poor countries will grow faster than rich countries because growth does not necessarily slow as incomes rise, so there is no expectation of convergence of incomes. Initial disparities in income can remain, or even enlarge, if richer countries make investments that encompass larger externalities.

Because growth can perpetuate in these models without relying on an assumption of exogenous technological change, they often are referred to as **endogenous**

growth models. They are potentially important for explaining continued growth in industrialized countries that never reach a steady state, especially those engaged in R&D of new ideas on the cutting edge of technology.

For developing countries, the new models reinforce some of the main messages of the Solow and Harrod-Domar models. Like their forerunners, these models show the importance of factor accumulation and increases in productivity in the growth process. The potential benefits from both of these sources of growth are even greater in endogenous growth models because of potential positive externalities. The core messages of saving, investing in health and education, using the factors of production as productively and efficiently as possible, and seeking out appropriate new technologies are consistent across all these models.

The applicability of endogenous growth models to developing countries is questionable, however, because many low-income countries can achieve rapid growth by adapting the technologies developed in countries with more advanced research capacities rather than making the investments in R&D themselves. For many low-income countries, the Solow model's assumptions of exogenous technological change and constant returns to scale in the aggregate production function may be more appropriate. Productivity growth in such countries may also depend as much on appropriate complementary investments and policies as on technological change itself.

SUMMARY

- Formal growth models provide a more precise mechanism to explore the contributions to economic growth of both factor accumulation and productivity gains. These models allow us to understand better the implications of changes in saving rates, population growth rates, technological change, and other related factors on output and growth.
- The Harrod-Domar model assumes a fixed-coefficients production function, which helps simplify the model but introduces strict rigidity in the mix of capital and labor needed for any level of output. In this model, growth is directly related to saving in inverse proportion to the incremental capital–output ratio.
- The Harrod-Domar model usefully emphasizes the role of saving, but at the same time overemphasizes its importance by implying that saving (and investment) is sufficient for sustained growth, which it is not. Also, the model does not directly address changes in productivity. In addition, the model's assumption of a fixed ICOR leads to increasing inaccuracy over time as the structure of production evolves and the marginal product of capital changes.

- The Solow model improves on some of the weaknesses in the Harrod-Domar framework and has become the most influential growth model in economics. The model allows for more flexibility in the mix of capital and labor in the production process and introduces the powerful concept of diminishing marginal product of capital. It allows for exploration of the impact on growth of changes in the saving rate, the population growth rate, and technological change. The model helps provide a deeper understanding of a much wider range of growth experiences than the Harrod-Domar model. Nevertheless, the model does not provide a full explanation for growth. It does not provide insights into the more fundamental causes of factor accumulation and productivity growth and, as a one-sector model, does not address the issue of resource allocation across sectors.
- The Solow model has several powerful implications, including (1) poor countries have the *potential* to grow relatively rapidly; (2) growth rates tend to slow as incomes rise; (3) across countries that share important common characteristics, the incomes of poor countries potentially can converge with those of the rich countries; and (4) acquiring new technology is central to both accelerating and sustaining economic growth.
- Some poor countries have achieved rapid growth and seen their incomes converge with those of richer countries, but many have not and have seen their incomes fall further behind. There is no evidence of unconditional convergence across countries, but there is strong evidence of *conditional* convergence, in which countries sharing certain characteristics are able to achieve rapid growth and begin to catch up with the richer countries. We find evidence of conditional convergence in the inverse relationship between initial income and subsequent growth rates, when we account for countries differing in key parameters and hence in their steady-state levels.
- In the 1990s, new theories of economic growth gained popularity. These theories sought to fill in key gaps in the neoclassical model. In particular, these models sought to explain the rate of technological progress, which Solow had identified as the determinant of long-term growth in per capita income but that remained exogenous in his model. These new growth theories are thus known as endogenous growth models.



States and Markets

Who is it that makes economic development happen? Is it all done by individuals pursuing their various goals, relying on the market to coordinate and make sense of their diverse actions? Or does the government play a central role in achieving sustained increases in per capita income and output? The debate over the proper role for government in achieving economic progress has been going on for more than two centuries. In the *Wealth of Nations* (1776) Adam Smith argued against conventional wisdom of the time, which assumed it was government that led the development process. For example, the king and queen of Spain supported Columbus's exploration of the Americas, and a royal monopoly gave the British East India Company its dominant role in development of trade with India and beyond. Smith proposed instead that economic progress was achieved by individuals who worked through the market to specialize in the production of one particular product (his example was the manufacture of pins) and then exchanged that product on the market for what else they, as producers, needed.

In England at least, Smith's view dominated development thinking for much of the country's growth in the 1800s, as the leaders of the nation promoted free trade and a minimal role for government in the economy. Elsewhere, the state often played a major role in the development of new industries. As one moved eastward from England to France, Germany, and finally Russia, the government played a steadily increasing role in nineteenth-century economic development.[1]

[1]This is the basic thesis of Alexander Gerschenkron in *Economic Backwardness in Historical Perspective* (Cambridge: Belknap Press of Harvard University Press, 1962).

After David Ricardo and later Karl Marx, however, the attention of most economists shifted to trying to understand how markets worked, and economic growth was more or less taken for granted, at least in Europe and North America.[2] This focus changed dramatically after the end of World War II, which had devastated much of Europe and Japan. The challenge was how to achieve economic recovery before wartime devastation brought to power new totalitarian regimes and further warfare. In what we now know as the developing countries of the world, former colonies were rapidly becoming independent states. These countries wanted programs that would facilitate their rise out of poverty and make it possible for them to catch up with the already industrialized nations. The United States, the one industrialized country not devastated by war, stood ready to help.

The first decades after the end of World War II witnessed a major effort to achieve modern economic growth throughout the world. There were some successes, mainly the rapid recovery of western Europe, but also many failures, and that set off a debate that continues to this day over just what countries need to do to achieve sustained economic development. At the center of this debate is the question of the proper role for government in development. In the remainder of this chapter we trace the evolution of ideas since the end of World War II about the proper role of the state in economic development. For the most part thinking about the role of the state has been a learning process because successes and failures have helped develop a more sophisticated understanding of the role of government and government-created institutions in the development process. In addition, old ideas that did not initially work well have crept back into the debate. And despite all of the development thinking since the 1950s, there are still a great many poor countries and poor people in the world. In the remainder of this chapter we describe and analyze how thinking about the role of the state has evolved through these debates. In the chapters that follow we return again and again in much greater detail to the role of government as part of the solution to the full panoply of specific development questions.

DEVELOPMENT THINKING AFTER WORLD WAR II

There were several key influences on development thinking after the end of World War II. For economists, the main ideas about growth dated back a century or more. Adam Smith had stressed markets and technological progress as the key, but David Ricardo writing three decades later stressed the importance of saving and investment and of the accumulation of capital. Karl Marx published *Das Kapital* in 1867 and also took the accumulation of capital as the central driving force of growth. Furthermore, the Harrod-Domar model (see Chapter 4) put investment and capital at the center of

[2]The notable exception in the first half of the twentieth century was Joseph Schumpeter.

growth. One influential book published in 1960 argued that the key to growth was to raise the rate of investment and capital formation to a level above 15 percent of gross domestic product (GDP), although there were numerous voices at the time that said the process was more complicated than that.[3]

If capital is at the center of the growth story, it would seem to follow that one key to growth would be to raise the rate of saving so that one could increase the rate of investment. But in poor countries, it was argued, there was a vicious circle of poverty that had to be broken first. Poor countries were too poor to save much and so did not have the resources to invest and so could not grow. The obvious solution was for rich countries to contribute some of their saving to poor countries to break this cycle. In the 1960s the idea was refined further with the development of Hollis Chenery's two-gap model.[4] There was not just a saving gap but also a foreign exchange gap, and the latter was the major constraint on development in most countries. Emphasis on the foreign exchange gap reinforced the view that the key to economic development was for rich countries to provide aid (in the form of dollars or another convertible currency) to fill both the savings and foreign exchange gaps.

This view of development was greatly reinforced at the time by the experience of the Marshall Plan in rescuing the economies (and the democracies) of western Europe and Japan. The Marshall Plan supplied $25 billion to rebuild western Europe, with immediate and dramatic results.[5] Europe recovered rapidly, with GDPs growing so quickly that these countries caught up to the level of income that they would have achieved if they had not been engulfed by war. Much the same experience occurred in Japan. In 1945 virtually all of the cities of Japan lay in ruins, and malnutrition was common throughout the country. But by the 1950s the nation was well launched on two decades of unprecedented 10 percent a year GDP growth. Large injections of capital, mostly from the United States, had done the trick, or so it seemed.

What was missing from the analysis of these success stories was the role of the supporting institutions required for economic development. Western Europe had fully developed legal systems designed for a modern industrial economy; had education systems capable of providing required training; and despite the enormous loss of life, had large numbers of people who knew how to organize a company (and a government) and who understood modern technology. Much of the physical infrastructure was gone, but many of the key institutions were embodied in people, and all that was needed was a little help to get started. When the same emphasis on foreign aid, capital, and foreign exchange was applied a decade later to developing countries,

[3]Walter W. Rostow, *The Stages of Economic Growth: A Non-Communist Manifesto* (Cambridge: Cambridge University Press, 1960).

[4]Hollis Chenery and Alan Strout, "Foreign Assistance and Economic Development," *American Economic Review*, 56, no. 4 (1966).

[5]Technically the Marshall Plan itself supplied only half of the $25 billion; the other half was contributed by the United States in the years between the end of World War II and the formal beginning of the plan in April 1948. The total amount of aid converted into today's dollars would be over US$200 billion.

many of the critical institutions were missing. *Developing* poor and underdeveloped nations proved to be a very different process from *reconstructing* already developed economies destroyed by war.

Most developing countries did not start with the Smith view that well-functioning markets were the key to successful development. To begin with, markets in developing countries were associated with capitalism, and capitalism in the minds of many was associated with colonialism and being kept poor by colonial powers. Many leaders in developing countries were looking for an alternative path, and they had plenty of support from people and events elsewhere.

In the industrialized West there was the obvious fact that markets had not functioned all that well before the war. The Great Depression of the 1930s was a market failure on an unprecedented scale. But economists and others had long been aware that all economies suffer market failures, albeit on a smaller scale (Box 5–1). The question was not whether these market failures existed but how prevalent they were and thus what (if anything) the government should do about them. Among the Fabian socialists in Britain whose ideas were reflected in the Labour Party then in power in Britain, market failures were common and called for active intervention by government in many areas of economic activity. The British Labour government of the time also supported outright government takeover of certain major businesses. The relevance of this to developing countries is that many were former British colonies whose leaders had been educated in Britain at universities where Fabian socialist views were common and widely discussed. It is interesting to note that this was true of Jawaharlal Nehru and other leaders of the Indian Congress Party, which ruled India for several decades after the country achieved independence from Britain in 1948.

Some of the most prominent theories of development economists in this early period also argued that the situation facing developing countries required broad government intervention. The **big push** idea of economist Paul Rosenstein-Rodan argued that it was difficult for one modern industry to start up in a developing country (his example was shoes) because there would not be enough people with the money to buy the products of one shoe factory.[6] Instead one needed to develop a wide range of industries all at the same time so that the workers in each factory would have the income needed to buy the products of the other factories. The only way to achieve this broad-based development of a wide range of industries was to have a government plan that could guide individual producers and guarantee that the demand would be there when they went into production. Economist Albert Hirschman countered this argument with the view that countries instead could start by emphasizing industries with strong **linkages** so the creation of one industry, machinery for example, would create demand for steel and steel would create

[6]Paul Rosenstein-Rodan, "Problems of Industrialization of Eastern and South-Eastern Europe," *Economic Journal* 53, nos. 210–11 (1943), 202–11.

BOX 5-1 MARKET FAILURE

The term *market failure* refers to situations in which markets fail to achieve efficient outcomes. *Efficiency* in this context has a precise definition, based on the concept of **Pareto efficiency**. A Pareto efficient market outcome is one in which no one can be made better off without making someone else worse off. (Note that this definition, conceived by nineteenth-century Italian economist Vilfredo Pareto, says nothing about equity.) Market failures can occur for various reasons, including the presence of externalities and the existence of monopolies, public goods, and information asymmetries. The common theme across these categories of market failure is the divergence in equilibrium between **marginal social costs** and **marginal social benefits**. Market equilibrium occurs when private actors equate their marginal costs with their marginal benefits. Under conditions of perfect competition, with no externalities and no public goods, the marginal costs and benefits faced by private actors equal the marginal social costs and benefits faced by society. Yet, imagine a situation in which a factory sits along a riverbank upstream from a fishery. That factory chooses a profit-maximizing quantity of output by equating its private marginal costs and benefits and thus releases the related quantity of effluent into the river. If the cost of the pollution is reflected in the loss of revenue to the fishery, that cost is borne by society but does not enter into the private cost–benefit calculations of the factory owner. It is an externality and leads to a divergence between marginal private costs and marginal social costs. This constitutes a market failure in which the factory overproduces both its output and pollution. On a larger scale, the overproduction of greenhouse gasses, widely seen as a cause of global warming, has been called, "the biggest market failure the world has ever seen."[a] (We return to this theme extensively in Chapter 19.)

A similar divergence between private and social marginal costs and benefits occurs in the presence of monopolies. Profit-maximizing monopolists restrict output below the level that would occur under perfect competition. In doing so, the monopolist is equating private marginal costs and benefits (the basic condition for profit maximization). Yet, reducing the quantity of product available in the monopolized market results in an equilibrium in which the marginal social benefits of an additional unit of output exceed the marginal social costs of producing that last unit, another instance of market failure. Society would be better off with additional output in this market.

The market failure associated with public goods arises from the incentive for individuals to act as free riders. Public goods have the characteristic that no

[a]N. Stern, *Stern Review on the Economics of Climate Change* (London: HM Treasury, 2006).

one can be excluded from consuming them (for example, national defense). The resulting distinction between public and private goods is that to consume the latter, each individual must reveal his or her true willingness to pay for that good, but with the former, because no one can be excluded, the incentive is for individuals to conceal their true willingness to pay for consumption. Markets thus fail by undersupplying public goods (here, too, implying a divergence between private and social marginal costs and benefits).

Information asymmetries are situations in which one party to a transaction has more information than the other party. Such situations commonly arise in insurance markets, where the insured knows more about his or her own risk profile and behavior than does the insurer. Thus, high-risk individuals will tend to drive up the price of insurance for low-risk individuals and drive them out of the market (the adverse selection problem). Similarly, having insurance might lead some to engage in riskier behavior than otherwise, thus driving up the price of insurance for those who continue to behave more cautiously (the "moral hazard" problem). In both cases, the information asymmetry creates an externality, leading to market failure.

Faced with these (and other) potential sources of market failure, the challenge for government policy makers lies in deciding if, when, and how to intervene. Potential policy interventions to correct market failures are varied and complex. They might include taxing polluters (to internalize the cost of externalities), breaking or regulating monopolies, and directly providing public goods. In practice, diagnosing market failures is rarely a straightforward task, and public intervention brings with it the risk of making the situation worse than it would have been without the intervention. Such instances are known as **government failure**. The tension between the risks of market failure and government failure lies at the heart of the debates described in this chapter.

demand for iron ore and coal (what Hirschman called **backward linkages**).[7] The big push theory implicitly assumed that solving the demand problem by exporting a new product was not a realistic option, an assumption that was not unreasonable given that Rosenstein-Rodan was writing during and shortly after World War II. Some analysts still use the term *big push,* but mostly when arguing for making a large effort, such as an unprecedented increase in foreign aid to get over one or another major

[7]Albert O. Hirschman, *The Strategy of Economic Development* (New Haven, CT: Yale University Press, 1958).

obstacle to growth. More recent uses of the big push terminology are discussed at greater length in Chapter 14.

Of comparable or greater influence than these intellectual debates was the perceived success of the model of growth pioneered by the Soviet Union.[8] This model had three basic components:

1. The proponents of the model, like Marx, emphasized the role of capital and high rates of investment achieved by limiting consumption by farmers and others.
2. The model gave priority to the development of heavy industry that would provide physical capital products, such as machinery, that were assumed not to be available unless produced domestically.[9]
3. Management of the economy rejected the use of market forces except in areas peripheral to most economic activity.

The allocation of goods was to be done by large government bureaucracies guided by a central plan drawn up in Moscow. The appeal of the Soviet approach was based on the widely held perception that it had transformed a poor and backward country into an industrial powerhouse that defeated the strongest military power in the world, Nazi Germany. Three decades later this economic model would appear to be far less formidable within the Soviet Union itself as well as elsewhere, but in the 1950s many developing countries were inspired by it. Ghana under Kwame Nkrumah, for example, the first African nation to achieve independence from its colonial ruler, focused on heavy industry during its initial growth phase. India went even further in attempting to adapt major parts of the Soviet economic model, and China, after the 1949 revolution, adopted the Soviet system in its entirety.

Finally, the attitude toward the prospects of achieving growth through reliance on foreign trade had limited appeal in most developing countries. The belief among many was that dependence on foreign trade would force these countries to be forever the source of raw material from their mines and agriculture and never industrialized countries in their own right. More important, the one region of the developing world that had begun industrial development before World War II (South America, in particular Brazil, Argentina, Columbia, and Chile) had done so by developing industry behind high protective barriers, called *import substituting industrialization* (discussed at greater length in Chapter 18). There was no belief at the time that growth could be achieved to a large degree by the export of manufactures, and there was a solid reason for this belief. During World War I and II, there were few international markets for South American manufactures because submarine war, among other

[8]Strictly speaking, the Soviet model of economic growth was patterned in part on the German wartime economic model used during World War I (1914–18).

[9]Strictly speaking, the Soviet model assumed that the economy was basically closed to foreign trade; hence exporting consumer goods or natural resources to purchase capital goods from abroad was not an option or was, at best, a very limited option.

things, prevented most normal trading relationships. In between the two wars was the Great Depression, when industrialized countries vied with each other to raise tariff barriers to protect their own domestic industry. The very high Smoot-Hawley Tariff, enacted in 1930, was the American contribution to this trade war.

By the 1950s and 1960s this emphasis on import substituting industrialization was further reinforced in South America. Industries developed behind high trade barriers were an important part of those economies and gave them substantial political influence to keep the trade barriers in place. In addition, economists such as Raul Prebisch, based at the Economic Commission for Latin America (ECLA) in Santiago, Chile, and Hans Singer, based at the United Nations Industrial Development Organization (UNIDO) provided an intellectual rationale for import substituting industrialization. The low income and price elasticities of demand for developing countries' exports of raw materials and the inability of such nations to compete against the economic head start of the industrialized nations meant that all or most of the benefits of foreign trade went to the already industrialized nations, or so Prebisch and Singer argued.

The result of these various influences is that most developing countries in the 1950s and 1960s believed that a high level of government planning and government intervention in the economy was desirable and even absolutely necessary if economic growth were to be achieved. All over the world, countries introduced one government intervention after another. Sometimes these interventions were based on strategy articulated in a national plan; at other times they had a much narrower focus, including the personal interests of the government officials introducing the measures. All of this was occurring, however, while the governments of the industrialized nations were in the process of dismantling most of their trade barriers under the auspices of the General Agreement on Tariffs and Trade (GATT), formed in 1949 and now superseded by the World Trade Organization (WTO). Economists from industrialized countries who had lived through the trade wars of the 1930s and a much wider group of political leaders felt that these trade disagreements had contributed directly to global warfare and were determined to move the industrialized world back to the open trading system that had existed before World War I. Some of these same economists argued vigorously against the protectionist trends in the developing world, but not many policy makers in developing countries were listening.

The main approach of economists working with developing countries in the 1950s and 1960s was not to try to reduce the role of government in the economy but to try to make government intervention more rational and geared to achieving well thought out development objectives. One such measure was to help countries set up planning commissions that would draw up economic development plans and provide sound economic advice to development officials in general. The Harvard Development Advisory Service (later the Harvard Institute for International Development), heavily funded by the Ford Foundation, for example, provided technical

support mostly in the form of economic analysis and training to planning commissions in Iran, Pakistan, Indonesia, Malaysia, Ghana, and other countries in the 1960s and early 1970s.[10]

By the early 1970s, therefore, a high level of intervention and a high level of government investment in the economy of developing countries was the norm, although there were exceptions. The role of industrialized countries was to provide aid mainly to governments either through bilateral or multilateral aid programs (for example, the World Bank) and to provide technical assistance to help design and manage aid projects and help more generally with government economic policy making. In countries within the sphere of influence of the Soviet Union, the full Soviet model, where the government controlled and administered most economic activity outside of agriculture, was the norm.

FUNDAMENTAL CHANGES IN THE 1970s AND 1980s

By the 1970s a reaction against the approaches of the earlier decades was building, particularly among economists and other policy makers in the bilateral and multilateral aid agencies and in the governments of the industrialized world. There was a growing disillusionment with the results of the considerable amounts of aid given out over the previous two decades. In Africa, most governments that had gained independence in the 1960s were experiencing little or no growth, and total factor productivity in the continent as a whole was negative. In Latin America import substituting industrialization had run into the limits caused by the small size of most of the domestic markets and growth had slowed in much of the region. By the 1980s (and in China by the late 1970s), interest in the Soviet model had waned because most of the countries with a Soviet-type economic system began to stagnate and in certain cases to build up unsustainable levels of international debt.

The analysis of what had gone wrong, however, had several different strains to it, and analysts and policy makers often did not agree about which one was the major source of weak performance. One strain of criticism of the policies of the 1950s and 1960s argued that too much attention had been paid to the growth rate of GDP and not enough to how the benefits of growth were distributed.

Too many of the benefits of growth, it was argued, were going to a narrow group of wealthy business people and their government and politician supporters. The share of the population trapped in poverty was not declining rapidly enough even in areas where GDP growth was relatively robust, and it was not declining at all in

[10]Edward S. Mason, *The Harvard institute for International Development and Its Antecedents* (Cambridge: Distributed by Harvard University Press, 1986).

much of the developing world.[11] Inequality also had a regional component. In many countries growth appeared to be concentrated in one part of the country, often near the capital city, while other regions stagnated. The breakup of West Pakistan and East Pakistan with the latter becoming Bangladesh in 1971, for example, was in part caused by the belief that all or most of the benefits of foreign aid and development were flowing to West Pakistan.

Economists had recognized for decades that an efficient economic system did not necessarily mean the system was equitable or that the fruits of the system were fairly distributed. Efficiency in economic theory was typically defined as a system that could not make anyone better off without making someone else worse off, a concept called **Pareto optimality**. Markets were efficient in this sense (or at least perfectly competitive markets were efficient), but that could be consistent with an economy with a concentration of income and wealth in the hands of the few. If markets could not necessarily solve the problem of inequality, it seemed to many economists that it was up to the government to correct this failure. They argued that government was best suited to address basic needs and target aid to the poor. Private charities and other nongovernmental organizations (NGOs) could play a role in these efforts, but they lacked the resources possessed by governments.

Counter to this argument for even more government intervention in the economy was a growing awareness based on research and experience that many government interventions did not work out as intended or planned. Many simply created obstacles to economic growth. Trade restrictions in Indonesia, for example, forced importers to obtain 70 different stamps of approval before a good could be brought into the country. This approach may have protected a few domestic industries, but it made it difficult for most businesses to get needed inputs in a timely manner, something that was essential if they were to compete in the international marketplace. Other interventions such as those of Kwame Nkrumah in Ghana were based on a flawed development strategy unsuited to the conditions then prevailing in the country (Box 5–2). Many governments had no economic rationale at all for their interventions or justified their actions as solving some particular market failure when the real motivation lay elsewhere. Often policies that had the stated goal of helping the poor were really designed to provide jobs or illicit income for well-off government officials. Other interventions, whether well intended or not, provided government officials with discretionary authority to issue permits, providing the opportunity for illegal payments.

For the increasing number of economists and policy makers who held these views, the clear implication was that governments of developing countries needed to

[11]One study at the time that had an important influence on thinking about redistribution, particularly in the World Bank, was the book by Hollis Chenery, Richard Jolly, Montek S. Ahluwalia, C. L. Bell, and John H. Duloy, *Redistribution with Growth: Policies to Improve Income Distribution in Developing Countries in the Context of Economic Growth* (World Bank, 1974).

BOX 5-2 GHANA AFTER INDEPENDENCE

On March 6, 1957, with great celebration, Ghana became the first sub-Saharan African country to gain its independence. Optimism was in the air: Ghana was the world's largest cocoa producer, it had the highest per capita income in sub-Saharan Africa (except South Africa), and its foreign exchange reserves were equivalent to over three years of imports. Ghana's prospects looked bright.

But it was not to be. President Kwame Nkrumah, who was brilliant in leading Ghana to independence, proved to be much less effective as an economic and political leader. In a push for rapid industrialization, Nkrumah introduced extensive state intervention in the economy, with many controlled prices, restrictions on trade and investment, and state-owned enterprises. Based on its cocoa earnings, Ghana borrowed heavily to establish a wide range of industries designed to substitute for imports and for processing many of Ghana's raw materials. But cocoa prices collapsed in the mid-1960s, and the state's heavy hand led to widespread corruption, poor investment decisions, distorted prices, and growing instability. In 1966, a group of army officers overthrew Nkrumah in what was only the first of a series of coups and countercoups over the following 17 years. The economy fluctuated wildly, with growth plunging to –6 percent in 1966, then climbing to 7 percent in 1970 as cocoa prices recovered, only to collapse again to –5 percent 2 years later and to a disastrous –14 percent in 1975. By 1983, the economy had collapsed. Cocoa production was only half as large as it once had been, inflation exceeded 120 percent, growth had been below –6 percent for 3 years in a row, and per capita income was about one-third lower than it had been 26 years before at independence.

In 1983, President Jerry Rawlings (who had taken power himself in a coup 2 years earlier as Flight Lieutenant Rawlings), with few options at his disposal, turned to the International Monetary Fund (IMF) and World Bank for financial assistance and, in return, accepted their strict conditions for changes in economic policy. Rawlings introduced Ghana's Economic Recovery Program, which in many ways was a prototype for reform programs introduced by many other developing countries under the auspices of the IMF and World Bank in the 1980s and early 1990s. The reforms were widespread: Ghana devalued its currency, freed interest rates from restrictions, and substantially reduced its budget deficit. It removed or modified a range of price controls (including raising the price paid to cocoa farmers), reduced tariffs on many imported products, and otherwise took steps to liberalize the trade regime. It partially or fully privatized dozens of enterprises over several years, including transport, the marketing of cocoa, banking, and other areas. The IMF and World Bank heralded Ghana's progress in reforms

and predicted that these changes would lay the foundation for greater macroeconomic stability, sustained economic growth, and a reduction in poverty.

The results were a clear improvement although not as dramatic as many at the time had hoped. The economy stabilized: Inflation fell (although it still remained above 20 percent), the trade deficit shrank, and foreign exchange reserves grew. Growth became far less volatile, and per capita GDP grew at 1.6 percent through the mid-1990s and at a higher rate of just over 2.9 percent from 1996 through 2010. Growth after 2004 was further boosted by the international community forgiving nearly $6 billion of Ghana's debt, although Ghana then borrowed another $750 million in 2007. The discovery, also in 2007, of major off-shore oil reserves promised further support for Ghana's economy. The first oil began to flow in 2010 and the per capita GDP growth rate in 2011 jumped to over 8 percent. Whether this oil windfall will lead to sustained higher growth or will be mismanaged with negative consequences for growth remains to be seen. The evidence as to whether this will be the case as of 2011 is mixed.

step back from meddling in the economy and better design whichever existing policies seemed to be working. The shortcomings of import substituting industrialization have been particularly well documented.[12] Even Nobel laureate Gunnar Myrdal, an early advocate of import substituting protection as a means of mobilizing underemployed labor, eventually recognized that intervention often leads to corruption, which in turn inhibits development.[13] By the 1970s and early 1980s the World Bank also began to push harder on these issues.

While these debates over the role of government were proceeding, two major developments occurred that were to have a profound impact on how the role of government in development was perceived. The first development was the rise in oil prices engineered by the Organization of the Petroleum Exporting Countries (OPEC) in 1973 and 1979. These price increases led to large amounts of money flowing into the treasuries of oil-exporting countries, expanding the role of government investment in those countries, often by a very large amount. Of equal if not greater importance over the

[12]One of the early influential articles questioning import substitution industrialization was John Sheahan, "Import Substitution and Economic Policy: A Second Review," Research Memorandum No. 50 (Williamstown: Williams College, Center for Development Economics, 1972). See also his study of import substitution in Latin America, John Sheahan, *Patterns of Development in Latin America: Poverty, Repression and Economic Strategy* (Princeton: Princeton University Press, 1987).

[13]In his *Asian Drama: An Inquiry into the Poverty of Nations,* 3 vols. (New York: Twentieth Century Fund, 1968), Gunnar Myrdal introduced the concept of the *soft state*, which was ineffective in carrying out its development goals and was typically riddled with corruption arising from these ineffective efforts.

longer run, however, was that several of the largest exporters acquired large reserves of foreign exchange that they reinvested with the cooperation of the international banks.

The enormous amount of money involved led to low interest rates on loans worldwide, and governments of developing countries were encouraged to borrow large sums to fund their development programs. Particularly in Latin America, debts to the major private international banks began to accumulate. By the mid-1980s, the situation had changed dramatically. Alarmed by rising inflation, due in part to the OPEC oil shocks, the U.S. Federal Reserve sharply contracted the U.S. money supply, causing interest rates to rise precipitously. This had two highly detrimental effects on developing nations that had borrowed abroad. First, many loans were written with variable interest rates; second, they were denominated in dollars, not local currency. When the U.S. dollar rapidly appreciated as markets responded to the policies of U.S. monetary authorities, heavily indebted nations had trouble servicing their debts as their interest payments rose and the value of their currencies against the U.S. dollar fell. The ensuing debt burden on Latin American countries forced several major countries to consider defaulting on their debts. The full nature of the debt problem in this period is discussed at length in Chapter 13; here we note that the buildup of debt first made it possible for developing country governments to finance a wide array of development projects, and then created a payment crisis that brought many of these projects to a premature end in what in Latin America has often been called the lost decade of the 1980s. The impact of this lost decade went far beyond the immediate debt problems. It called into question the whole subject of the appropriate role for government in development.

The other profound event of the 1970s really began in the 1960s, but its impact on development thinking was felt only a decade later, and its lessons are still being debated today. This event was the rise of the Four East Asian Tigers: South Korea, Taiwan, Hong Kong, and Singapore. By the late 1970s these economies had been growing at rates comparable to those of Japan for over a decade and were increasingly seen as economic success stories. The basis for their growth was a direct challenge to the view that the path to successful development rested on an emphasis on import substituting industrialization. All four built their rapid economic growth on the export of labor-intensive manufactures. There was a period of import substitution, but the four quickly moved to a reliance on exports, manufactured exports in particular. The model for the economies of South Korea and Taiwan was Japan, which had pioneered a similar strategy during the first phases of its development and again after World War II. Hong Kong and Singapore had long been trading entrepôts and possessed all of the infrastructure and institutions that trading economies required. What was added in the latter two economies was a manufacturing sector oriented toward export markets. In Hong Kong's case, it was manufacturers in Shanghai who moved their base of operation to Hong Kong after the Communist revolution in China in 1949. All Four East Asian Tigers exported first and foremost to the United States whose market, unlike in the 1930s, was wide open to foreign imports. These nations demonstrated that a development strategy based on the export of manufactures was

not only possible but superior to the import substituting strategies then prevailing in much of the developing world.

More controversial is the influence of the Four East Asian Tigers on the role of government in their economic development strategy. The facts themselves are not really in dispute. The governments of South Korea and Taiwan were actively involved in export promotion and in industrial development more generally. They devalued their currencies, subsidized exports in various ways, and promoted education and research in technology in support of industry. Government support, including temporary protection against imports, was predicated on performance, and performance by Korean and Taiwanese firms was measured by their ability to export. By any standard these governments were interventionist, but they had a clear strategy behind their intervention, which was not allowed to be diverted toward the private interests of government officials. Corruption existed in the two economies, but it had little influence on industrial and export policies. It is also the case, however, that state-led economic development became less and less viable as these economies became more sophisticated and complex. By the first decade of the twenty-first century, both South Korea and Taiwan had eliminated much of the intervention that characterized their first decades of rapid growth (Box 5–3).

The experience with the role of government in development in Hong Kong and to a large degree Singapore as well could not be more different. Hong Kong was as close to a pure laissez faire market economy as any in the world. The government of Hong Kong maintained a completely open free trade economy. In the few areas in which it did intervene, such as land policy, the government imposed policies unrelated to the main development strategy. The role of government was to maintain a stable macroeconomic environment, support education and health, and maintain security. Singapore's government was similar in most respects, but there the government did play a role in supporting particular industries and several of the key corporations (Singapore Airlines, for example) were government owned. For the most part, however, Singapore's industrial policy focused on creating a favorable climate for foreign direct investment (see Chapter 10).

Some studies have tried to press these rather different experiences of the Four East Asian Tigers into a story consistent across all the groups.[14] More common has been the selective choice of one experience over another to support a particular preconceived view of the role of government intervention (Hong Kong for those who support laissez faire markets;[15] South Korea's heavy and chemical industry drive of

[14]World Bank, *The East Asian Miracle: Economic Growth and Public Policy* (New York: Oxford University Press, 1993), ably describes many of the reasons for the success of the four economies but less successfully tries to fit the experiences of the four into a single development model that relied mostly on market forces for its success. This reasonably portrays the policies of Hong Kong and Singapore but not those of South Korea and Taiwan.

[15]See for example Milton Friedman, "The Real Lesson of Hong Kong," *National Review,* December 31, 1997.

BOX 5-3 THE DECLINING EFFECTIVENESS OF GOVERNMENT INTERVENTION IN THE MARKET: KOREA, 1960s–2010

South Korea in the 1960s and 1970s pursued an industrial policy that involved heavy intervention by the government to promote exports in general and specific industries in particular. The government provided bank loans to favored industries at below-market rates, gave these industries special access to government-controlled foreign exchange and imports, and helped build the infrastructure they required. The president of the country even met monthly with the heads of the major industries to review and help solve their problems, many of them related to intervention by government in their efforts to expand. The government saw itself as correcting for market failures, even though many of these market failures were the creation of earlier and continuing government controls that were part of its import substituting policies.

These industrial policies appear to have worked reasonably well, although the heavy industry and chemical industry drive of the 1970s remains controversial to this day. President Park Chung Hee, who led this effort, kept politics and corruption out of most of these industrial policy decisions. Because he ran an authoritarian regime supported mainly by a modern army, he felt no need to use industrial policies to pay off his political supporters. By the early 1980s, however, many Korean economists and others felt that this activist government industrial policy was causing too many economic distortions and the country should move toward a more market-oriented system. This change proved to be difficult because of promises made to private companies that had carried out the government's wishes in the past. Such companies felt that they had an implicit guarantee that they would be helped out if, by doing the government's bidding, they got into trouble. In the 1980s, they did get into trouble, and the government spent much of its time in that decade providing them with subsidies and other means of support.

By the 1990s, Korea had a democratic government led, from 1992 to 1997, by President Kim Young Sam, who had little interest in an activist industrial policy. Like his predecessors, however, he saw how these policies could be used to raise large sums of money to support his political campaigns. Increasingly the principal criterion for getting government support for a particular firm or industry depended on whether that firm contributed to the political campaigns. Such contributions existed in the 1960s and 1970s as well, but they were not allowed to influence major industrial policy decisions. By contrast, in the 1990s, industrial policy interventions were driven more by politics than by a concern for building efficient, competitive industries. Industrial policy in the earlier period had produced

such firms as POSCO, one of the largest and most efficient steel companies in the world, a firm that remained profitable right through the Asian financial crisis of 1997–98. The industrial policy in the 1990s produced such firms as the Hanbo Steel company, whose bankruptcy contributed in a significant way to Korea's involvement in the Asian financial crisis.

Finally, in the 1998–2008 period, the Korean government under two new presidents, Kim Dae Jung and then Roh Moo Hyun, began to make a major effort to get themselves out of the business of supporting particular industries and firms. The objective was to create a level playing field for all, but to do so without triggering another financial crisis caused by the collapse of the older firms that had relied so much on government support. Moving away from extensive government intervention and relying entirely on market forces, however, is not as simple as simply telling the government to stop intervening. In many cases, the government felt it had to introduce new interventions to undo the damage of previous ones. This same conflict of goals continued during the government of President Lee Myung-Bak who took office in 2009.

the 1970s for those who believe in extensive government intervention). As the experience of these four economies became better understood, however, two conclusions became increasingly clear. One was that developing countries could grow at very rapid rates with the right policies. The other was that growth could be based on exports, at least on exports of manufactures. Exports based on natural resources (discussed in Chapter 17) were another matter, but none of the four tigers had significant exports of natural resources.

The influence of the experience of the Four East Asian Tigers was particularly strong regionally, notably in China after 1978 and Vietnam after 1986, but it also changed the thinking of many economists and policy makers in South America and elsewhere in the developing world.

STRUCTURAL ADJUSTMENT, THE WASHINGTON CONSENSUS, AND THE END OF THE SOVIET MODEL

The response of the multilateral and bilateral aid agencies, particularly the World Bank, was to use their aid programs to support changes in the approach to development in countries whose economic performance was weak. By the late 1980s the list of countries struggling but largely failing to achieve rapid economic development included most of those in Africa, Latin America, and South Asia. The term used by

the World Bank for many of the changes it proposed was *structural adjustment*, a broad term that includes reforms aimed at changing the structure of the economy to be more market based, more efficient, more open to international trade, and more focused on the production of tradable commodities. The result was a wide range of reform proposals, initially somewhat ad hoc, that were often made conditional if a country wanted to receive assistance from the World Bank or other aid agencies.

A major objective of structural adjustment reforms was to reduce or remove distortions generated by government intervention. In the majority of developing countries, the list of distorted prices was long and included high and uneven tariffs, cheap interest rates on loans (which discourage saving and lead to poor loan decisions), high minimum wages that help some workers but hurt others, subsidized fuel or food, and limits on prices paid to farmers to control food prices. For every government-determined price, there are winners who receive more for their output or pay less for their inputs, and losers who experience the opposite. Price distortions most often occur when the beneficiaries of a price change are few and concentrated, and the losers are many and dispersed. A tariff raises profits for the well-connected factory owner. But even though millions of people might be paying a higher price for the tariff-protected goods, they are scattered unevenly across the country and unlikely to organize to protest the increase. Because the tariff is built into retail prices, consumers may not even be aware that government intervention caused the higher prices. The owners of the company, however, know and are happy to share their increased profits with those who helped them, thus creating political barriers to reform.

Sometimes, correcting a distortion may lead to lower prices, as in removing a tariff, but it can also lead to higher consumer prices, as in the removal of food or fuel subsidies. These subsidies are often paid for out of the budget and can be major contributors to budget deficits and inflation, or they can come from scarce tax revenues that might be better spent on schools, roads, or health clinics. However, removing such subsidies can create enormous political pressure because consumers immediately see the costs but not always the benefits. Removal of a food or fuel subsidy frequently triggers rioting that can topple governments.

Thus there is often stiff resistance to structural adjustment reforms designed to correct price distortions, either because they hurt powerful and well-connected people or because they affect many consumers. In central and eastern Europe and in Russia and China, most state enterprises saw themselves as beneficiaries of state price controls because it meant that they received key inputs at artificially low prices. To get around the resistance of large state enterprises, China created a dual price system in the 1980s by which steel going to a large enterprise in accordance with the annual plan was charged a low state-set price and all other steel was sold on the market at much higher market-determined prices. This dual system overcame much of the political resistance to price reform, but it also created opportunities for corruption by those who could use their influence to buy at the low state price and quickly resell at the high market price, corruption that contributed to the discontent that

fueled the demonstrations on Tiananmen Square in 1989. Many of the other socialist and former socialist countries, from Russia to Vietnam, therefore, opted for eliminating state-set prices altogether and using one market-determined price for most goods and services. By the latter half of the 1990s, China had largely eliminated the two-price system in favor of a unified market-price system. The gradual elimination of the dual price system resulted in part from deliberate decisions by the Chinese government to reduce the quantity of goods sold at low state-set prices. For the most part, however, the state-set prices disappeared because producers increasingly were unwilling to sell their goods at below market prices, and they found numerous ways to circumvent government instructions requiring them to do so.

There were many other elements to most structural adjustment programs, and over time what started as an ad hoc country-by-country approach gradually evolved into a comprehensive view of what successful development involved. In later years this came to be known as the Washington Consensus, but a comprehensive view of what was wrong began to appear well before that term was coined. In 1981, for example, the World Bank published an influential report, *Accelerated Development in Sub-Saharan Africa: An Agenda for Action*, better known as the Berg Report after its principal author, economist Eliot Berg. It presented a wide range of policy changes that would be needed if Africa were to reach its potential growth. The report focused heavily on agriculture because that is where most people lived and worked. The Berg Report called for an end to overvalued exchange rates then common in Africa, and an elimination of trade restrictions that favored inefficient industries over efficient farms. The expanding role of the state in managing the economy, particularly the role of state marketing boards as substitutes for private marketing mechanisms, came in for criticism. These marketing boards, such as the Ghana Cocoa Board, often extracted high taxes from farmers by setting output prices far below the true export price for the commodity. There also were recommendations for better debt and aid management and more emphasis on improving human resources through education and health measures. Similar reports followed in subsequent years but through the 1980s African development remained sluggish.

It was a seminal essay by economist John Williamson, however, that laid out clearly the view of what the international and bilateral aid agencies thought was the core set of problems that countries needed to face if they were to accelerate their development. This growing strong agreement among the International Monetary Fund (IMF), the World Bank, the U.S. government, and other key international actors focused on the steps that were necessary for Latin American countries to reform their economies, but the implications went beyond Latin America. Williamson was the first to label this broadly agreed reform agenda *the Washington Consensus.*[16] Williamson's

[16]The original article is John Williamson, "What Washington Means by Policy Reform" in Williamson, ed., *Latin American Readjustment: How Much Has Happened* (Washington, DC: Institute for International Economics, 1989).

purpose was to identify "the lowest common denominator" of policy advice prevailing at the time. In so doing, he highlighted how these policies differed from the conventional wisdom that had prevailed 20 years before. The 10 components of Williamson's Washington Consensus are as follows.

- *Fiscal discipline.* Balanced government budgets are essential to avoid inflation, balance of payments deficits, and capital flight. Government budget deficits of more than 1 or 2 percent of GDP are considered to be excessive and evidence of policy failure.
- *Reordering public expenditure priorities.* Expenditures on subsidies of various kinds (gasoline, food, and many other products) and politically sensitive areas, including the military, should be cut back and instead should be concentrated on education, health, key infrastructure, and maintaining essential public administration (but cutting back on bloated bureaucracies).
- *Tax reform.* To alleviate budget deficits, improve incentives, and achieve equity, reforms call for modest marginal tax rates and a broadening of the tax base.
- *Liberalization of interest rates.* Market forces should determine interest rates and the allocation of credit. Interest rates need to be positive in real terms—that is, the nominal interest rate should be higher than the rate of inflation.
- *Competitive exchange rates.* The exchange rate also should be determined by market forces and at a level that maintains a balance-of-payments equilibrium and, in particular, facilitates the expansion of exports.
- *Trade liberalization.* Elimination of quantitative restrictions (import quotas and import licensing) and a move toward uniform and low tariffs on imported goods. Such measures, in part, would encourage domestic firms to sell to foreign markets rather than focus on protected domestic ones.
- *Liberalization of foreign direct investment.* Restrictions on foreign direct investment should be removed so as to facilitate the inflow of needed capital, skills, and know-how.
- *Privatization.* Public enterprises should be privatized, to both relieve pressures on the government budget (from subsidies to loss-making public enterprises) and promote greater efficiency. This reflected the prevailing view that private firms generally are more efficient than those in the public sector.
- *Deregulation.* The primary vehicle for promoting competition within the domestic economy is to greatly reduce or eliminate government regulation, especially those that limit the entry and exit of firms. However, regulations that ensure safety, environmental protection, and the supervision of the banking sector remain justified.
- *Secure property rights.* Private property rights matter and steps should be taken to strengthen these rights, including providing them at reasonable cost to the small-scale, informal sector.

The common thread of all of these recommendations was that growth was being inhibited by major distortions in the market, distortions for the most part introduced by inappropriate government policies. Removing these distortions was seen as the key to accelerating growth. Government failure, more so than market failure, was viewed as the primary impediment to economic growth and development.

Williamson's original essay was written at the end of what was referred to above as the *lost decade* for economic growth and development in Latin America. The paper was meant both as a *description* of the policy reforms recommended to Latin American economies and a *prescription* for reforms needed in the region. This package of reforms, or close variants of it, which advocated prudent macroeconomic policies, greater outward orientation, and increased reliance on free markets, was also applied outside of Latin America. After the fall of the Berlin Wall in 1989, a large group of transition economies emerged, and they were encouraged to pursue similar policies. Sub-Saharan Africa, which had been stuck in its own growth crisis, also was urged to follow this same approach. When financial crises reemerged, first in Mexico in 1994, then in East Asia in 1997, followed soon thereafter by crises in Russia, Brazil, Turkey, and Argentina, elements of the Washington Consensus were applied to these situations, even though the underlying reasons for financial crises (see Chapters 12 and 13) were often different from a failure to achieve sustained economic growth.

The 1980s and 1990s witnessed a significant shift toward the reform agenda captured by the Washington Consensus. The pendulum swung away from state planning toward the market. Fiscal discipline became more widespread, with budgets moving in the direction of greater balance; hyperinflation, which especially had plagued Latin America, became less common; exchange rates worldwide became less overvalued, with black market premiums tending to fall if not disappear entirely; financial repression (by which nominal interest rates are held below the rate of price inflation, with governments allocating scarce credit to favored firms) fell out of favor; produce marketing boards were disbanded and other forms of direct price control dismantled; trade barriers were reduced; and privatization of state-owned enterprises proceeded on all continents.[17]

SOVIET COMMAND MODEL TO MARKET ECONOMIES: THE GREAT TRANSITION

As the Latin American economies were beginning to pull out of the debt crisis of the 1980s and were moving in the direction of the Washington Consensus, one of the

[17]In Latin America, fiscal discipline reduced the average budget deficit from 5 percent of GDP to about 2 percent—and lowered public external debt from about 50 percent of GDP to less than 20 percent. Trade liberalization brought average tariffs down from more than 40 percent to nearly 10 percent. . . . Banks, power plants, telecommunication systems, and even roads and water and health services were sold off to the private sector. More than 800 public enterprises were privatized between 1988 and 1997. Nancy Birdsall and Augusto de la Torre, *Washington Contentious: Economic Policies for Social Equity in Latin America* (Washington, DC: Carnegie Endowment for International Peace, 2001).

major events of the late twentieth century was occurring in Eastern Europe. Under the leadership of Mikhail Gorbachev of the Soviet Union, the countries in Eastern Europe were freed up to go their own way. In rapid succession beginning in 1989, the wall separating East Germany from West Germany was torn down and throughout the countries of Eastern Europe, the ruling authoritarian communist parties were replaced by multiparty parliamentary systems. Soon thereafter, the Soviet Union itself collapsed and was broken up into 15 separate and independent countries.

For the countries of central and eastern Europe that had been independent before World War II and in the newly independent three Baltic states, the economic goal of the new governments was clear—they wanted to become as much like western Europe as possible, and that included the rapid replacement of Soviet-style central planning and administrative allocation of most industrial products with a market economy. In Russia there was more division of opinion, but there too under President Boris Yeltsin the country set out to transform itself into a market economy as well as a more democratic polity.

The question in all of these transition economies was how to carry out this transformation. Should the transition be carried out across the board in one dramatic leap, as many foreign and domestic economists were advising? Or should the process be gradual? What exactly did the transition to a market economy involve? Some of these economies had allowed the market to exist alongside the command economy, notably in agriculture and in the distribution of many consumer products, but now the challenge was how to introduce market forces into the economy as a whole. Introducing a market system into an economy governed primarily through nonmarket mechanisms is not the same as removing distortions in an existing market system, but that difference was not well understood at the time. In a distorted market system, the institutions needed for a market system to function are often in place but are being distorted by government interference—for example, in the setting of prices. In a command economy in which most goods are distributed through government bureaucracies and not through markets, many of the institutions needed for a well-functioning market system do not exist.

There are five key elements required for a market to function well:

1. *There must be a reasonable degree of macroeconomic stability*. High inflation, large balance of payments deficits, and large-scale unemployment are generally not consistent with a well-functioning market economy.
2. *Most goods and services must be distributed through the market*. This seems obvious, but remember that the Soviet-type economy mostly distributed goods and services using administrative means implemented by large government bureaucracies. Even in Western economies during wartime, it is typical for a large part of distribution to be handled by government officials rather than by the market. Commanders in the field do not buy their weapons on the market—the weapons are allocated to them by a government bureaucracy, such as a defense ministry. For goods to be distributed on the market, a

market must exist. In the case of simple markets (for example, a farmer selling an egg to a consumer), markets can spring up overnight. More complicated markets, however, require institutions that often do not exist. Stock markets and futures markets, for example, operate in accordance with an elaborate set of rules that define what kinds of activities are allowed and what kinds are not allowed. Even the sale of automobiles typically requires a distribution network of dealers who not only sell the car but also service it. Because automobiles are expensive relative to the income of the typical consumer, there must also be a mechanism set up to help finance the purchase of the auto.

3. If the goods are going to go to their most efficient uses, *prices must be freed up to reflect the relative scarcities in the economy.* Prices are the signals that guide producers on what to produce and consumers on what best to consume. The wrong prices send the goods to the wrong (less efficient) uses.
4. *There must be competition in the markets.* Monopoly power distorts prices and hence distorts the signals sent by prices to producers and consumers. A mechanism thus needs to be in place that prevents producers from colluding with each other to establish monopoly control of particular markets.
5. Getting prices right in part through competition, however, is only half of the battle. If goods and services are to be directed to their most efficient uses, producers must respond to these prices in the appropriate way—they must behave in accordance with the rules of the market, which in their most basic form say that *producers should attempt to maximize profits*. Furthermore they need to maximize profits by increasing their sales or cutting their costs, not by getting larger subsidies through their connections with government.

The standard advice of economists involved in the Eastern European and Russian transitions was similar in many key respects to the program of the Washington Consensus, and it dealt with many key aspects of what a well-functioning market requires but not all. These economic advisers typically recommended that governments should eliminate high levels of inflation and restructure the international debts that had built up because of long periods with balance of payments deficits. Assuming that as background, or going ahead even if the macro situation was not under control, the remaining steps were to have all or most goods and services sold on a market and prices freed up to find their appropriate level. The final step was to privatize as much of the economy as possible so that producers would thus focus on the maximization of profit. All of these steps were to be taken as quickly as possible, preferably all at once in a year or two.

Not all of the countries involved in this transition process followed this advice closely, but many did. The result in almost all cases was a catastrophic drop in GDP and a rise in unemployment. The transition economies fell into an economic depression that lasted for years. What went wrong? Were the recommendations about how to make the transition from a state managed economy to a market economy wrong or

did the transition uncover problems in several of these economies that had been hidden from view? The answer involves some of both.

The Soviet Union and several Eastern European countries had industries whose products were no longer as much in demand as in the past. The end of the cold war, for example, meant that the demand for military products fell off sharply in both Eastern Europe and the former states of the Soviet Union. Many other products had enjoyed captive markets through the trading system developed by the Soviet Bloc. The quality of many of those goods was too low to compete with similar products from Western Europe and other market economies, and it would take considerable time and learning to upgrade them to a competitive level. Many enterprises simply could not accomplish the necessary upgrading, and they lost their markets as consumers in the former Soviet Bloc switched to higher quality goods that were available from elsewhere if one had the necessary foreign exchange to pay for them.

But the model of how to make the transition to a market economy was itself seriously flawed in two ways. One was the view that everything had to be done together and rapidly more or less at the same time, the so-called big bang or shock therapy approach to reform. It was easy enough quickly to free up the prices of many products. It was also plausible to try to get inflation under control quickly. Bringing down inflation gradually from say 1,000 percent the first year to 500 percent the next, 200 percent the next, and so on seldom makes sense. The anti-inflationary effort is much more likely to be credible and sustaining if the money supply is brought down sharply to a level compatible with price stability in a matter of weeks or months.

What does not make sense is the view that one can change the behavior of the management of a large industrial sector over night simply by privatizing the industries involved. Most of the enterprises in the Soviet Union and Eastern Europe were run by managers who had learned to manage in the context of the Soviet command economy. They knew little about markets or how one organized a company to meet market demand. The concept of marketing as something that could be taught in business schools was unknown to most of these people. They produced what the plan told them to produce, and another bureaucracy delivered the end product to its user. Privatization could be a first step toward changing the behavior of these managers, but there were many other steps required to obtain the necessary knowledge and skills, and those took time.[18]

It is also the case that markets that did not exist in the past do not necessarily spring up quickly. This is particularly true for financial markets such as stock markets for which elaborate rules to guide and control transactions on the stock exchange need to be devised and enforcement mechanisms need to be introduced. But it also applies to less complex markets. In the early stages of the reforms in eastern Europe, the rush

[18]Privatization of state-owned enterprises was done for political and economic reasons. Privatization was seen by many reformers as a way of breaking the power of the existing political system in which senior government and party officials were closely allied with state enterprise managers.

to change sometimes led to the abolition of the existing state-run purchasing and distribution system (for grain in one case) before an alternative system was in place.

In terms of the five criteria needed for a functioning market economy, the reforms advocated by those reflecting an approach similar to that of the Washington Consensus were on target with respect to items 1 (control of macro imbalances), 3 (freeing up prices), and 4 (promoting market competition), but were less than adequate with respect to 2 (creating the institutions needed for well functioning markets) and fell far short of what was required with respect to item 5 (changing the incentives and hence the behavior of plant managers in an appropriate way).

While Latin America was recovering from its lost decade, and the countries of eastern Europe and the former Soviet Union were undergoing their radical transformation, at the other end of the Eurasian continent a similar and equally radical economic transition was already under way, only this one by any reasonable measure was a success. China began its reform effort in late 1978 with decisions to open up its economy to foreign trade and foreign direct investment and to return the management of agricultural production back to the household where before it had been managed by production teams of 20 to 30 households working together collectively. China did not begin these reforms with a complete well thought out theory or view about how reform should proceed. What China's leaders did know was that the previous system had worked badly and they wanted a system that would make China wealthy and powerful in as short a time as possible. The reforms in agriculture were followed in 1984 by replacing allocation of products by the administration with allocation through markets at market prices. That led to a boom in what was called township and village industries. By the 1990s the legal and incentive environment for foreign investors had improved to a point where foreign companies were investing more than $50 billion a year in China—a figure that would more than double by 2010. GDP growth continued at over 9 percent per year for three decades, and China was well on its way to transforming a poor rural economy into a middle-income urban and industrial economy.

China, therefore, demonstrated that a sustained high growth rate was possible if one opened up the economy to the outside world, promoted the export of manufactures, and introduced competitive market forces into most transactions. For simpler markets and production, China made some of these changes rapidly, but when large enterprises had to be transformed from creatures of central planning into internationally competitive market-oriented firms, change was very gradual. Further, China still lacks, for example, many of the institutions needed for a Western-style market economy and a legal system capable of settling major economic disputes without political interference. The more complex the institutions required, the longer it takes to introduce them and make them function properly.

Beginning in 1986 and accelerating after 1989, Vietnam followed a reform path similar to that of China. This demonstrated, among other things, that the success of China's step-by-step approach was not a fluke.

The other major country that introduced broad-based reforms designed to increase the role of market forces and free up technical and entrepreneurial energy was India. India, as pointed out earlier, had been heavily influenced by both the Soviet economic model and the British Fabian socialists' views on the importance of active regulation of the economy. The result for decades had been a slow rate of growth of GDP that raised per capita incomes by a little under 2 percent per year between 1952 and 1980 a rate similar to that of China in the same period. Modest reforms reducing regulation and freeing up markets led to some acceleration in per capita growth in the 1980s to a bit over 3 percent a year; then major reforms beginning in and after 1991 led to East Asian–style growth rates of nearly 5 percent per capita through 2010. (Total GDP grew at nearly 8 percent per year over the reform period.) Unlike China where the most rapid growth was in industry, in India the highest growth sector was in services. The reason is that industry still retained a high degree of regulation and India's political leaders had not yet made a full commitment to opening their economy. Hence, Indian manufactured exports remained far below those of the Four East Asian Tigers and particularly China, but India did achieve significant success in exporting services. India's lower level of dependence on exports, it should be noted, also reduced the impact of the 2008–10 world recession on India.

China and India thus present important examples of large economies that grew with a mixed model of state involvement, both adopting selective dimensions but neither adopting all of the Washington Consensus. Together India and China account for 37 percent of the world's population, in contrast to 16 percent for all of the high-income countries of Europe, North America, and Japan combined. Arguably their rise out of extreme poverty into middle-income status is the most significant event of the past half century.

WAS THE WASHINGTON CONSENSUS A SUCCESS OR A FAILURE?

During the first decade of the twenty-first century there has been a growing backlash against the forces of globalization, which in effect is also a backlash against many elements of the Washington Consensus. Demonstrations, often violent, have disrupted some international meetings set up to further the liberalization of international trade, for example in Seattle in 1999 and Cancun in 2003. The global recession of 2008–10 also led many economists to question the effort of the past decades to reduce sharply the amount of government regulation in the economy. All of the high-income countries introduced new regulatory measures for their financial markets, and many developing countries have pulled back from efforts to further liberalize their systems. Was the earlier effort to reduce the government regulatory role in the economy wrong? Was it wrong for all countries or for only the few that took the process too far?

The issue for developing countries as of 2010 was whether the Washington Consensus approach to development was wrong or whether it furthered their economic goals, even if it did not offer a complete solution. To begin with we note that many

developing countries across the globe did make an effort to implement many of the recommended policies contained in the Washington Consensus. In Latin America in particular most countries succeeded in ending decade after decade of high rates of inflation. Exchange rates for the most part were no longer overvalued and import barriers were greatly reduced. Fiscal discipline was much improved, although this was not universal. Argentina, for example, tried to ensure stability by pegging its currency to the dollar from 1991 to 2002 in what amounted to a currency board arrangement (that is, allowing full convertibility between the peso and the dollar at a fixed rate of exchange), but the central government could not control the spending of local governments, and maintaining the currency peg became impossible.

The data in Table 5–1 compare Latin America's ability to control inflation and its rate of per capita GDP growth starting with the lost decade of the 1980s through to 2008 and the beginning of the global financial crisis and inflation. The data are for the region as a whole (including the Caribbean nations) and for several of its largest and most important countries. The rate of inflation is chosen not only because it is an important indicator in its own right of the quality of government economic management but because high rates of inflation are closely associated with government budgets that are out of control and with a wide variety of other undesirable interventions, many of which are designed to ameliorate the effects of inflation.

The data in Table 5–1 make clear what was meant by the lost decade of the 1980s. Per capita GDP growth in the region as a whole and in most of the major countries was low or negative. In per capita terms the region as a whole suffered a decline in income. The 1980s were also a period of hyperinflation, the kind of inflation in which prices change almost daily (see Chapter 12). During the 1990s, particularly during the later years, Latin America began to get control of inflation, and by the first decade of the next century the rate of annual price increase was in the single digits throughout the region.

TABLE 5–1 Economic Growth and Inflation in Latin America

	PER CAPITA GDP GROWTH (%)			INFLATION RATES (%)		
REGION/COUNTRY	1980–90	1991–2000	2001–08	1980–90	1991–2000	2001–08
Latin America and Caribbean	−0.4	1.6	2.3	14.6	17.5	6.1
Brazil	0.1	1.0	2.3	396.4	308.9	8.1
Mexico	0.3	1.8	1.3	60.7	16.3	7.8
Colombia	1.5	0.7	3.0	24.7	22.5	6.7
Argentina	−2.4	2.2	3.2	490.7	10.3	12.7
Venezuela	−2.4	0.0	2.7	22.1	40.1	25.0
Chile	2.5	4.8	3.1	20.9	8.5	5.7

GPD, gross domestic product.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

Did control of inflation together with all of the other reforms introduced in the 1990s and after improve the overall performance of the economy? For the most part economic stagnation came to an end in the region as growth resumed, but the growth rates achieved were modest, particularly compared with what was happening in East and Southeast Asia and in India. The most plausible explanation for this performance was that the reforms designed to improve the functioning of markets in Latin America were necessary if higher growth was to be achieved, but were not by themselves sufficient to achieve the very high catch-up growth rates experienced in Asia in recent years. There is more to achieving a high rate of GDP growth than simply getting the prices right by removing market distortions caused by inappropriate government interventions. There is a need for new technologies appropriate to the conditions in specific countries. It is also essential to have high-quality human resources, people who are well educated and healthy. Investors also require a stable political environment. These other elements in the development equation will come up again and again in the chapters that follow. Here we simply note that the level and nature of government intervention in Latin American in the 1980s and before was excessive, and the efforts to reduce and redirect the nature of government intervention in the 1990s and during the first decade of the twenty-first century appear to have had a positive impact but not as large an impact as many of the advocates of the Washington Consensus reform program had hoped for.

It is also the case that market-oriented reforms in Latin America did little to improve the highly unequal distribution of income that characterizes many of the countries in the region. This in turn is part of the reason for the reaction in the first decade of the twenty-first century that brought many leftist governments to power. Some of these governments, notably Brazil, retained market oriented reforms while making a greater effort to improve social welfare initiatives. Others, notably Venezuela, reverted to earlier approaches that used high oil prices to increase subsidies to political supporters, an approach that has generally proved to be unsustainable once oil prices fall.

Any attempt to assess the impact of market reforms on development in Africa is even more complex than for Latin America. Many countries in sub-Saharan Africa during the turn of the century did make a major effort to remove distortions in markets caused by inappropriate government intervention. Overvalued exchange rates were devalued, and the role of government marketing organizations was reduced. Trade restrictions were also cut back in much of the region. But sub-Saharan Africa is made up of 47 countries, some of which have made a major effort to reform, while others are mired in civil war or rampant corruption. To understand the impact of market-oriented reforms, therefore, one needs to separate out the countries that have made a major reform effort from those that have not. A recent study by Steven Radelet has done precisely that, dividing the countries in sub-Saharan Africa into three groups: emerging nations that have made major reforms and have enjoyed sustained increases in per capita income, oil-exporting countries whose income has

TABLE 5–2 Emerging African Countries and Growth in Average Incomes, 1996–2008

EMERGING COUNTRIES	ANNUAL PER CAPITA INCOME GROWTH (%)	CUMULATIVE INCREASE IN AVERAGE REAL INCOME (%)
Botswana	4.1	68
Burkina Faso	2.8	43
Cape Verde	4.0	67
Ethiopia	4.1	65
Ghana	2.6	40
Lesotho	2.3	33
Mali	2.5	37
Mauritius	3.7	61
Mozambique	5.3	96
Namibia	2.4	36
Rwanda	3.7	60
São Tomé and Principe	5.0	40
Seychelles	2.5	37
South Africa	2.0	29
Tanzania	3.0	46
Uganda	3.8	61
Zambia	1.8*	25
Average	3.2	50

*We include Zambia even through its 13-year growth rate is slightly lower than 2 percent because its annual average growth rate for the 10-year period was 2.3 percent.

Source: Steven Radelet, *Emerging Africa: How 17 Countries Are Leading the Way* (Washington, DC: Center for Global Development, 2010).

tended to fluctuate with the ups and downs of petroleum prices, and the others that have not made much of a reform effort and/or are mired in civil war. He identifies 17 emerging African countries that have carried out substantial reforms and have achieved sustained per capita growth since 1996 (Table 5–2).

These reforming or emerging African economies have lowered tariff rates, eliminated overvalued exchange rates, and in a variety of other ways improved the climate for private businesses more than in the countries in Africa that did not perform as well (Table 5–3). Among the countries that had generally not done as well were the oil exporters that took in large amounts of foreign exchange from oil revenues; African governments typically could not manage the funds efficiently or without large-scale corruption. The poorest performers included countries wracked by civil war (notably the Congo) or those that were badly managed by an oligarchy willing to do almost anything to hold on to power (for example, Zimbabwe).

As in the case of Latin America, market-oriented reforms in parts of Africa did not eliminate all of the barriers to high growth in the region. The growth rates achieved by the emerging African countries were far below the per capita GDP growth rates achieved in Asia. But the growth rates in the 17 emerging African countries were also

TABLE 5–3 Average Rank in Doing Business of Sub-Sahara African (SSA) Countries out of 181 Countries Worldwide, 2009

INDICATOR*	EMERGING SSA	OTHER SSA	LOW INCOME (WORLDWIDE)
Total ease of doing business	108	158	141
Starting a business	97	146	118
Dealing with construction permits	109	121	124
Employing workers	100	132	114
Registering property	113	129	118
Getting credit	100	135	123
Protecting investors	87	131	115
Paying taxes	85	128	120
Trading across borders	128	139	141
Enforcing contracts	92	134	113
Closing a business	108	140	135

*The lower the ranking, the greater the ease of doing business.

Source: Steven Radelet, *Emerging Africa: How 17 Countries Are Leading the Way* (Washington, DC: Center for Global Development, 2010).

far above what those countries had achieved in the past before 1996 and were also significantly above what other sub-Saharan African countries were achieving after 1996.

What then can one conclude about the appropriate mix of market influences and the role of government intervention? Was the Washington Consensus approach and its variant dealing with reform of former Soviet-type economies the right answer? Or was that approach far off the mark? Economists and others continue to actively debate this question. Discussions of a post–Washington Consensus have tended to moderate the original's exclusive focus on market liberalization, highlighting instead the role of good governance and strong institutions as pillars of development.[19]

The first conclusion that can be derived from an overview of the actual experience of developing countries is that there was considerable variation in the degree to which these countries, including the most reform minded among them, applied the policies recommended by the Washington Consensus. China and Vietnam, for example, did not carry out privatization of most of their state-owned enterprises. Other countries retained restrictions on imports, and the Four East Asian Tigers model that inspired many reform efforts actually retained tight restrictions on imports during the first three decades of rapid growth. While few if any of the countries that attempted to implement market-oriented reforms carried out all of them, virtually all of the more successful countries, whether in Asia, Latin America, or sub-Saharan Africa, did make significant moves in the direction of reduced government intervention and greater

[19]A good review of these debates as seen by one analyst is Derek Heady, "Appraising a post-Washington Paradigm: What Professor Rodrik Means by Policy Reform," *Review of International Political Economy* 16, no. 4 (2009), 698–728.

reliance on market forces. No country, however, had to move all the way to a perfect market economy to achieve positive results.

When it comes to the question of what is the appropriate level and nature of government intervention in the economy of a developing country, there is no one-size-fits-all formula. The governments of some countries such as South Korea and Taiwan have actively promoted particular industries using a variety of subsidies and other interventions and have done so successfully, at least in the early stages of their accelerated growth period. Governments elsewhere, Indonesia in the 1990s, for example, pursued government-led industrial policies that put increasing emphasis on large state-supported import substituting industrial projects with disastrous results. Yet, Indonesia before the 1990s had done well as a result of policies that had steadily reduced the regulation of trade and industry and had created an environment conducive to the export of labor-intensive products. Governments in most developing countries have played a dominant role in the provision of infrastructure (roads, railroads, electric power, and so on), and some have carried out these large capital investments with comparative efficiency, while in others the process has been riddled with corruption and waste.

The point here is that some governments in developing countries are capable of carrying out a relatively efficient interventionist industrial and infrastructure development policy, but other countries lack this ability. The terms sometimes used to describe these two kinds of governments are **hard governments**, which can intervene actively in a way that promotes growth, and **soft governments**, which are not able to do so. Hard governments can carry out Korean-style interventionist policies with success; soft governments can also achieve successful development but must rely as much as possible on market forces to guide the effort. Being hard or soft is not something that is a permanent condition for the government of a particular country. Governments can move from being unable to set priorities and stick to them when implementing a development program to being able to do so and vice versa. What makes some governments hard and others soft at a point in time takes one back to the underlying conditions prevailing in particular countries—the nature of their colonial experience, their culture and politics, their natural resources, and much else. These topics go far beyond what can be covered in this book. Hard or soft, however, as economies become more advanced and sophisticated intervention becomes more and more difficult, market forces need to play a larger and larger role. Government intervention should be carefully targeted toward areas in which market guidance alone is clearly insufficient (the financial sector, environmental controls, and amelioration of extreme poverty are three important areas).

A second conclusion is that market-oriented reforms do not require that everything be done at once and as quickly as possible. Reform can be achieved step by step over a substantial period of time—the big bang or shock therapy approach was not necessary, and in most cases was probably not feasible or even desirable. Finally, making greater use of market forces and reducing the role of government intervention

in the economy is not a magic bullet that solves all of the problems of development. A highly interventionist approach by which government actions are driven as much by politics and rent seeking or corruption can slow or even stop growth, but reliance on market forces by itself does not guarantee high economic growth rates. A large degree of reliance on market forces is generally helpful to development, but there is much more to the development process, as the chapters that follow make clear.

SUMMARY

- The respective roles of states and markets in shaping economic development have been a central focus of debate since the time of Adam Smith. This debate took on new urgency in the period after World War II, when today's developing countries began to emerge from colonial rule.
- Early postwar development thinking focused on the accumulation of capital as the key to economic growth. This idea was formalized in early growth models, such as Harrod-Domar; yet, the intellectual origins of the focus on capital accumulation arose from such divergent sources as David Ricardo and Karl Marx.
- The experience of rebuilding western Europe after the war bolstered the capital-oriented view of development, but early analysts failed to note the important role played by supporting institutions that had long existed on the Continent.
- The experience of the Great Depression and massive market failure before World War II contributed to widespread skepticism among economists and policy makers in both emerging industrialized countries and developing countries about the ability of markets to guide development. The perceived early success of the Soviet growth model that rejected reliance on market forces in favor of an economy run almost entirely by large state bureaucracies also influenced developing country thinking, particularly in China and India.
- Market skepticism extended to international trade as well, leading many countries to turn inward and erect trade barriers to protect their domestic economies from what appeared to be unreliable international market forces. This import substituting approach to development was particularly prevalent in South America, where early industrialization had begun in the first half of the twentieth century, when two world wars and the Great Depression made export-oriented industrial growth virtually impossible; but industry began to stagnate as small domestic markets became saturated.

- By the late 1970s, poor economic performance in many developing countries following government-led models of development and import substitution, together with the rising debt burden of many Latin American countries, led to a broad questioning of that approach and a reexamination of the potential benefits of markets and trade for development. At the same time, economists and policy makers were becoming more aware of alternative and successful growth models being pursued in East Asia. These models stressed the export of manufactures as central to growth and limited import substituting protection to the first few years when a new industry was learning how to compete. The role of government in these successful countries, however, varied considerably from the highly interventionist model of South Korea, Taiwan, and Japan to the almost laissez faire model of Hong Kong.
- By the 1980s and early 1990s, the perceived development failures in the developing world outside of East Asia had led the major international development agencies and the U.S. government, among others, to be advocates of a development model that came to be known as the Washington Consensus. This model called for developing countries to practice fiscal discipline and rely for the most part on market forces to guide the economy. Foreign assistance and debt relief was often made conditional on a country implementing the components of this consensus.
- The Washington Consensus was first promoted in sub-Saharan Africa, even before the term was coined, but it was later central to international efforts to help Latin America recover from the lost decade caused by the debt crisis of the 1980s. This same set of policy prescriptions was then advocated for and to some degree implemented by the states of Eastern Europe and the former Soviet Union as they rejected the Soviet-style command economy in favor of market forces. The economic performance of the countries that adopted many or even most of the policies called for by the Washington Consensus, however, was often disappointing.
- While much of the international aid effort was focused on implementing the goals of the Washington Consensus, China, beginning in 1978, was implementing a model that had some of the components of that consensus but lacked many others. China moved step by step toward a model that relied increasingly on markets but with a high degree of state intervention and continued state ownership of many of the larger enterprises. By 2011, that model had produced decades of high, near double-digit growth.
- India and Vietnam by the 1990s were also moving away from the extreme levels of government intervention in the economy that characterized their earlier development strategies, but like China they implemented only some of the components of the Washington Consensus. Like China, they began to

grow very rapidly throughout the 1990s and the first decade of the twenty-first century.

- Was the Washington Consensus a failure? There is no question that it oversimplified what was required to achieve sustained economic development. Not all of the elements in the consensus were necessary in all developing countries, and certain key elements, notably the quality of governance, were largely absent from the consensus. Nevertheless, when one looks at the performance of nations in Africa and Latin America, it is the countries that implemented much of what was called for by the consensus that have performed better than those that did not do so.

PART TWO

Distribution and Human Resources



Inequality and Poverty

Bangladesh is one of the world's most heavily populated low-income nations. In 2010, per capita income among Bangladesh's 164 million people was $1,486 (US$, PPP). On a daily basis, this amounts to a bit more than $4 per person per day, which would permit a meager standard of living at best. But $4 per day refers only to the average, the income level that would prevail if all of GDP was available for consumption and then was divided equally among the entire population, which, of course, it is not. What income level is earned by those who fall below the average? How much do the poorest Bangladeshis depend on for their survival? Evidence from the World Bank suggests that 50 percent of Bangladeshis survive on $1.25 a day or less, the standard international definition for absolute poverty.[1]

The most recent estimates for global poverty are for 2005, although projections have been made for more recent years. In 2005, 1.4 billion people, or more than one out of every five people, were estimated to live below the $1.25 a day poverty line; 2.6 billion people, about half the population of the developing world in 2005, survived on less than $2 a day. For most readers of this book, who are enrolled in a college or

[1]The World Bank's international poverty line originally referred to $1 a day, in 1985 US$ (PPP). It has been revised several times. The most recent estimates identify two poverty lines: $1.25 a day in 2005 US$ (PPP) is the mean of the national poverty lines for 15 of the world's poorest countries and is considered a measure of extreme poverty; $2.00 a day is the mean national poverty line among all developing countries. The selection of these poverty lines is discussed in more detail later in the chapter. Global poverty estimates refer to 2005 and are from S. Chen and M. Ravallion, "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight Against Poverty," *Quarterly Journal of Economics* 125, no. 4 (2010). Estimates for Bangladesh and other countries referred to in this chapter are from World Bank, "World Development Indicators," http://databank.worldbank.org unless otherwise indicated.

university, the idea of living on $1 to $2 a day is close to inconceivable. But these are the circumstances facing tens of millions of Bangladeshis and billions of people worldwide.

The major explanation for the degree of absolute poverty in Bangladesh and other low income nations is the low level of total production per capita. But this is not the only factor. Mexico is an upper-middle-income nation with a 2010 GDP per capita income of US$15,224 (PPP), more than 10 times the level in Bangladesh. If GDP were distributed equally in Mexico, each Mexican would have about $42 per day. But income is not equally shared in Mexico, or in Bangladesh, or for that matter, in any other country. The richest 20 percent of Mexicans receive over 50 percent of total household income, almost *15 times* as much as the poorest 20 percent who receive only 3.5 percent of total income. The distribution of income in Mexico results in over 3.5 million Mexicans (about 3.4 percent of the population in 2008) living below the $1.25 a day poverty line and 8.1 percent living below $2 a day.

Raising people out of poverty requires economic growth. Increases in GDP per capita typically benefit those below the poverty line as well as those who live near or considerably above it. Without sustained economic growth, the most Bangladesh could achieve is the low level of income that $4 a day permits. But the distribution of national income plays a vital role, too. Inequality affects the amount of poverty generated by a given level of income. It may affect growth just as growth may affect levels of inequality, and inequality *itself* is something people care about, independent of its effects on poverty and growth.

If both economic growth and distribution affect poverty levels, what does this suggest about policy? Toward the end of this chapter, we consider potential elements of a pro-poor development strategy. These elements include encouraging more rapid economic growth, improving opportunities for the poor via investments in basic education and healthcare, and designing social safety nets and other programs for especially vulnerable groups.

MEASURING INEQUALITY

Economists often are interested in the distribution of *income* among households within a nation. But these are not the only dimensions of inequality we might want to investigate. Instead of income, development economists often look at the distribution of household *consumption*, usually measured by household expenditures, whether in-kind or in money terms. In poor countries income can be hard to measure, especially for subsistence farm households who consume rather than market most of what they produce. Consumption also may be a more reliable indicator of welfare than is income, in part because consumption tends not to fluctuate as much as income from one period to the next.

One might also be interested in the distribution of *wealth,* which always is more unequal than the distribution of either income or consumption. Distributions of

assets, whether of land or education, are useful in understanding the opportunities individuals have to be productive and generate household income. The distribution of income depends on ownership of factors of production (including the value of the labor services that one "owns") and the role each factor plays in the production process. Ownership of land and capital often is highly concentrated, so anything that enhances the relative returns to these factors makes the distribution of income more unequal. Conversely, relatively higher wages for unskilled labor, the most widely distributed factor of production in developing nations, tend to lead to a more equal distribution.

In addition to deciding whether to look at the distribution of income, consumption, or wealth *within* one nation, one might want to look at how each is distributed *among* nations. We assess the level of global inequality at the end of the chapter. One can also look within the household at patterns of intrahousehold inequality, which are critical for understanding gender issues and the welfare of children.

No matter what dimension of distribution one is interested in, one needs a set of analytical tools for describing and understanding distributional outcomes. The simplest way of depicting any distribution is to display its **frequency distribution**, which tells us how many (or what percent of) families or individuals receive different amounts of income. Figure 6–1a presents the frequency distribution of household consumption per capita for Bangladesh. This distribution is based on a survey of about 7,000 Bangladeshi families, selected to represent the over 24 million households in Bangladesh. Surveyed households responded to detailed questionnaires about their sources of income and consumption of a wide variety of goods. Researchers used this information to derive an estimate of each household's per capita consumption.

Figure 6–1a tells us the percentage of individuals with different levels of annual consumption starting from the lowest reported level and rising in increments of 650 taka, the Bangladeshi currency. Almost 1 million people, less than 1 percent of Bangladesh's population, reported the lowest annual consumption expenditures in the survey, under 3,250 taka per year (less than $270, PPP); 8.4 percent (the highest bar in the figure) had per capita consumption of between 5,850 and 6,500 taka (around $500, PPP); and less than 0.5 percent had the top amounts reported in Figure 6–1a, over 22,750 taka (almost $1,900, PPP). There are households with even higher consumption in Bangladesh, but Figure 6–1a reports the distribution for only 95 percent of individuals ranked by per capita consumption. Had the top 5 percent been included in the figure, the tail of the distribution would continue to extend much farther to the right of the diagram.[2]

[2]Household income and expenditures surveys, like the one for Bangladesh, often fail to capture two groups of households. Both the poorest and richest families are likely to be underrepresented. The poorest families, including those who are homeless and live in places such as railway stations or river gullies, tend not to be adequately enumerated. Similarly, the most affluent, few in number, are unlikely to be part of the statistical sample. There also is a tendency for more affluent households not to respond to such surveys or to underreport their incomes. Thus even in the best of surveys, the degree of inequality may be underestimated.

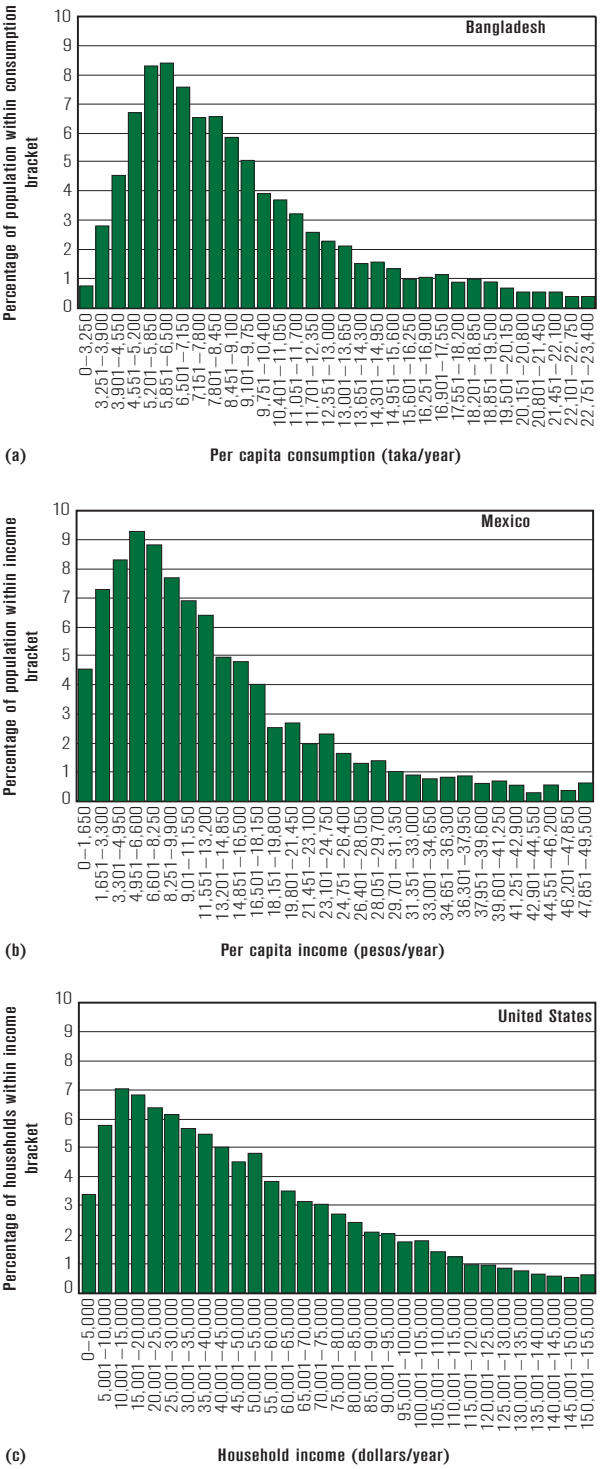


FIGURE 6–1 The Distribution of Income: Bangladesh, Mexico, and the United States, 2000s

Sources: Collaboration with Claudio E. Montenegro, World Bank; *U.S. Current Population Survey* (March 2004).

Take a look at the shape of the frequency distribution for Bangladesh in Figure 6-1a. The distribution is not rectangular, with every level of consumption represented by the same percentage of individuals, nor was it expected to be. The height of the bars, each bar representing a different consumption level, initially rises and then grows progressively shorter. If distributions were very equal, there would be a small number of very tall bars in the middle of the diagram, indicating that just about everyone had nearly the same level of consumption. The distribution also is not a normal distribution, the so-called bell curve, with which you may be familiar from courses in statistics. If consumption or income were distributed normally, there would be an equal number of households on either side of the mean and symmetric tails to the distribution suggesting equal percentages of the population at both low and high consumption and incomes. IQs are distributed normally, but consumption and income are not. The distribution instead is described as lognormal, meaning that if you took the logarithms of household consumption or incomes and redrew the frequency distribution it would approximate the more familiar bell curve.

The lognormal distribution captures what you already know about incomes. In virtually all societies there are a relatively small number of rich households (captured in the long, flat tail to the right) and a much larger number of lower-income families who make up "the hump" of the distribution, which is located at the lower end of the income range. This particular shape of the distribution of income is not unique to Bangladesh. Similar distributional outcomes characterize low-, middle-, and high-income economies. Figure 6–1b–c illustrate this point. These income distributions are based on micro-level household surveys for Mexico and the United States. Note that all three nations exhibit similar lognormal distributions, each with its own long, flat tail to the right.

Bangladesh, Mexico, and the United States all have similarly shaped frequency distributions, but they are not identical. This means that the degree of inequality in the three nations varies. To get a sense of the differences we need to rearrange the data contained in the frequency distributions; these distributions have complex shapes and are difficult to compare across nations or within nations over time. Calculating the **size distribution** provides an easier way of identifying the degree of inequality present in the underlying distribution.

Size distributions tell us the share of total consumption or income received by different groups of households, ranked according to their consumption or income level. One can rank households or individuals by deciles or even percentiles, but the convention is to report on quintiles, ranking households from the poorest 20 percent, to the next 20 percent, all the way to the richest 20 percent of households. In the case of Bangladesh, each quintile represents about 30 million people. Summing all individual consumption expenditures in each quintile and dividing by the country's total consumption yields each quintile's share.

Table 6–1 contains World Bank estimates of the size distribution of consumption for Bangladesh and household income for Mexico and the United States. This way of

TABLE 6–1 Size Distributions of Consumption or Income within Quintiles in Bangladesh, Mexico, and the United States, 2000s

	SHARE OF TOTAL CONSUMPTION OR INCOME		
QUINTILE	BANGLADESH	MEXICO	UNITED STATES
Bottom 20%	9.0	3.5	3.4
Second 20%	12.5	8.2	8.7
Third 20%	16.0	13.3	14.8
Fourth 20%	21.5	21.2	23.4
Top 20%	41.0	53.7	49.8

Sources: Collaboration with Claudio E. Montenegro, World Bank; *U.S. Current Population Survey* (March 2004).

presenting the data makes it clear that, of the three nations, Bangladesh has the relatively most equal distribution because the quintile shares are closer to one another than is the case for either Mexico or the United States. (If the distribution were completely equal, each quintile would receive 20 percent of the total.) In Bangladesh, the top 20 percent receive 41 percent of total consumption, about 4.5 times the amount received by the poorest 20 percent. In Mexico and the United States, the ratio is much larger, roughly 15:1.[3] Some of these differences are due to what precisely is being measured—consumption or income and whether on a per capita or household basis—but much of the difference is due to the underlying distribution within each nation.

The size distribution provides a means for introducing some other techniques commonly used to measure inequality, including some that reduce the entire distribution to a single number. Data from the size distribution can be used to draw a **Lorenz curve** (Figure 6–2), named after Max Lorenz, a statistician. Income recipients are arrayed from lowest to highest income along the horizontal axis. The curve itself shows the share of total income received by any cumulative percentage of recipients. Its shape indicates the degree of inequality in the income distribution. By definition, the curve must touch the 45-degree line at both the lower-left corner (0 percent of recipients must receive 0 percent of income) and the upper-right corner (100 percent of recipients must receive 100 percent of income). If everyone had the same income, the Lorenz curve would lie along the 45-degree line (perfect equality). If only one household received income and all other households had none, the curve

[3]Unlike most low- and middle-income nations, the United States has a comprehensive system of taxation of household incomes and government transfer payments (such as Social Security). The data reported here are for money incomes before taxes and after cash transfer payments. Noncash transfers, such as food stamps and Medicare, are not accounted for. Once taxes and all government transfer payments are included, the distribution becomes somewhat more equal. The ratio of the shares of the top 20 to the bottom 20 falls to closer to 10:1 than the 15:1 reported in Table 6–1.

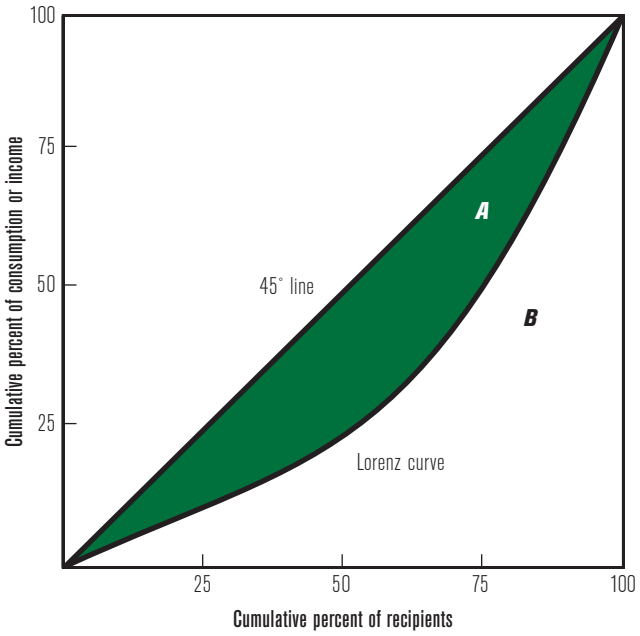


FIGURE 6–2 **Lorenz Curve**

would trace the bottom and right-hand borders of the diagram (perfect inequality). In all actual cases the Lorenz curve lies somewhere in between. The farther the Lorenz curve bends away from the 45-degree line of perfect equality (the larger the shaded area, *A*) the greater the inequality. When comparing Lorenz curves, either for the same country at different points of time or for different countries, it is possible for the separate curves to intersect. When this happens it is ambiguous if inequality has increased or decreased over time (or if one country has more inequality than the other).[4]

Single numbers can also be used to describe the distribution of income. One commonly used statistic is the ratio of the income share of the top 20 percent of households to the share received by the bottom 20 or 40 percent. The most frequently used statistic, the **Gini coefficient** (named after statistician Corrado Gini,), can be derived from the Lorenz curve. This ratio is understood most easily as the value of area *A* divided by area *A* + *B* in Figure 6–2.[5] The larger the share of the area between

[4]To see the reason for this ambiguity, draw two Lorenz curves and label the point of intersection *X*. Below point *X*, there is relatively more equality along the Lorenz curve closer to the 45-degree line; but above point *X*, there is relatively more equality along the other Lorenz curve. Because both Lorenz curves have sections of greater equality, there is no way to decide, overall, if one represents a more equal distribution than the other.

[5]The Gini coefficient can be calculated using a relatively complex formula based on the absolute income differences across all observations in a population, which are then normalized by both the size and mean income of the population.

the 45-degree line and the Lorenz curve, the higher the value of the Gini coefficient. The theoretical range of the Gini coefficient is from 0 (perfect equality) to 1 (perfect inequality). In practice, values measured in national income distributions have a much narrower range, ordinarily from about 0.25 to 0.65.

Collapsing all the information contained in the frequency distribution into a single number inevitably results in some loss of information about the underlying distribution. Argentina and Kenya both report similar Gini coefficients of around 0.48, but the underlying distributions are not identical. Both nations have considerable amounts of inequality, but the lowest quintile received 4.7 percent of income in Kenya, and only 3.6 percent in Argentina. From the perspective of the poor in these countries, a more than 1 percentage point difference in income shares is a significant amount. Another criticism of the Gini coefficient is that it is more sensitive to changes in some parts of the distribution than in others. Despite these shortcomings of the Gini, the desire among researchers to summarize inequality in a single number combined with some other attractive properties of the Gini, including its geometric interpretation using Lorenz curves, have encouraged its widespread use.[6]

PATTERNS OF INEQUALITY

Simon Kuznets, one of the early Nobel Prize winners in economics and a pioneer of empirical work on the processes of economic growth and development, was one of the first economists to investigate patterns of inequality. In his presidential address to the 1954 meeting of the American Economic Association, Kuznets reported on historical data on income shares for England, Germany, and the United States. He then introduced a few data points for the developing world: Ceylon (now Sri Lanka), India, and Puerto Rico. The data were so limited that Kuznets did not include them in a table or figure; he simply listed them in the text.[7] Some 50 years later, researchers are far less constrained. Data on inequality are available for most nations of the world, although for most countries they are updated only once or twice a decade. The quality and comparability of these data also are sometimes a concern: The World Bank, one of the primary compilers of such information, offers more than the usual caveats about data comparability across countries and over time.

Figure 6–3 presents one set of estimates of Gini coefficients of the distribution of household consumption or income for countries by region. What is immediately

[6]There are other single-number measures of inequality in addition to the Gini coefficient and top 20 to bottom 20 or 40 ratios. Discussion of the alternatives and the desirable properties of measures of inequality can be found in Gary Fields, *Distribution and Development: A New Look at the Developing World* (Cambridge: MIT Press, 2001).

[7]Simon Kuznets, "Economic Growth and Income Inequality," *American Economic Review* 45, no. 1 (1955).

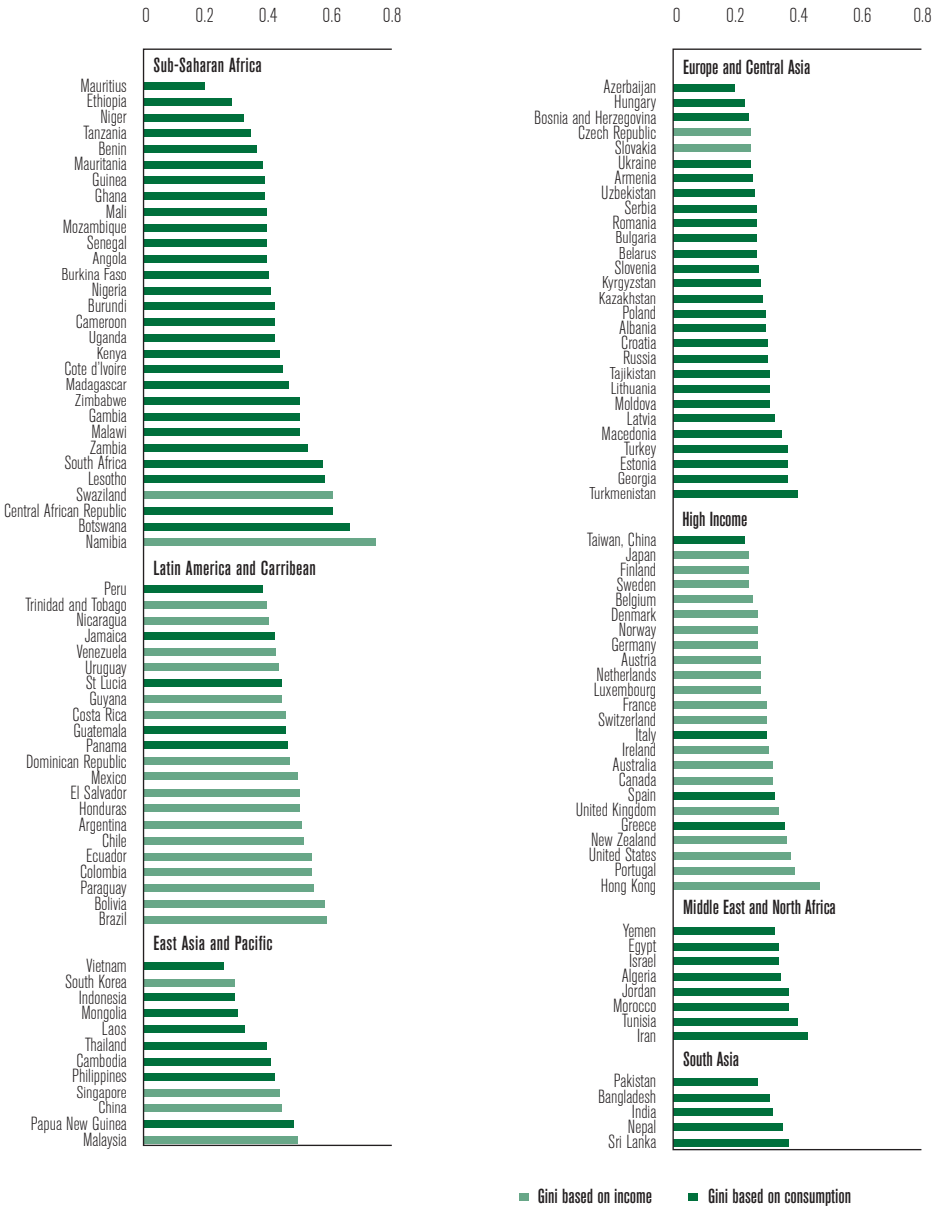


FIGURE 6–3 Gini Coeffiecients by Country and Region

Sources: Figure 2.9 from The International Bank for Reconstruction and Development. The World Bank: *World Development Report*, 2006. Reprinted with permission.

apparent is the variance in inequality both within and across regions. There is evidence of a tendency toward higher inequality, with the Gini coefficient generally above 0.40 and in some cases approaching 0.65, in Latin America and parts of Africa (especially southern Africa). Low inequality is characteristic of the transition economies of eastern Europe and central Asia, many of the high-income economies of the Organisation for Economic Co-Operation and Development

(OECD) economies, and much of South Asia. Nations in East Asia generally range from low to moderate levels of income inequality; the Middle East and North Africa report moderate levels.

What explains these differences? Are they related to specific characteristics of different regions or are regions a proxy for something else—income level, perhaps? One idea with a long history is that economic growth itself may be associated with the degree of inequality. The basic intuition is that growth is an inherently unbalanced process. Some individuals capture the benefits of growth early on, and it takes time for others to benefit and for returns to equalize.

GROWTH AND INEQUALITY

Kuznets was one of the first economists to speculate on the relationship between growth and inequality, suggesting that inequality might first increase as a nation makes the transition from a mostly agricultural economy to an industrial one. The underlying mechanism for this rise in income inequality is the result of differences in the returns to factors of production between agriculture (where they are lower and less dispersed) and industry. When everyone works in agriculture, income is distributed relatively equally, but as industrialization and urbanization progress, inequality rises. As more factors make the transition from farm to factory, inequality may then start to fall. Kuznets readily acknowledged that the basis for this relationship was "perhaps 5 percent empirical information and 95 percent speculation, some of it possibly tainted by wishful thinking."

Other economists offered alternative explanations for an association between growth and inequality. W. Arthur Lewis, a Nobel Prize winner, developed a theoretical model that predicts rising inequality followed by a "turning point," which eventually leads to a decline in inequality. Employing a two-sector model, the modern or industrial sector faces "unlimited supplies" of labor as it is able to draw workers with low or even zero marginal product from agriculture. With wages held down by the elastic supply of workers, industrial growth is accompanied by a rising share of profits. As average incomes rise, labor receives a smaller share of the total, increasing inequality. The turning point is reached when all the surplus labor has been absorbed and the supply of labor becomes more inelastic. Wages and labor's share of income then start to rise and inequality falls.[8]

In Lewis's **surplus labor model**, inequality is not just a necessary effect of economic growth; it is a cause of growth. A distribution of income that favors high-income groups contributes to growth because profit earners save to obtain funds

[8]W. Arthur Lewis, "Economic Development with Unlimited Supplies of Labor," *Manchester School* 22 (1954). The Lewis model is more fully developed in Chapter 16.

for expanding their enterprises. The more income they receive, the more they invest. Their saving and investment increase the economy's productive capacity and thus bring about output growth. Not only does inequality contribute to growth according to Lewis, but attempts to redistribute income prematurely run the risk that economic growth will be slowed. These were powerful conclusions. Could they be maintained in light of the empirical evidence on economic growth and inequality?

The ideas of Kuznets, Lewis, and others about growth and inequality held considerable sway among development economists for several decades. During the 1960s, a period of strong growth in many regions, some economists wondered why growth was not yielding more rapid reductions in poverty. One idea was that the relationship that came to be known as Kuznets inverted-U, or the **Kuznets curve**, might be at work. Twenty years after Kuznets's original paper, researchers were armed with more data on inequality and reexamined the relationship using primarily cross-section analyses of countries, including many developing countries. A key assumption in this approach was that nations at different levels of per capita income could approximate what individual nations might experience over time as they achieved economic growth. Studies using this approach supported the existence of the Kuznets curve.[9] The tendency for inequality to rise and then fall with rising levels of per capita income was maintained as a stylized fact about development until the late 1980s. Subsequent research has overturned this perspective.

Better and more abundant data on income inequality, especially time-series data on individual countries, coupled with more rigorous econometric methods, permitted researchers to identify patterns over time within individual nations. In India, a low-income nation with generally low-income inequality, there is evidence of some decline in inequality from 1950 until the mid-1960s, but at least into the 1990s there was no distinct trend in either direction.[10] Figure 6–4 illustrates the trend in the Gini coefficient since 1980 for Chile and Taiwan. Chile, one of Latin America's most successful economies, has been a middle-income nation throughout this period; Taiwan has gone from middle- to high-income status. Chile's Gini coefficient fluctuates minimally from year to year but exhibits no particular pattern over time, except perhaps for a recent slight downward trend. What is most apparent in Chile's case is the persistence of a relatively high level of inequality. Taiwan also exhibits little fluctuation over time. The most notable feature of Taiwan's

[9]Montek S. Ahluwalia, "Inequality, Poverty and Development," *Journal of Development Economics* 3 (1976).

[10]Michael Bruno, Martin Ravallion, and Lyn Squire, "Equity and Growth in Developing Countries: Old and New Perspectives on the Policy Issues," in Vito Tanzi and Ke-young Chu, eds., *Income Distribution and High-Quality Growth* (Cambridge: MIT Press, 1998).

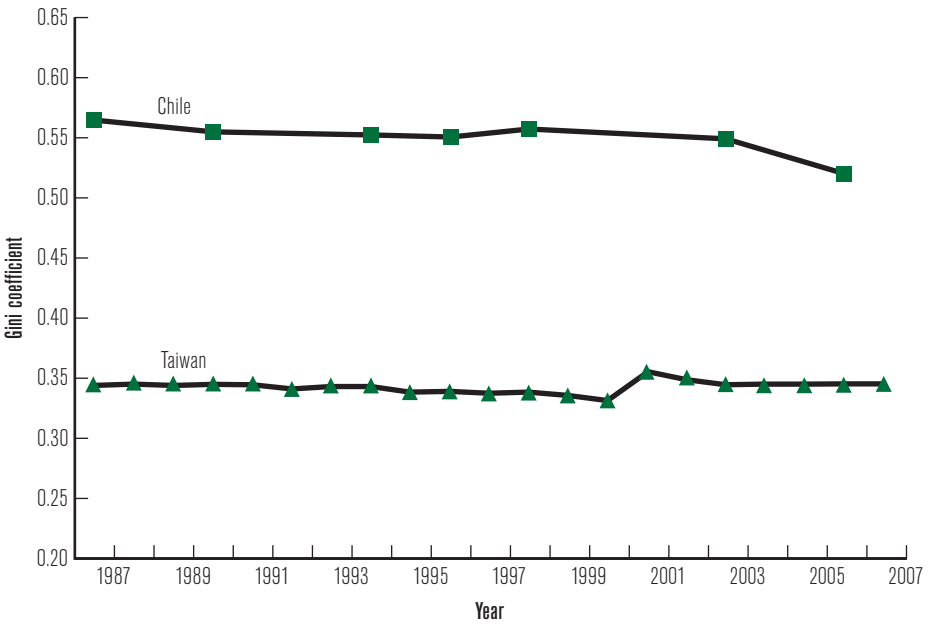


FIGURE 6–4 Trends in Inequality: Gini Coefficients in Chile and Taiwan

Sources: Data for Chile: World Bank, "World Development Indicators," http://databank.worldbank.org. Data for Taiwan: Statistical Bureau of the Republic of China (Taiwan), Report on the Survey of Family Income and Expenditure, "Percentage Share of Disposable Income by Percentile Groups of Income Recipients and Measures of Income Distribution," 2007, link available at http://eng.stat.gov.tw/ct.asp?xItem=3417&CtNode=1596, accessed June 2009.

experience is the persistence of a very low level of inequality.[11] There is no evidence of any inverted-U for India, Chile, or Taiwan.

One should not generalize from a few cases, so it is important to look at the experience of a larger number of nations. Researchers who have done so find little evidence of a general tendency for income inequality to first rise and then fall with economic growth. It is as common for Gini coefficients to rise as per capita income grows as it is for them to fall.[12] The persistence of a given level of inequality within nations may be the strongest trend. Support for the Kuznets curve in earlier cross-section analysis was driven by the higher inequality of a subset of middle-income nations. Better econometric tests reveal that the inverted-U was driven not by

[11]Taiwan's record of rapid economic growth with low income inequality is somewhat unique. Some transition economies have similarly low or even lower levels of income inequality but none has yet achieved Taiwan's per capita income level. Among rapidly growing economies, China has achieved similar growth rates, although not yet for as many years, but China has done so with significantly higher inequality. The same is true for Malaysia, Thailand, and Singapore. Korea's record on growth and inequality comes closest to Taiwan's.

[12]Klaus Deininger and Lyn Squire, "New Ways of Looking at Old Issues: Inequality and Growth," *Journal of Development Economics* 57 (1998); Martin Ravallion and Shaohua Chen, "What Can New Survey Data Tell Us about Recent Changes in Distribution and Poverty?" *World Bank Economic Review* 11, no. 2 (1997).

economic growth but by the coincidence of Latin America's high levels of inequality and middle-income ranking. Investigation of the Kuznets curve offers a cautionary tale for empirical work on developing nations. Patterns observed across nations may not always provide reliable insight into what ultimately are dynamic processes that occur within nations.[13]

Rejection of the Kuznets curve as an overall tendency does not imply that economic growth has no impact on inequality, nor does it deny that something like the Kuznets curve might occur in some countries. The lack of one general pattern simply confirms the complexity of the process that determines inequality. Research on the Kuznets curve identifies the large role played by country- or region-specific circumstances, which appear far more important in determining distributional outcomes than the level of per capita income. This is an encouraging finding. That all nations do not follow a similar distributional path suggests that there may be greater scope for government policy to influence the distributional outcomes that accompany economic growth.

WHAT ELSE MIGHT CAUSE INEQUALITY?

If inequality is not systematically associated with the level of income, what else accounts for the observed differences across countries and regions shown in Figure 6–3? There is little doubt that history and politics have played an important role. An obvious example is South Africa, which has one of the world's highest levels of income inequality. For decades, the apartheid government excluded blacks and other nonwhite South Africans from owning prime agricultural land, getting a decent education, and living in major urban areas. The legacy of these policies remains today, reflected in South Africa's highly unequal distribution of income. According to the World Bank's country brief on South Africa, the Gini coefficient in 2008 was 0.67—one of the highest in the world.

History and politics also played key roles in other parts of the world. At the end of World War II, the United States administered a land reform in South Korea as part of the dissolution of 50 years of Japanese colonial rule. These reforms redistributed land from Japanese to Korean households, providing millions of Korean families with a key asset from which they could earn a living, resulting in a relatively equal distribution of rural incomes. Low levels of inequality in eastern Europe and central Asia are, in part, the product of years of legislated wages and state ownership of the means of production. Citizens of these countries had limited opportunity to accumulate

[13]Even the experience of the United States now rejects Kuznets's original hypothesis. Income inequality in the United States fell over the four decades after the great Depression. But, after 1975, during a period of continued growth in the U.S. economy, household incomes became increasingly *unequal,* the opposite of the predictions of the Kuznets curve. Economists believe technological change, immigration, increased international trade, and a slowdown in the growth of school attainment are some of the factors that explain this post-1975 shift.

any productive assets other than an education. It is not surprising that, with the turn toward the market, inequality is now rising in many of these countries; in some of them this is happening alongside renewed economic growth.

Some of the high levels of inequality in Latin America may be traced to patterns of land ownership dating back centuries. Not only was the colonial legacy part of this process but the demands of specific crops, including the advantages of plantation-style agriculture, were also a factor. In East Asia, rice cultivation was better suited to family farming and established a basis for higher income equality in this region. Conversely, mineral wealth, whether in diamonds or oil, tends to produce higher inequality. Factor endowments shaped distributional outcomes in the past and continue to do so today.

Resource endowments and the "persistence of the past" play significant roles in shaping a nation's distribution of income, but policy choices made today also affect these outcomes. Government policy influences the accumulation of assets, including education. Policy decisions affect the diffusion of technology and access to markets, which condition productivity growth and the returns to factors of production. Taxes and government spending, including expenditures on social safety nets, directly influence how income is divided. The level of inequality in any nation is the result of complex interactions among history, politics, resource endowments, market forces, and government policies.

WHY INEQUALITY MATTERS

If economic development requires a reduction in the amount of poverty, then the simplest explanation for why income inequality matters is that the degree of inequality plus the level of income determines the extent of poverty. Even without a discussion of how to define poverty (which we will get to shortly), some examples of the relationship between economic growth, income inequality, and poverty should be clear. If an economy grows and inequality remains unchanged, the income of the poor has grown in line with everyone else's. The poorest quintiles have more income, which potentially helps raise some households out of poverty. If per capita income remains unchanged and inequality rises, the poorest quintiles have less income and some households probably have fallen into poverty. These propositions merely express the basic mathematics governing the relationship among poverty, growth, and inequality.

But inequality and growth are not determined independently of one another. There may not be one systematic tendency for how inequality changes as nations grow, but inequality often changes with rising per capita incomes. Similarly, inequality may affect the rate of growth an economy achieves. Do nations with more inequality tend to experience slower growth, thus doubly hurting the well-being of the poor?

Finally, income inequality also matters in its own right. Societies have preferences concerning inequality and may (or may not) wish for their governments to intervene to achieve distributional outcomes.[14]

Just as there has been a long debate over the impact of growth on inequality, there is ongoing debate over how inequality affects economic growth. As discussed previously in this chapter, some early theories of economic development concluded that inequality might raise growth rates. By concentrating income in fewer hands, there might be more savings available to finance investments critical for capital accumulation. But this simple view of distribution and growth fails to capture other channels in which income inequality can be a drag on economic growth. Contemporary discussion of the relationship between inequality and growth explores these channels.

When inequality is high, worthwhile investments may not be undertaken. Poor people may have promising investment opportunities. Buying a farm animal or improving irrigation, investing in a piece of equipment or building a store, or sending a child to school all may yield a good economic return. But the individuals or families may not undertake these investments because they cannot afford them. Credit market imperfections and the inability of the poor to offer lenders collateral lowers the amount of productive investment they engage in and leads to less economic growth. If the economy had a more equal income distribution, more of these productive investments could be financed and pursued.

Another channel that links inequality and growth is the political process. There are numerous "political economy" connections between distribution and growth. Some argue that, when inequality is high, the rich use their wealth to secure outcomes favorable to their interests, influencing everything from government spending (disproportionate amounts spent on public universities, which the children of the rich attend, as compared to primary schools) to trade policy (using tariffs and other forms of trade protection to maintain domestic monopolies). These policies can lead to inefficient outcomes that lower growth rates. Others argue the opposite political response occurs. When inequality is high, populist movements may arise that focus more on redistribution and less on growth, leading, for example, to higher taxes and less investment. High inequality also tends to be associated with more violence, both personal and political, which in turn can reallocate expenditures to less productive activities (more police and private security services) and discourage greater investment.

Given the multitude of connections between inequality and growth, the net effect is an empirical matter. Studies that attempt to sort out the relationship tend to rely on the cross-country growth regressions introduced in Chapter 3. Early studies found statistical support that high initial inequality, especially of landholdings, was

[14]Evidence that people prefer fairness is presented in World Bank, "Equity and Well-Being," *2006 World Development Report: Equity and Development* (Washington, DC: World Bank, 2005.)

associated with slower subsequent growth. But later studies, using larger data sets and different econometric techniques, either found no such effect or even an opposing one.[15] The inconclusive nature of these results is not surprising, given both the complexity and potential circularity of the relationship. Inequality affects growth and growth affects inequality, complicating statistical identification. It also is unlikely that one systematic pattern describes the relationship between inequality and growth for all countries at all times. This does not imply that, in a specific-country context, inequality and growth are unrelated. High inequality may be a constraint on growth, but easing this constraint is always a daunting challenge for policy makers.

MEASURING POVERTY

People sometimes refer to inequality and poverty as if they were the same thing. They are not. Inequality is an important determinant of poverty, but the two concepts are distinct. To see why, consider the following. Assume your lot in life was to fall in the bottom quintile of the income distribution. If you could pick the nation you would live in, would it be one where the bottom 20 percent received 3.4, 3.5, or 9.0 percent of household income? If your answer is the nation where the poorest receive 9.0 percent, you are confusing inequality with poverty. You also probably are forgetting the results presented in Table 6–1. Recall from that table that the share of household income of the poorest quintile in Bangladesh was 9.0 percent; in Mexico, 3.5 percent; and in the United States, only 3.4 percent!

In the United States the bottom 20 percent refers to over 60 million people. Some of these people are destitute but the overwhelming majority is not. Almost all live in a permanent dwelling with electricity, a gas or electric stove, clean water, and indoor plumbing. Most have access to medical care (even if it means an emergency room at a local hospital) and, during childhood, receive a full regimen of vaccinations against many infectious diseases. The likelihood of contracting malaria or dying of a diarrheal disease is remote, although both were common in the United States in its earlier history. For those with children, the probability of an infant dying before its first birthday is low, and that child is likely to receive at least 12 years of education. Those in the bottom quintile in the United States are likely to own a television and a landline or cellular phone and perhaps a car, and have some access to a computer at a public library if not in their own homes.

[15]Two frequently cited papers that reach opposite conclusions relating inequality and growth are Alberto Alesina and Dani Rodrik, "Distributive Politics and Economic Growth," *Quarterly Journal of Economics* 109, no. 2 (1994); and Kristin Forbes, "A Reassessment of the Relationship between Inequality and Growth," *American Economic Review* 90, no. 4 (2000).

The poorest 20 percent are poor relative to most other Americans and may find this demoralizing, but they have a substantially higher material standard of living than the poor in either Bangladesh or Mexico. No one in the bottom quintile in Bangladesh (or in most any quintile) is likely to receive the health or education benefits or the material goods consumed by America's poorest individuals. The 30 million individuals who make up the bottom quintile in Bangladesh are likely to live in the most rudimentary of dwellings, those that can be washed away in a bad storm. Food is often scarce and clean water unavailable. Living with intestinal parasites is a regular occurrence; infectious diseases take a regular toll on young and old and infant deaths are a common event. School enrollment rates are rising, but the educational attainment of poor Bangladeshis, especially females, is well below that of their American counterparts. Consumer goods consist of a few articles of clothing, some cooking utensils and little else. Most of those in the bottom 20 percent have never placed a phone call or clicked on a mouse.

The bottom quintile in the United States receives only 3.4 percent of household income, whereas the bottom quintile in Bangladesh consumes 9.0 percent of total consumption expenditures in Bangladesh. But America's bottom 20 percent commands a much larger amount of *total* income and, therefore, enjoys a higher standard of living even if its relative share is so much less. The Gini coefficient is thus a measure of relative equity; it describes the outcomes for one group relative to other groups but provides no information about absolute standards of living.

POVERTY LINES

Just as a set of analytical tools is needed for describing and understanding distributional outcomes, a similar set of tools is needed to define and measure poverty. We focus mostly on a consumption or income definition of **absolute poverty**, but it is important to recognize that poverty is multidimensional and encompasses deprivations not readily captured by income measures alone. This should be a familiar idea because it parallels the debate over how to define economic development. Both the human development index (HDI) and the millennium development goals (MDGs), discussed in Chapter 2, go well beyond GDP per capita as a measure of well-being and similar approaches are used in defining poverty.

Poverty lines, defined as having a certain amount of taka or pesos or dollars to spend per day, can capture the degree of material deprivation but may not reflect securing basic health and education. A family may have sufficient funds to purchase a minimal basket of food, but if they have no ready access to safe drinking water, food purchases are no guarantee of meeting nutritional needs because waterborne microbes may result in gastrointestinal illness and reduce the absorption of key nutrients. In this critical sense, access to safe drinking water joins money income as a determinant of absolute poverty. The availability of public services, including basic health and education, can also have an impact on poverty status today and the

transmission of poverty across generations, independent of current consumption levels. Another dimension of poverty is vulnerability to adverse shocks. Expenditures in one period may raise a family above the poverty line, but in a subsequent period, natural disasters, economywide downturns, or even the ill health or death of a family's breadwinner can push the family below the poverty line. Families often move in and out of poverty and reducing vulnerability is intrinsic to improving well-being.

Poverty is multidimensional, and it is possible to quantify many of its dimensions.[16] A great deal of attention is paid to quantifying income or consumption poverty. Development economists often use a definition of absolute poverty by which a specific monetary value is defined as a dividing line between the poor and nonpoor. Most nations define their own poverty lines, usually basing the amount on the per capita cost of some minimal consumption basket of food and a few other necessities (Box 6–1). Food dominates these consumption bundles because it may account for two-thirds to three-quarters of poor people's total expenditures. In many low-income nations, poverty lines are based on a standard of obtaining 2000 or more calories per day. While these caloric requirements seem "scientific," the actual poverty line remains a social construct. The food purchased to achieve these calories depends on what individuals actually choose to buy. Expenditures even lower than the poverty line might achieve required calories but hardly anyone would actually purchase such a consumption basket.

Often governments specify more than one poverty line. Because of regional price differences, distinct poverty lines may be applied for urban versus rural areas or, as is the case in Bangladesh, for different regions of the country. Once a poverty line (or lines) is established and expressed in a nation's own currency, that level of consumption or income has to be adjusted on an annual basis to account for changes in the price of the underlying bundle of goods. The goal is to maintain a constant poverty line over time, holding constant the ability to purchase the core consumption basket of food and other necessities.[17] This permits policy makers and researchers to chart the progress a country or region is making in lifting people out of absolute poverty.

[16]Multidimensional indices of poverty have been developed by scholars at the Oxford Poverty & Human Development Initiative (www.ophi.org.uk). With their assistance, the multidimensional poverty index (MPI) was introduced by the UN Development Programme (UNDP) in its 2010 human development report and serves as an alternative to measuring income poverty. The MPI defines someone as poor if he or she experiences deprivation in a number of areas, including education (for example, "if no household member has completed five years of schooling"), health (for example, "if any household member is malnourished"), and standard of living (for example, "if the household has no electricity" or "if the household has a dirt, sand, or dung floor").

[17]Poverty lines also can be expressed in relative terms. In the European Union, poverty is sometimes defined as living below 60 percent of median income. With this definition, the poverty line does not represent the ability to purchase a fixed bundle of goods but changes as median incomes change. Using this approach, absolute poverty declines only if incomes become more equally distributed, not if there is a general increase in per capita incomes.

BOX 6-1 NATIONAL POVERTY LINES IN BANGLADESH, MEXICO, AND THE UNITED STATES

Instead of one official poverty line, Bangladesh has many. Separate poverty lines exist for each of 14 regions to reflect varying costs. Regional lines are further divided into upper and lower levels to capture different intensities of poverty. All poverty lines are based on securing a minimum daily caloric intake of 2,112 calories. The representative bundle of food to obtain these calories was specified in the early 1990s and is made up of 11 items: rice, wheat, pulses, milk, oil, meat, freshwater fish, potatoes, other vegetables, sugar, and fruit. The cost of this bundle is adjusted using a domestic price index. The lower poverty line in each region represents the level of poverty at which a person does not have the resources to meet both food and nonfood requirements and must sacrifice some minimum daily caloric requirement to afford essential nonfood needs. The upper poverty line represents a level of poverty at which a person is able to meet minimum daily food requirements and afford some nonfood expenditures.

Mexico has three official poverty lines that capture a range of conditions of poverty. Within these lines, there is differentiation between rural and urban populations. The lowest poverty line is estimated by calculating the cost of a representative bundle of food, taking into account the differing nutritional requirements of rural versus urban dwellers in terms of daily calories and grams of protein. Falling below this poverty line indicates that a person cannot meet even these minimal daily nutrition requirements. Falling below the second poverty line means the person does not have the resources to meet both daily nutritional requirements and minimum health and educational expenses. The third line indicates that resources are insufficient to pay for all necessary costs of living, including food, education, health, clothing and footwear, housing, and public transportation expenses.

The United States also specifies multiple poverty lines that vary, not by location, but according to household size and age of household members. U.S. poverty lines, like those in Bangladesh and Mexico, start with the cost of a basket of food items. Designed to meet a person's nutritional needs at minimum cost, the bundle of food items used is still based on a 1955 survey of household food consumption. It is made up of servings of milk, cheese, and ice cream; meat, poultry and fish; eggs; dry beans, peas, and nuts; flour, cereal, and baked goods; citrus fruit and tomatoes; dark-green and deep-yellow vegetables; potatoes; other vegetables and fruits; fats and oils; and sugars and sweets. The cost of the bundle of food was multiplied by three to arrive at the poverty threshold because the 1955 survey found that the average family of three or more people spent

approximately one-third of its disposable income on food. Since adopting these poverty lines in 1965, the dollar value is adjusted annually to account for price inflation. Neither the bundle of food items nor the portion of income a family spends on food has been adjusted in almost 60 years despite changes in diets and evidence that even poor Americans spend less than one-third of their after-tax income on food.

Sources: Fernando Cortés, Daniel Hernández, Enrique Hernández Laos, Miguel Székely, and Hadid Vera Llamas, "Evolución y características de la pobreza en México en la última década del siglo XX," Economia Mexicana NUEVA EPOCA, vol. XII (2003), available at www.economiamexicana.cide.edu/num_anteriores/XII-2/Fernando_Cortes.pdf. Eloise Cofer, Evelyn Grossman, and Faith Clark, "Family Food Plans and Food Costs: For Nutritionists and Other Leaders Who Develop or Use Food Plans," Home Economics Research Report 20 (Washington, DC: U.S. Government Printing Office, November 1962), available at http://aspe.hhs.gov/poverty/familyfoodplan.pdf, accessed July 2005; Constance Citro and Robert T. Michael, *Measuring Poverty: A New Approach* (Washington, DC: National Academy Press, 1995); World Bank, "Poverty in Bangladesh: Building on Progress," Report 24299-BD (Washington, DC: World Bank, December 2002).

Most nations have their own poverty lines, and these could be used to make international comparisons. One could combine the number who are deemed poor in Bangladesh (daily per person regional poverty lines of between roughly 19 and 32 taka, or US$1.70 and $2.80, PPP) with the number said to be poor in Mexico (poverty lines of 30 to 45 pesos, or US$4.60 to $6.60, PPP) and in the United States (daily per person poverty line of around $15).[18] This would offer a measure of poverty as perceived by each nation. But the resulting differences in poverty rates across nations would themselves be functions of the poverty lines the nations choose. An alternative and more widely adopted approach is to establish a single global poverty line. By applying one common poverty line, often the $1.25-a-day or $2-a-day measure, it may be possible to obtain a more consistent picture of the degree of absolute poverty across countries and regions and of how the number of poor is changing over time.

Before investigating the origins of the $1.25-a-day line and regional trends in poverty, it is worth examining the use of poverty lines in a bit more detail and defining some alternative measures of poverty that can be based on such lines. Figure 6–5 reproduces the frequency distribution of consumption per capita in Bangladesh with

[18]The United States defines poverty depending on the size and composition of a household. Households with children or elderly members are assumed to have different food requirements, leading to different poverty lines. Households with more members are assumed to achieve economies of scale in consumption and this, too, affects their poverty lines. In 2009, daily per capita requirements ranged from $30 for a household with one nonelderly member to $12 for a household of eight with six children. The $15 refers to the average poverty line for households with four members. The reported values for Bangladesh and Mexico refer to their upper poverty lines, as discussed in Box 6–1, for the years 2000 and 2002, respectively.

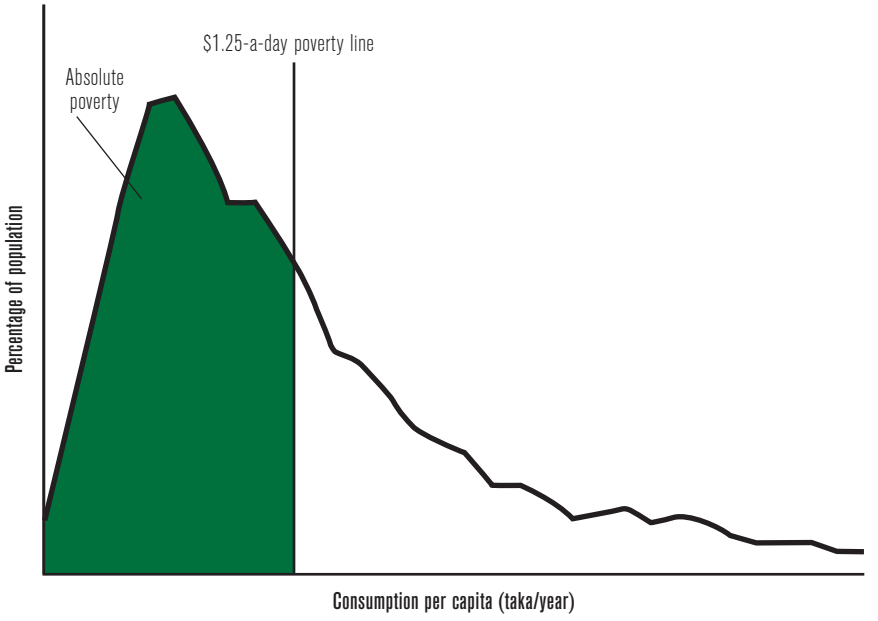


FIGURE 6–5 Absolute Poverty in Bangladesh
Source: Collaboration with Claudio E. Montenegro, World Bank.

the addition of the $1.25-a-day poverty line. Individuals with consumption below the poverty line, some 50 percent of all Bangladeshis, are considered absolutely poor. It should be evident that there is something arbitrary about this distinction between the poor and the nonpoor. Is someone just above the poverty line, with a few taka more of consumption expenditures, living that much differently from someone just below the poverty line? This arbitrary character of poverty lines is inevitable. But poverty lines still are useful in providing a sense of the extent of absolute poverty, a means for assessing the success of policies designed to alleviate poverty, and a mechanism for calling attention to and mobilizing support for reducing human deprivation. Strategies that succeed in reducing the numbers below the poverty line usually spill over and also help the "near poor," those just to the right of the poverty line in the frequency distribution.

Once a poverty line is selected, the extent of absolute poverty can be identified in a number of ways. The simplest is to report the number of people below the poverty line. Equally straightforward is the **head-count index**, the ratio of the number below the poverty line to total population. A third measure, the **poverty gap**, describes the severity of poverty. The severity of poverty refers to both how many people fall below the poverty line and how far they are from that line. Look again at Figure 6–5. Imagine if the frequency distribution below the $1.25-a-day poverty line was somewhat higher closer to the origin and somewhat lower closer to the poverty line. This would mean that poverty was more severe. The same number of individuals might

fall below the poverty line but the total amount of income they would need to get to the poverty line is greater. The poverty gap (*PG*) captures such differences and can be calculated as

$$PG = [(PL - MC)/PL] \times H$$

where *PL* stands for the poverty line, *MC* is mean consumption per capita of all individuals below the poverty line, and *H* is the head-count index. The bracketed term in the equation indicates in relative terms how far the average poor person is from the poverty line; the head-count index then weights this amount by the percent of poor people in the population. The poverty gap, a measure of how much income is needed to get the poor to the poverty line, increases the farther mean consumption of the poor is from the poverty line and the higher the share of the population is below the poverty line.

The head-count index, tells us what proportion of the population is poor, and the poverty gap tells us proportionately how far below the poverty line the mean income of the poor falls. Neither indicator is sensitive to changes in the income distribution among the poor, and by relying on the mean income gap, *PG* places equal weight on an individual just below the poverty line and on another individual who might be quite far below the line. These shortcomings are addressed simply by squaring *PG*. This third indicator, known (it's not surprising) as the poverty gap squared, places greater weight on incomes that fall farther below the poverty line and thus more fully captures the severity of poverty. These three indicators together provide a much richer picture of the various dimensions of poverty than any one of them alone.

WHY $1.25 A DAY?

The first widely used global poverty line was a $1 a day and had its origins in the late 1980s. To determine the extent of absolute poverty in the world, the World Bank's 1990 World Development Report (WDR) examined a set of 34 country-specific poverty lines from both developing and developed nations. As expected, these poverty lines generally rose with income level. Focusing only on the low-income nations in this group, the country-specific poverty lines tended to fall within a range of $275 to $370, measured in terms of 1985 PPP dollars per person per year. The upper bound of this range, just over $1 a day, was adopted as a global poverty threshold.

To chart changes in poverty over time, it is necessary to increase the poverty line in local currencies in response to changes in domestic prices. Ideally, this would be done using a price index based on the goods the poor tend to consume. In practice, a nation's consumer price index is used. To assess what happened to regional and global poverty since 1985, researchers have done more than just adjust the original $1-a-day poverty line by domestic price inflation. The most recent estimate defines *extreme poverty* as living below $1.25, measured in terms of 2005 PPP dollars.

To compute this latest global poverty line, researchers were aided by an expanded compilation of poverty lines, now including 74 developing countries, as well as new estimates of purchasing power parity, which, for the first time, included price surveys for China. The new compilation of poverty lines was especially important because the composition of low-income nations changed between 1985 and 2005. The original $1-a-day poverty line reflected national poverty lines, including those of Bangladesh, Egypt, India, Indonesia, Kenya, Morocco, and Tanzania. Economies in sub-Saharan Africa were underrepresented. Rather than relying on an inflation-adjusted value of the poverty line defined by this earlier group of countries, researchers at the World Bank repeated the original exercise and determined a new poverty line based on 15 of today's poorest countries.[19] The new group includes 13 sub-Saharan countries, Nepal, and Tajikistan. The average of the national poverty lines of this group is $1.25 a day, which is thought to be more representative of how the poorest nations define absolute poverty. The median poverty line of all developing economies in the sample was $2.00 a day, which is often used as a measure of poverty in middle-income economies, especially in Latin America and eastern Europe.

New estimates of purchasing power parity were equally important in establishing the $1.25-a-day poverty line and in assessing the extent of poverty. If two people have the same purchasing power, they should be considered poor or not poor independent of where they live. But if they live in different countries they use different currencies and face different prices. As discussed in Chapter 2, we cannot rely on comparisons made using market exchange rates as a way to ensure that these two people living in different countries will be evaluated similarly. This is because of the importance of nontraded goods in consumption. To capture PPP, economists rely on detailed price surveys in individual countries. Prices in shops and stalls of specific items, including everything from 500-gram packages of durum spaghetti to low-heeled ladies' shoes, are collected. These surveys are conducted under the auspices of the United Nations International Comparison Program (ICP) and offer estimates of PPP.

The most recent poverty estimates rely on ICP surveys from 2005, which have several important advantages over previous rounds. One problem in comparing prices in different countries is accounting for the quality of goods. Nontraded goods may be cheaper in poor countries than in rich countries, but some of this difference may be due to inferior quality, leading to an underestimation of PPP in poor countries. The 2005 ICP made corrections for this problem. It also expanded country coverage; China participated for the first time. Given the size of China's population and its success in reducing poverty, accurate measures of its PPP were critical for improved estimates of global poverty. Finally, the most recent estimates of global poverty were aided by improved coverage of the household surveys needed to make poverty estimates. Over 1.2 million households were part of 1 of 675 surveys taken

[19]S. Chen, M. Ravallion, and P. Sangraula, "Dollar a Day Revisited," *World Bank Economic Review* 23, no. 2 (2009).

in 115 countries and covering over 90 percent of the population of low- and middle-income nations. Based on these new sources of information, it was possible not only to estimate poverty in 2005 but also to reestimate values back to 1981.

Figure 6–6 presents the most recent poverty estimates by World Bank economists Shaohua Chen and Martin Ravallion. The headline news from these new estimates was that "the developing world is poorer than we thought, but no less successful in the fight against poverty." The key reason for finding more poverty was that the latest PPP estimates indicated that the cost of living in developing countries, including China and India, was higher than previously thought. Consumption expenditures therefore purchased less, causing poverty levels to rise, not only for 2005 but also in the revised poverty counts going back to 1981. As far as the world's poor was concerned, nothing had changed. The difficulty of their circumstances remained the same. It was only the official count of the number of poor that had gotten worse.

Employing the $1.25-a-day poverty line, the good news is that the number of people living in absolute poverty fell by almost 520 million people, from 1.90 in 1981 to 1.38 billion in 2005. This represents an incredible achievement in reducing human deprivation. The bad news is that 1.38 billion people in poverty still accounts for more than one out of every five people living in developing nations. A closer look at regional patterns also reveals how isolated and uneven the fall in poverty has been. Almost the entire decline occurred in East Asia, and within East Asia most of the

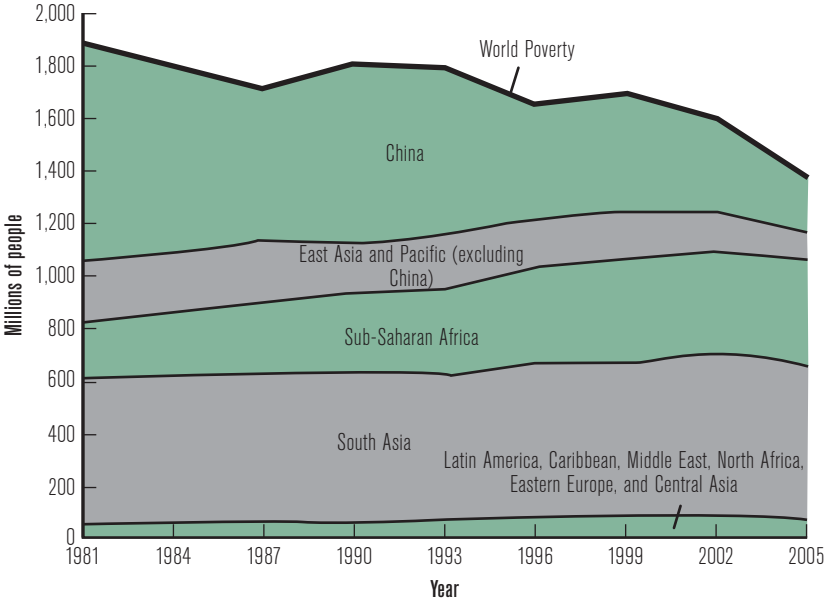


FIGURE 6–6 Number of People Living Below $1.25 a Day

Source: S. Chen and M. Ravallion, "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight Against Poverty," *Quarterly Journal of Economics* 125, no. 4 (2010).

decline is due to China's success. Over 600 million fewer Chinese people lived below $1.25 a day in 2005 than did in 1981. Many observers trace the start of China's success to the economic reforms of the late 1970s, which decollectivized agriculture and encouraged farm households to produce and market more of their output, pulling them out of poverty.

Poverty reduction was dramatic throughout all of East Asia, with large declines in Indonesia, Malaysia, South Korea, Taiwan, and Thailand. But in absolute terms, given the size of China's population, China's success dominates the global decline. By comparison, South Asia, dominated by India, saw an increase of almost 50 million in the number of poor. In sub-Saharan Africa, the trend was worse. As the population of this region grew, so did absolute poverty, from 214 million in 1981 to 391 million in 2005.[20] Trends in other regions add little to the aggregate picture for two reasons: Their population size is relatively small and, as mostly middle-income nations, the share of their populations living at only $1.25 a day was and remains low.

Trends in absolute numbers are one way to mark the progress made against absolute poverty. But it is not the only way. The head-count index presents poverty relative to population size and reinforces both the good and bad news on poverty (Table 6–2). In 1981, China's head-count index was 84 percent; by 2005, it was down to only 16 percent. In South Asia, the incidence of poverty also declined significantly, from just over 59 percent to 40 percent. Sub-Saharan Africa showed a modest decrease, from 54 to 51 percent. Latin America also had limited success at decreasing its poverty rate; the head-count index fell from 11.5 to 8.4 percent. Given their generally middle-income status, $1.25 a day is not the most meaningful poverty line for eastern Europe and central Asia (ECA) or for the Middle East and North Africa (MENA). Both regions have had poverty rates in the single digits for the past 25 years, although ECA experienced an increase in poverty since 1981 while MENA saw a decline. For the developing world as a whole, there has been a significant decline in the percentage of people who experience the grinding poverty of living below $1.25 a day. Approximately 52 percent of the developing world fell below the $1.25-a-day poverty line in 1981; 24 years later, this ratio stood at 25. Can similar progress be made over the next 20 years?

Before trying to answer that question, it is worth considering the severity of poverty as measured by the poverty gap (Table 6–2). In 1981, absolute poverty was severe in the most populous regions of the developing world. In East and South Asia, as well as in sub-Saharan Africa, the poverty gap ranged from 20 to 36 percent. But, by 2005, the poverty gap had fallen to 10 percent or less everywhere but in sub-Saharan Africa,

[20]S. Chen and M. Ravallion, "How Have the World's Poorest Fared Since the Early 1980s?" *World Bank Research Observer* 19, no. 2 (2004). The authors point out that the earlier their estimates, the less confident they are in the results because of the paucity of household surveys from the early 1980s. This is especially true in Africa. The trend toward increasing absolute poverty in sub-Saharan Africa is not in doubt, but the magnitude of this increase may not be precise.

TABLE 6–2 Absolute Poverty* by Region, 1981–2005

	NUMBER OF POOR (MILLIONS)			HEAD-COUNT INDEX (PERCENT)			POVERTY GAP (PERCENT)		
REGION	1981	1990	2005	1981	1990	2005	1981	1990	2005
East Asia	1,071.5	873.3	316.2	77.7	54.7	16.8	35.5	18.2	4.0
(China only)	(835.1)	(683.2)	(207.7)	(84)	(60.2)	(15.9)	(39.3)	(20.7)	(4.0)
South Asia	548.3	579.2	595.6	59.4	51.7	40.3	19.6	15.2	10.3
Sub-Saharan Africa	213.7	299.1	390.6	53.7	57.9	51.2	22.9	26.6	21.1
Latin America and Caribbean	42.0	42.9	46.1	11.5	9.8	8.4	4.0	3.6	3.2
Middle East and North Africa	13.7	9.7	11.0	7.9	4.3	3.6	1.6	0.9	0.8
Eastern Europe and Central Asia	7.1	9.1	17.3	1.7	2.0	3.7	0.4	0.6	1.1
Total†	1,896.2	1,813.4	1,376.7	51.8	41.6	25.2	21.3	14.2	7.6
(Total excluding China)	(1,061.1)	(1,130.2)	(1,169.0)	(39.8)	(35.0)	(28.2)	(n.a.)‡	(n.a.)	(n.a.)

*Absolute poverty refers to a poverty line of $1.25 a day (PPP, 2005).
†Total refers to low- and middle-income nations only.
‡n.a., not available.

Source: S. Chen and M. Ravallion, "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight Against Poverty," *Quarterly Journal of Economics* 125, no. 4 (2010).

where it remained at over 20 percent. Almost 1.4 billion people still lived on less than $1.25 a day in 2005, but most of them had gotten a good deal closer to this bare minimum of consumption.

China's success at lowering its poverty gap from 39 to 4 percent is an unprecedented achievement in human history. It also holds some promise for what is possible and, therefore, what nations in sub-Saharan Africa might achieve in the next few decades. Much of sub-Saharan Africa began the twenty-first century with a poverty gap well below China's level 25 years earlier. However, the challenge facing Africa is greater. In 1981, China had a more equal distribution of income than is typical of most African nations. Therefore, African growth rates have to be even faster than China's or policies of redistribution greater for absolute poverty to fall as rapidly. It is hard to envision such outcomes. China benefited from a fundamental transformation in its economy, which brought tremendous economic progress. It is hard to identify anything in Africa comparable to China's transition out of socialism and toward the market that has the potential to produce sustained growth rates in output approaching double digits. Reducing absolute poverty in sub-Saharan Africa remains a huge challenge to both African nations and the global community.

DISSENTING OPINIONS ON THE EXTENT OF ABSOLUTE POVERTY

With something as complex as estimating the amount of absolute poverty in the world, it should come as no surprise that not everyone agrees with the numbers. Some criticize the estimates as too low; others claim they are too high.

The somewhat arbitrary nature of any poverty line already has been identified. Is someone living on just less than $1.25 a day poor, whereas someone consuming just over $1.25 a day not poor? Princeton economist Angus Deaton has a different concern. He devoted his 2010 presidential address to the American Economic Association to the latest round of poverty estimates and argued that the $1.25-a-day poverty line is *too high*! The latest poverty estimates raise the number of poor in 2005 by close to half a billion people compared to using the previous method and applying it to 2005. This is a huge increase for a measure of *absolute* poverty, which is intended to stay constant in real terms over time.

Deaton traces the problem to how the poverty line was constructed. He is not persuaded that the 15 countries used to establish the new poverty line were the right ones. He notes that the small nation of Guinea-Bissau, with a population of about 1.5 million people, was included in the construction of the new poverty line, but India, with a population exceeding 1 billion people, was excluded. India has a national poverty line below $1.25 a day but is now judged as having many more poor people because of a standard determined by the poverty conditions prevailing in Guinea-Bissau (where the poverty line is above $1.25 a day) and other smaller economies. Deaton favors using a weighted average of national poverty lines from a much larger group of developing nations so that the weights reflect the number of poor people living in each nation. These and other recommended adjustments result in a global poverty line of less than $1 a day and a world poverty count in 2005 of well under 1 billion.[21]

Some commentators disagree with Deaton and see $1.25 a day as *too low* a threshold for defining absolute poverty at a global level. This cutoff has been described as "destitution" and may be too low to serve as an effective benchmark for poverty alleviation. Others are critical of what they consider the particularly arbitrary nature of any international poverty line. They argue that the use of one global poverty line bears too little relationship to national poverty lines and therefore defines poverty in way that may have little relevance to the actual bundles of goods poor people need to purchase to attain their basic needs.[22] This latter argument has special merit for middle-income nations whose poverty lines tend to be above $1.25 a day. Still others contend that even $2 a day is too low a threshold for a *global* poverty line (Box 6–2).

[21]A. Deaton, "Price Indexes, Inequality, and the Measurement of World Poverty," *American Economic Review* 100, no. 1 (2010).

[22]See *In Focus: Dollar a Day: How Much Does It Say?* International Poverty Centre, UN Development Programme, September 2004. www.ipc-undp.org/pub/IPCPovertyInFocus4.pdf.

BOX 6-2 WHO IS *NOT* POOR?

Economist Lant Pritchett argues against $1.25 a day and even $2 a day as legitimate measures of *global* poverty. He suggests instead a poverty line no less than $15 a day (2000, PPP), close to the lower bound of the prevailing poverty lines in high-income nations. Pritchett argues as follows:

> Because poverty is a social construct each country should be free to set its own definitions of poverty and its own poverty line. . . . But for setting a common, international standard for income poverty—for what constitutes "unacceptable" deprivation in the human condition or inadequate income in a globalized world—it seems grossly unfair that a person is "poor" if born in one country and yet is "not poor" with a level of real income *ten times* lower if born in another. That is, while India might set a poverty line that is attuned to its capabilities and circumstances and the USA another, for international comparisons choosing the lower line implies that what is "unacceptable" deprivation for a US resident is acceptable for another human being simply because of their residence.

Pritchett goes on to demonstrate that the World Bank's poverty lines are grossly inconsistent with achieving minimally acceptable levels of such indicators of physical well-being as infant mortality and stunting, the latter referring to the fraction of children whose height for age is less than two standard deviations below medical norms.

Pritchett recommends defining $1.25 a day as "destitution," $2 a day as "extreme poverty," and $15 a day as *global* poverty. He concludes, "This simple shift in definitions allows continuity and comparability with previous measures of poverty while embracing a new bold vision of what the dream of a world free of poverty really means."

Sources: Lant Pritchett, "Who Is *Not* Poor? Proposing a Higher International Standard for Poverty" CGD Working Paper 33 (Washington, DC: Center for Global Development, November 2003); Lant Pritchett, "Who Is *Not* Poor? Dreaming of a World Truly Free of Poverty," *World Bank Research Observer* 21, no. 1 (2006).

POVERTY TODAY

In 2005, 2.6 billion people, 47 percent of the population of all low- and middle-income economies, fell below the $2-a-day poverty line. Regional head-count indices reached as high as 73 percent in South Asia and sub-Saharan Africa. China stood at 36 percent, and Latin America at 17 percent. When comparing world regions at $2 a day, poverty remained most severe in sub-Saharan Africa (it has the highest poverty gap), but there were still more than three times the number of poor in all of Asia than in sub-Saharan Africa (Figure 6–7). What has happened since 2005?

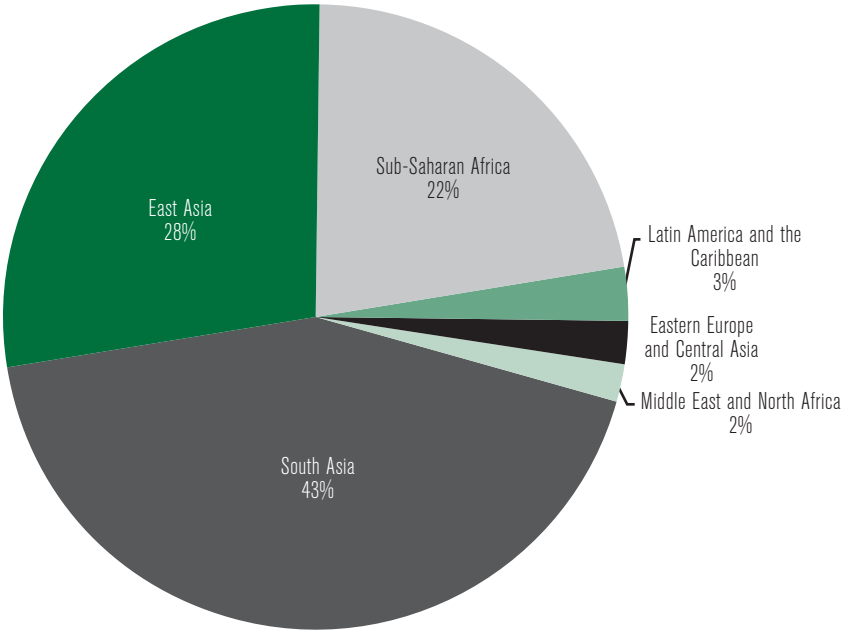


FIGURE 6–7 Regional Distribution of People Living below the $2-a-Day Poverty Line, 2005

Source: S. Chen and M. Ravallion, "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight Against Poverty," *Quarterly Journal of Economics* 125, no. 4 (2010).

The simple answer is that we do not really know. The data required for making estimates of global poverty are considerable and the exercise is not undertaken frequently. Following the financial crisis of 2008, which began in the United States and quickly spread to other nations, there was considerable fear that developing nations would be hit hard. Many were but certainly not all; China and India continued to grow rapidly. The World Bank estimated that 64 million people in developing nations were pushed into "extreme poverty" (under $1.25 a day) because of the crisis. But the years leading up to the crisis were years of strong growth in many regions of the world, and 2010 saw a faster rebound for developing regions than many expected. Far more people are thought to have escaped poverty between 2005 and 2010 than the number pushed back into poverty because of the global financial crisis. Because economic growth tends to be good for the poor, a point we return to below, there is every reason to believe that the number of absolutely poor people as well as the percentage that are poor will be much lower in the second decade of the twenty-first century than in the first.

WHO ARE THE POOR?

If the World Bank's estimates for 2005 are correct and nearly three out of every four Africans and South Asians lived below the $2-a-day poverty line, it might be easier to identify the nonpoor than the poor. This is a somewhat cynical response

to the question, Who are the poor? Even though the majority of African and South Asian populations live in conditions that qualify as poverty, some live well below the poverty line and are poorer than others. There are defining characteristics of extreme poverty.

The twenty-first century is the first time in human history that the majority of the world's population lives in areas defined as urban. Nevertheless, rural poverty, in terms of absolute numbers and rates, tends to be higher than urban poverty. This may come as a surprise to the casual observer who contrasts the image of rural villages, with their open spaces and often picturesque arrangements of straw huts and basic dwellings, with the urban squalor and densely settled and seeming disorganization in the shantytowns and favelas[23] common to large cities in low- and middle-income nations. What the casual observer fails to see are the lack of opportunities in rural areas. There are fewer ways to earn income, less education and healthcare, and often more vulnerability to the weather and forces of nature. Within rural areas, poverty often is most common among landless casual laborers or, in sub-Saharan Africa, among pastoralists. Individuals often choose to leave the countryside for the slums of the city because there are more economic opportunities in urban areas.

In addition to the rural–urban divide, poverty rates vary by regions within countries. Two examples are northeast Brazil and the Indian state of Uttar Pradesh, which for decades have been pockets of deep and persistent poverty compared to the rest of their respective countries. The persistence of regional poverty reflects both the limits of spreading development from one region to another and the constraints individuals may face in escaping poverty by migrating from one region to another. Poverty also has a racial and ethnic face. Scheduled castes in India, certain ethnic minorities in eastern Europe including Roma, and indigenous groups in the Andean region all experience poverty rates in excess of others in their societies.

Looking at poverty from the perspective of gender requires consideration of intra-household distribution, the sharing of resources within family units. Most studies of gender roles and opportunities in developing and developed nations conclude that women are disadvantaged relative to men along many dimensions.[24] Women have less access to property rights, including ownership of land, and often are denied inheritance. Girls have tended to receive less primary and secondary education than boys, although in many countries and regions this is no longer the case. Labor markets tend to discriminate against women, paying them less than men for the same work. Combining the work done at home with that done for income,

[23]The origin of the word *favela* is the Morro de Favela hillside in Rio de Janeiro, Brazil, where freed slaves established a community of squatters in the late nineteenth century. Over time the term has been adopted to describe any urban slum, especially in Latin America.

[24]This discussion draws from World Bank, *Engendering Development* (Washington, DC: Oxford University Press, 2001).

women tend to work many more hours per week. Domestic violence against women is all too common. Sex-selective abortion favors the birth of boys over girls in many parts of the world.

Given all these disadvantages, the feminization of poverty seems straightforward. But it is not. Measures of individual consumption are based on household data, and it is difficult to disentangle who consumes what within the family. Some goods, including housing, are jointly consumed. Given the lack of data on individual consumption by gender, some studies compare poverty rates between households headed by men and women. The results are inconclusive. Some categories of women-headed households do quite poorly. For example, widows without male heirs in India and elderly women on pensions in eastern Europe have particularly high poverty rates. But other women-headed households, including unmarried women working in urban areas and married women with husbands working abroad and sending remittances, may experience lower-than-average poverty rates. Studies of nutrition that assess the degree to which the needs of females versus males are met find some evidence of the relative deprivation of females but the results vary widely across regions of the world and even within countries. Despite the lack of evidence on the feminization of poverty, it is hard to imagine that reducing gender inequality would not also help reduce poverty in general.

LIVING IN POVERTY

You now know how economists define poverty, how many poor people there are, and a bit about the characteristics of the poor. What do we know about the lives of the poor? You might think living below $1.25 a day implies living on the edge of subsistence and leading a monotonous life with few choices. But this is not the case. The poor are a heterogeneous group who do not devote all their resources to food, who engage in a variety of economic activities, and who pursue strategies to reduce the risks associated with their vulnerable economic situations.

Social scientists, especially anthropologists, have been studying the poor for a long time.[25] This work has been complemented more recently with the collection of household-level surveys that include the poorest families. Many of these surveys have been part of the World Bank's Living Standards Measurement Study (LSMS) surveys, the first of which were undertaken in Côte-d'Ivoire and Peru in 1985. The LSMS and other similar surveys have been used to obtain counts of the poor and to understand the circumstances and behavior of poor households. MIT economists

[25]In 1953, Sol Tax, an anthropologist, published *Penny Capitalism* (Washington, DC: Smithsonian) based on his observation of the lives of a poor indigenous community in Guatemala. Tax observed that members of the community were efficient in their use of time and other resources despite their poverty.

Abhijit Banerjee and Esther Duflo employ 13 of these surveys from around the developing world to create a profile of the economic lives of the poor.[26]

Banerjee and Duflo find that, on average, poor households do not put every penny into purchasing more calories. Food represents about half to three-quarters of total consumption expenditures. For those living under $1.25 a day, the elasticity of spending on calories relative to total food expenditures is about 0.5,[27] suggesting that the poor care about what they eat as well as how much they eat; how food tastes matters even when one is poor. The poor, like the rest of us, spend money on nonessential items, including alcohol, tobacco, festivals (funerals, weddings, and religious holidays), radios, and televisions. The health of the poor is far more compromised than that of the nonpoor. In a number of surveys that ask about health, Banerjee and Duflo find that a family member often is reported to have been bedridden for a day or more during the past month. In the Udaipur district in India, 43 percent of households reported that members did not have enough to eat throughout the year, and 55 percent of adults were anemic.

In their productive activities, the poor tend to diversify their labor time, pursuing multiple opportunities rather than specializing. Among rural households, almost all farm their own land, but this is not their only or even their primary source of income. Poor men often work as daily laborers; poor women might sew, gather fuel, or run small eateries (for example, selling dosas in the morning in India). In one survey of rural West Bengal in India, the median family had three working members engaged in *seven* different occupations. Household members in poor families may engage in temporary migration to obtain work, usually for not more than a month or two. Permanent migration of the poor out of West Bengal is rare, and most migration for work is not far from where the poor live. Reliance on multiple occupations and a willingness to temporarily migrate to find work are strategies that reduce risk. If one activity is no longer available or remunerative, another can be pursued. Such a strategy has its costs. By not specializing and investing in specific skills, it is hard to achieve economies of scale and to gain the productivity required for higher income.

While the poor are often entrepreneurial and lead diverse lives, it would be a mistake to romanticize their situation. Living under $1.25 is a constant struggle filled with uncertainty and vulnerability. Among the poor we can identify a group of "ultra-poor," who find themselves barely surviving. A study by the International Food Policy Research Institute (IFPRI) found that in 2004, 162 million people were ultra-poor,

[26]A. Banerjee and E. Duflo, "The Economic Lives of the Poor," *Journal of Economic Perspectives* 21, no. 1 (2007). Also see their book *Poor Economics*: *A Radical Rethinking of the Way to Fight Global Poverty* (New York: Public Affairs, 2011).

[27]Calories are obtained from different types of food. Coarse grains, like rice and wheat, are inexpensive to buy and offer an individual relatively cheap calories. Meat is expensive making the calories obtained from meat consumption more costly ones. By investigating the foods poor households consume to obtain their calories, it is possible to estimate the elasticity of spending on calories relative to total food expenditures.

living on *less than* 50 cents a day.[28] In another series of studies, researchers asked the poor to describe poverty. Some of their responses follow:[29]

> Don't ask me what poverty is because you have met it outside my house. Look at the house and count the number of holes. Look at my utensils and the clothes that I am wearing. Look at everything. . . . What you see is poverty.—A man in Kenya
>
> Poverty is hunger, loneliness, nowhere to go when the day is over, deprivation, discrimination, abuse, and illiteracy.—A mother in Guyana
>
> Poverty is pain; it feels like a disease. It attacks a person not only materially but also morally. It eats away one's dignity and drives one into total despair.—A woman in Moldova

When asked to describe poverty and its causes, poor people reveal poverty to be more than just a lack of money. Not only do the poor suffer from material deprivation but poverty exacts an often severe psychological toll. There are many causes of poverty. The poor lack assets, especially land, and are aware that illiteracy often constrains their opportunities. Illness is dreaded because it can drive a family into destitution. The poor are vulnerable to environmental risks as well, whether droughts or floods or the degradation of their land. The poor often feel exploited by markets, in which private traders charge exorbitant prices for necessities or levy usurious rates for credit, and by governments, who do not provide essential services to the essentially voiceless poor unable to affect change.

STRATEGIES TO REDUCE POVERTY

The World Bank's *1990 World Development Report* not only provided an estimate of the amount of world poverty but also outlined a strategy for alleviating it. The strategy had two main elements: (1) promote market-oriented economic growth and (2) direct basic health and education services to the poor. Market-oriented growth included many of the familiar recommendations of the Washington Consensus discussed in Chapter 5: macroeconomic stability, greater economic openness to trade and investment, increased public investment in infrastructure, improved credit markets, and the

[28]Estimates of the number of ultra-poor are from Akhter U. Ahmed, Ruth Vargas Hill, Lisa C. Smith, Doris M. Weismann, and Tim Frankenberger, "The World's Most Deprived," 2020 Discussion Paper 43, International Food Policy Research Institute, Washington, DC, 2007. This study was undertaken before the latest round of the ICP. With the new PPP estimates, the number of ultra-poor would be higher. At over 160 million, the ultra-poor collectively represent a population similar to that of Pakistan's, the world's sixth most populous nation.

[29]The quotations are from United Nations Development Programme, *Human Development Report 1997* (New York: Oxford University Press, 1997), and D. Narayan, *Voices of the Poor: Can Anyone Hear Us?* (Washington, DC: World Bank, 2000).

like. Combined, these policies would lead to labor-demanding growth, which would benefit the poor because the primary asset the poor rely on is their labor. The second element of the strategy called for investing in people. Directing government health and education services to the poor would increase their productivity and thereby contribute to poverty reduction. A third but less emphasized part of the strategy was to develop social safety nets to assist individuals unable to take advantage of market opportunities. This group includes the sick and the old but also all those who suffer from systemic shocks, such as natural disasters and macroeconomic crises.

Most strategies for reducing poverty call for increasing the rate of overall economic growth, because the poor are expected to benefit from economic growth along with the rest of the population. To quote a popular metaphor for development economists, "A rising tide lifts all boats." But there are dissenting voices. In 1996, the United Nations Development Progamme (UNDP) wrote in its flagship publication, *The Human Development Report*, "Policy-makers are often mesmerized by the quantity of growth. They need to be more concerned with its structure and quality." The report goes on to specify problems with economic growth: "*jobless*—where the overall economy grows but does not expand opportunities for employment," "*ruthless*—where the fruits of economic growth mostly benefit the rich, leaving millions of people struggling in ever-deepening poverty," "*voiceless*—where growth in the economy has not been accompanied by an extension of democracy or empowerment," *rootless*—which causes people's cultural identity to wither," and "*futureless*—where the present generation squanders resources needed by future generations."[30]

The UNDP report raises many concerns. One of them is to question whether (or at least how often) economic growth is good for the poor or whether the benefits are more likely to be concentrated on the rich. The basic mathematics of growth, distribution, and poverty suggests that, generally, growth should be good for the poor. Earlier in this chapter, we saw that inequality does not systematically increase with economic growth. This implies that in most cases growth should benefit the poor just as it benefits others in a society. As long as GDP grows faster than the population, average incomes within each quintile usually also increase. Numerous studies support this conclusion.

GROWTH IS GOOD FOR THE POOR

Figure 6–8 reproduces the data from one of the studies supporting the notion that growth is good for the poor.[31] This figure compares the growth in income of the poor, here defined as the bottom 20 percent of the income distribution, with overall eco-

[30]United Nations Development Programme, *Human Development Report* (New York: Oxford University Press, 1996), p. 3–4.

[31]David Dollar and Aart Kraay, "Growth Is Good for the Poor," *Journal of Economic Growth* 7, no. 3 (2002). Earlier studies that reached similar conclusions are Michael Roemer and Mary Kay Gugerty, "Does Economic Growth Reduce Poverty?" CAER Discussion Paper No. 5, Harvard Institute for International Development, Cambridge, 1997; and John Gallup, Steve Radelet, and Andrew Warner, "Economic Growth and the Income of the Poor," CAER Discussion Paper No. 36, Harvard Institute for International Development, Cambridge, 1999.

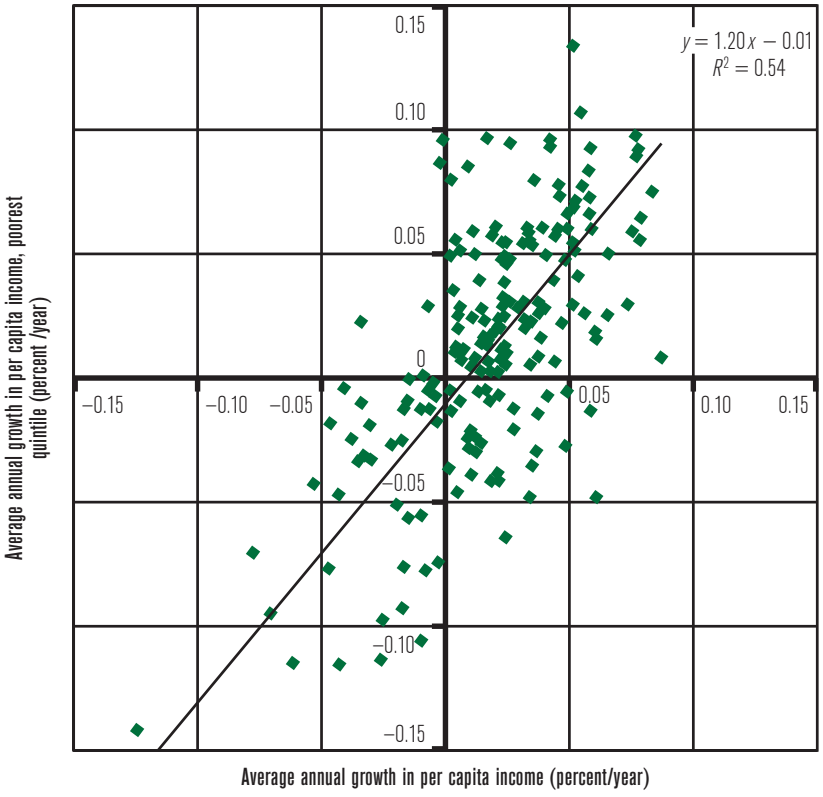


FIGURE 6–8 Is Growth Good for the Poor?

Source: David Dollar and Aart Kraay. "Growth Is Good for the Poor," *Journal of Economic Growth* 7 (September 2002). Dataset available at http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/0,,contentMDK:22802584~pagePK:64165401~piPK:64165026~theSitePK:469372,00.html.

nomic growth (growth in GDP per capita). Each observation refers to a growth spell of five or more years in a given country. In many cases, there is more than one observation per country. This occurs when information on the distribution of income is available for multiple years. Growth spells have a median length of six years. These spells occur between the years 1956 to 1999, with most taking place in the 1980s and 1990s. Not all countries are represented in the data set, but there is good coverage by region and for the world population as a whole.[32]

The trend shown in Figure 6–8 is consistent with the view that growth is generally good for the poor. The fitted regression line has a *positive* slope of 1.2, indicating that, *on average*, the incomes of the poor grow at a slightly higher rate than does GDP per capita. The slope of 1.2 means that if GDP per capita grows, say, at 5 percent, the regression predicts that the incomes of the poor will grow by 6 percent. Most observations lie in the quadrant in which both the growth in per capita income

[32]Departing from the original study, the Organisation for Economic Co-Operation and Development (OECD) economies are excluded from the figure, as are a few statistical outliers.

and the growth in the incomes of the poor are positive. This quadrant can be further divided in two. More than half the observations lie above the 45-degree line (not shown), where the incomes of the poor grow faster than average incomes. Growth is still "good for the poor" among the observations that lie below the 45-degree line but above the *x*-axis. In these cases the incomes of the poor grow at a rate somewhat less than the rest of the economy.

Returning to Figure 6–8, there also are cases of **immiserizing growth**, situations in which the poor witness a decline in their average incomes despite growth in the economy. The UNDP's concern over *ruthless* growth appears to have merit in some cases. In about one in five cases of positive economic growth, the income levels of the poor decrease. For Latin America this occurs about one-third of the time. Rapid growth, however, minimizes such outcomes. In these data, when growth is over 5 percent, there are only two cases of immiserizing growth.[33] Rapid growth seems especially good for the poor and nonpoor alike. The opposite also is true. The poor are particularly vulnerable during periods of economic decline. This is shown in the lower-left-hand quadrant, where many of the observations lie well below the regression line (a substantial number from eastern Europe and central Asia during the late 1980s and early 1990s). Although there are important exceptions, the empirical record suggests that economic growth benefits the poor more often than it leaves the poor behind.

SOMETIMES GROWTH MAY NOT BE ENOUGH

The centrality of economic growth in achieving economic development is a major theme of this textbook, as is the role of markets in achieving more rapid economic growth. But, in some situations, more than economic growth and market forces may be needed to reduce absolute poverty. First, there are the cases just identified, in which the well-being of the bottom quintile falls despite increases in GDP per capita. These are cases in which the increase in income inequality that accompanied economic growth was so large as to reverse any potential gains for the poor from rising GDP per capita. Second are cases in which, despite economic growth, too few of the poor cross the poverty line because their initial share of income is so low. This situation is not depicted by the data in Figure 6–8 because the vertical axis measures changes in the mean income of the bottom 20 percent, not changes in the headcount index or poverty gap. This makes a difference. Imagine an economy in which average incomes are low and inequality is high. If the bottom 20 percent receives only 2 to 3 percent of the total income, as they do in many countries in Latin America and southern Africa, the well-being of the poor makes little progress even if their

[33]The two cases are Puerto Rico (1963–69) and Singapore (1978–83).

incomes grow at the same rate as the nation's average income. If high levels of inequality further reduce the prospects for more rapid economic growth, the situation of the poor is that much worse.

A third situation is where distributional changes reduce the amount of poverty alleviation generated by economic growth. Latin America during the 1990s experienced only modest economic growth accompanied by rising income inequality. Even the modest growth that was achieved, assuming no change in inequality, would have reduced the numbers living under $2 a day by 90 million people.[34] But, because inequality increased during the 1990s, the number of people living in poverty fell by only 45 million. Poverty in Latin America would have fallen further had there been both faster growth and less inequality.

The data in Figure 6–8 report on rates of growth. But how does growth affect the *absolute amount* of income individuals in different quintiles actually receive? Even when the growth in income of the poor exceeds that of the rich, higher income quintiles usually still receive a much larger increase in absolute income. This is because a small percentage of a large number always represents much more additional income than a larger percentage of a small number. Consider the cases of Bangladesh and Mexico presented at the beginning of this chapter. If economic growth is distributionally neutral (GDP per capita rises but inequality remains the same), then for each increase of 100 taka or 100 pesos going to someone in the bottom 20 percent, someone in the top 20 percent, on average, receives roughly 450 more taka in Bangladesh and 1,500 more pesos in Mexico. In other words, even when growth is good for the poor and the number below the poverty line falls,[35] the absolute income gap between rich and poor continues to widen.

PRO-POOR GROWTH

With increasing attention paid to global poverty, captured by the MDG of cutting poverty in half by 2015, development experts frequently refer to the need for development strategies that adopt a poverty focus (Box 6–3). Sometimes this is referred to as **pro-poor growth**. This is not a well-defined term. Pro-poor growth is sometimes described as the situation in which income growth among the poor is faster than average income growth. The problem with this definition is that it favors circumstances in which the incomes of the poor grow, for example, at 3 percent while the GDP per capita grows at 2 percent over situations in which the incomes of the poor

[34] "Is Growth Enough?" *Latin American Economic Policies* 14 (2001).

[35] The elasticity of the poverty rate with respect to growth in GDP per capita is reported by Ravallion to be around –2 but with lots of variation. With a 95 percent confidence interval, a 2 percent rate of growth in average household income generates a 1 to 7 percent decline in the poverty rate. Martin Ravallion, "Poverty and Growth" in David Clark, ed., *The Elgar Companion to Development Studies* (Cheltenham, UK: Elgar Publishing, 2006).

BOX 6-3 WHY SHOULD DEVELOPMENT STRATEGIES HAVE A POVERTY FOCUS?

In an interview appearing in *Finance and Development*, a journal of the International Monetary Fund, Harvard University economist Dani Rodrik makes a compelling case for reform strategies to adopt a poverty focus.

> First, in considering social welfare, most people, and democratically elected governments in particular, would give more weight to the well-being of the poor than to that of the rich. The economy's growth rate is not a sufficient statistic for making welfare evaluations because it ignores not only the level of income but also its distribution. A policy that increases the income of the poor by one rupee can be worthwhile at the margin, even if it costs the rest of society more than a rupee. From this perspective, it may be entirely rational and proper for a government considering two competing growth strategies to choose the one that has a greater potential payoff for the poor, even if its impact on overall growth is less assured.
>
> Second, even if the welfare of the poor does not receive extra weight, interventions aimed at helping the poor may still be the most effective way to raise average incomes. Poverty is naturally associated with market imperfections and incompleteness. The poor remain poor because they cannot borrow against future earnings to invest in education, skills, new crops, and entrepreneurial activities. They are cut off from economic activity because they are deprived of many collective goods (such as property rights, public safety, and infrastructure) and lack information about market opportunities. It is a standard tenet of economic theory that raising real average incomes requires interventions designed to close gaps between private and social costs. There will be a preponderance of such opportunities where there is a preponderance of poverty.
>
> Third, focusing on poverty is also warranted from the perspective of a broader, capabilities-oriented approach to development. An exclusive focus on consumption or income levels constitutes too narrow an approach to development. As Nobel Laureate Amartya Sen has emphasized, the overarching goal of development is to maximize people's ability to lead the kind of life they value. The poor face the greatest hurdles in this area and are therefore the most deserving of urgent policy attention.

Source: Excerpt from Dani Rodrik, "Growth Versus Poverty Reduction: A Hollow Debate," *Finance and Development* 37, no. 4 (December 2000). Reprinted by permission of the International Monetary Fund.

grow at 4 percent while the GDP per capita grows at 6 percent, even though more rapid poverty reduction would occur in the latter case. This is not a hypothetical distinction; China's experience reflects the second scenario. An alternative and more general definition of pro-poor growth is economic growth that includes any growth in the incomes of the poor.

While not well-defined, at least the intentions of pro-poor growth are clear. What can governments do to achieve economic growth that rapidly improves the well-being of the poor? Answers lie in better understanding the complex interactions among growth, inequality, and poverty. Pro-poor growth does not represent a choice between being pro-growth and being pro-poor. The strategy calls for combining more rapid growth with increased opportunities for the poor to participate in that growth. Policies that both accelerate growth and address inequality may be needed to achieve these goals.

Those who question whether growth is good for the poor may also be skeptical about the ability of markets to help poor people. Competition may be seen as exploiting poor people, by paying them too little for their labor or by charging too much for the inputs they purchase or the goods they buy. Policies associated with the Washington Consensus are often viewed critically as protecting the interests of rich countries or rich citizens in poor nations at the expense of the masses of poor people in low- and middle-income countries. But are these criticisms valid? Can market-friendly strategies serve the interests of the poor?

Consider China's experience. Before 1980, China's farm system was dominated by collectives and communes. Market forces played little role in the allocation of farm inputs or in returns to agricultural work. Market reforms under the household responsibility system changed everything. It transferred control (but not ownership) over farmland to individual farmers. After meeting output quotas to the state, farmers were free to keep or sell additional output on relatively open markets. The unleashing of personal incentives resulted in a massive supply response and lifted hundreds of millions of Chinese peasants out of poverty. This outcome was supported by a relatively equal distribution of land, which occurred after the breakup of the communes. Low inequality in access to farmland was essential for the pro-poor growth that followed.[36]

Macroeconomic stability and economic openness are major elements of the Washington Consensus. What impact do these policies have on the poor? Maintaining fiscal discipline and reducing price inflation are key features of macroeconomic stability. Such stability usually entails reducing budget deficits, achieved by cutting government expenditures. This in turn often brings forth protests that the poor are hurt by reduced government spending on basic services. There can be considerable truth to such concerns, but it is not the whole picture. Government spending frequently fails to reach the poor, as programs, including those on health and education, often are targeted at higher income quintiles.

Maintaining unsustainable budget deficits also leads to price inflation, which tends to hurt the poor much more than the rich. Higher income individuals have ways to avoid the negative consequences of rapid price inflation. They can send

[36]Martin Ravallion, "A Comparative Perspective on Poverty Reduction in Brazil, China and India," *World Bank Research Observer* 26, no. 1 (2011).

their savings abroad or hold their wealth in land or real estate, which often are a hedge against inflation. But the poor lack these options. What minimal savings they have can disappear under persistent price inflation, the costs of even their minimal borrowing escalate, and any type of fixed income, like a pension, for example, soon becomes relatively worthless. Price inflation usually leads to the depreciation of the currency, which raises the price of imported goods, including fuel and other essential commodities. When fuel prices go up, transportation costs increase, affecting many items the poor depend on. Price inflation also acts like a brake on investment and economic growth, and slower economic growth is not in the interests of the poor. Overall, macroeconomic stability tends to be good for both economic growth and the poor.

Trade liberalization and greater economic openness entail reducing trade barriers, encouraging foreign direct investment and freeing up exchange rates. Antiglobalization advocates see such measures as hurting the poor, who cannot protect themselves from the vagaries of world capital markets or the onslaught of cheap exports. Once again, there is merit to this position in some circumstances. The Asian financial crisis of the late 1990s hurt both the rich and the poor in many economies. Cheap crops, the result (in part) of farm subsidies in the European Union and the United States, hurt farmers in some economies. Cotton growers in West Africa are often cited as an example.

But there is another side to increased economic openness. If nations export goods in which they have a comparative advantage, many low-income nations specialize in goods that rely on unskilled labor. This leads to an increase in the employment of such workers and, over time, to an increase in their wages. Because poverty is likely to be common among the least skilled, trade increases labor demand and reduces poverty. Greater openness, including the exchange of technology and the capital accumulation that results from foreign direct investment, can improve productivity and raise overall economic growth. In addition, trade may also reduce the prices the poor must pay for goods and services, further improving their welfare.

The textbook case for trade is compelling, and the success in poverty reduction throughout East Asia often identifies trade, especially exports of labor-intensive manufactured products, as an engine of growth. But there is more to economic openness than increasing exports, and a full account of the impact of globalization on poverty does not yield one simple conclusion. Trade reform can help some of the poor while hurting others. In Mexico, the North American Free Trade Agreement (NAFTA) brought more liberalized trade, which hurt farmers who previously were protected by tariffs on corn at the same time that it helped those who found employment in the growing export sector. (See Chapters 18 and 19 for a fuller account of the impact of trade on growth, development, and poverty.) Trade reforms illustrate an important principle: Market forces create new opportunities that benefit some and hurt others. Helping the poor get the most out of new opportunities is a challenge facing all economies.

IMPROVING OPPORTUNITIES FOR THE POOR

Improving the operation of markets helps the poor only if they can take advantage of these opportunities. This is why expanding basic education and health services is often a main element of strategies to reduce poverty. Education tends to make people more productive. It permits them to access new information and helps them take advantage of new opportunities, from new seed varieties that increase farm yields to new jobs that require the ability to read and write and new medicines that increase child survival or improve adult health. Education is not, however, a panacea. In a bad economic environment, education can yield a low return (see Chapter 8). But without an education, many of the poor may get trapped in intergenerational cycles of poverty. Similar arguments can be made for the importance of basic health (see Chapter 9). The debilitating effects of a host of diseases, including malaria and HIV/AIDS, prevent those who are infected from engaging in work and seizing new opportunities. No matter how well markets work, if someone is too sick to take advantage of such opportunities, poverty persists.

Investing in education and health—that is, in the human capital of the poor—is part of an agenda for improving the opportunities of the poor. Other reforms also warrant consideration. Most of the poor live in rural areas and either directly or indirectly depend on agriculture as a source of income. Studies of India find that growth in the rural economy alleviates both rural and urban poverty, whereas urban growth primarily reduces only urban poverty.[37] Supporting the rural economy is not something governments in low- and middle-income nations always do. More attention needs to be paid to rural infrastructure, including better roads and telecommunications, so that poor farmers can more easily market their crops and obtain information about prices. Tube wells for safe drinking water, improved irrigation, agriculture extension services, research and development on crop varieties, and expanded access to credit are other interventions that can contribute to the improved performance of the rural economy and a decline in rural poverty (see Chapters 16 and 17 for further discussion of agricultural development and its impact on the poor.)

Spending more on the poor, whether on their education or healthcare, on the infrastructure on which they depend, or in the rural economy where most of the poor live, means spending less on other groups in society. Reforming the allocation of government expenditures can be both pro-poor and pro-growth. But it is bound to encounter resistance as interest groups act to maintain government expenditures they have come to expect. Even more controversial are proposals to redistribute assets, most often land, to the poor.

Explanations for the economic growth and success at poverty alleviation among East Asian countries, including China, Korea, and Taiwan, often point to the role

[37]Martin Ravallion and Gaurav Datt, "Why Has Economic Growth Been More Pro-Poor in Some States of India than Others?" *Journal of Development Economics* 68 (2002).

played by earlier policies of land redistribution. These were fairly radical interventions that included the expropriation of land by the state with minimal or no compensation to owners. Land reforms in East Asia occurred during times of extreme political upheaval, whether social revolution in China or the end of foreign occupation in Korea. Despite the ultimate success of these economies and the role redistribution of agricultural land played in these nations' subsequent economic development, land reforms and other types of asset redistribution receive less support today. They often are seen as politically difficult, if not unfeasible, by both national governments and multilateral institutions like the International Monetary Fund and World Bank. Land reforms also have gone badly in many countries. Zimbabwe's land reform is but one element of the destructive actions taken by the government of Robert Mugabe that turned Zimbabwe from a net food exporter into a nation dependent on food aid. Zimbabwe's experience is an extreme example; it demonstrates that solely redistributing assets offers little promise of alleviating poverty. A much broader set of complementary policies is needed to permit the poor to take advantage of any increase in their assets. But even when such policies are in place, asset redistribution is often politically too difficult to play much of a role in a pro-poor growth strategy.

INCOME TRANSFERS AND SAFETY NETS

Every nation has individuals who are poor because they lack economic opportunity or are beyond the reach of the market. The latter include those who are too old, too young, or too sick to work and are without family networks to care for them. There are also situations in which systemic shocks, whether due to natural catastrophes or economic crises, require government action because the marketplace cannot resolve them.

Situations of chronic poverty tend to call for income transfers. These may take the form of cash grants or food pricing/distribution programs. The challenge facing all such programs is making them cost effective and having them reach the target population and not "leak" to higher-income groups. Food price subsidies, for example, aid the poor but also subsidize the purchases of the nonpoor. The net result can be an unsustainable fiscal burden on the budget.

Conditional cash transfers (CCTs) increasingly are employed in all regions to address both current and future poverty. These programs provide cash payments to eligible households and in return families must satisfy program goals, usually enrolling children in school and requiring them to attend at least 80 to 85 percent of school days. Some CCTs also require families to visit health clinics for periodic checkups, vaccinations of young children, prenatal care for expecting mothers, and other services. Cash payments address current poverty. By encouraging school enrollment and better healthcare, the programs simultaneously increase the human capital of children with the goal of reducing the transmission of poverty from one generation to the next.

Two of the most well known programs are Bolsa Familia Program (family allowance) in Brazil and Oportunidades (opportunities) in Mexico (originally called PROGRESA: Programa de Educación, Salud y Alimentación). Oportunidades uses geographic targeting based on census data to identify the poorest rural areas and urban blocks. It then uses household surveys on income and assets ("Do you have a dirt or cement floor?" "Do you own a hot-water heater?") in the identified areas to direct benefits to eligible households. This approach reduces the leakage of program benefits to the nonpoor. Transfers are made to the female head of a household (usually the mother, but sometimes a grandmother), consistent with evidence that women are more likely than men to spend financial resources on their children and to improve the well-being of their families. Banks rather than government agencies handle the transfer of funds to individual recipients, reducing the number of intermediaries involved and cutting down on corruption. A high percentage of the program's budget goes directly to beneficiaries who spend the money on local businesses, often small in scale and located within the poor Mexican communities where recipients live. From its inception, Oportunidades set up systems for data collection and rigorous program evaluation so that the impact of CCTs could be determined and verified.[38]

Bolsa Familia offers eligible households about $13 a month per child, helping to lift many families above Brazil's poverty line. The program has been especially effective in rural areas but somewhat less so in urban ones, where costs are higher and the $13 does not go as far. Problems of drug addiction, violence, child labor, and family breakdown also compound the causes of income poverty in urban areas. Bolsa Familia today covers families representing almost 50 million Brazilians, about one quarter of the population, but the aggregate cost is small, about 0.5 percent of Brazil's GDP and 2.5 percent of total government expenditures. This modest amount is seen as having played a role in reducing poverty *and* inequality in Brazil.

Social safety nets are similar to income transfers but, in their design, recognize that household poverty often is transitory rather than chronic. Panel data that trace individuals or households over time find that there are fewer families who are always poor than there are families who are poor some of the time. This outcome is true in low- and middle-income economies as well as in high-income settings. The transitory nature of poverty does not minimize the hardship families endure nor does it imply that those living on amounts just over a nation's poverty line are satisfying their material needs. But it does recommend designing policies that help individuals and households when income and consumption shortfalls occur, not in a permanent fashion. Public employment schemes are one example.

[38]Because of the linkage with schooling, we discuss CCTs again in Chapter 8; see especially Box 8–3. A comprehensive review of CCTs is provided by Ariel Fiszbein and Norbert Schady, *Conditional Cash Transfers: Reducing Present and Future Poverty* (Washington, DC: World Bank, 2009).

The Employment Guarantee Scheme (EGS) in the Indian state of Maharashtra and the Trabajar program in Argentina were designed to ensure poor people a source of income and reduce the variability of their incomes. The EGS guarantees employment within a few weeks of the individual's request and provides a job relatively close to a person's home. These jobs involve public works, such as road construction and repair, irrigation systems, and prevention of soil erosion. To solve the targeting problem, wages in public employment schemes must be kept low relative to market alternatives. By offering the prevailing market wage for unskilled rural labor, EGS encourages self-targeting, which increases the likelihood that those who choose to participate are individuals the program is intended to benefit. This maintains the cost-effectiveness of the scheme by minimizing the number of nonpoor people seeking these jobs. There also is a ceiling on the number of days per year such employment will be available.

To encourage self-targeting, Trabajar in Argentina offered a monthly wage set at 75 percent of the average monthly earnings of workers in the bottom 10 percent of households living in and around Buenos Aires. Evaluation of these public employment schemes finds that they were well used and well targeted. ESG provides about 100 million person-days of employment, varying both by season and by year. Participation falls during the busy season in agriculture, confirming that poor families use the program to counter the variability in monthly incomes. The majority of participants in the ESG and Trabajar programs were from the lowest income deciles. Participants realized significant increases in their incomes, lifting many above poverty.[39] The experience with ESG is one of the reasons why India launched a similar nationwide initiative, the Mahatma Gandhi National Rural Employment Guarantee Act, which promises up to 100 days of unskilled manual labor on rural public works projects per family per year. Pay is equal to the official minimum wage rate for agricultural labor.

GLOBAL INEQUALITY AND THE END OF POVERTY

Our discussion of inequality and poverty reduction has so far focused on nations. We looked carefully at levels and trends in inequality within nations and pro-poor policies that governments might pursue. Most policy making occurs at a national level, so this focus is warranted. But there also is a global dimension to issues of inequality and poverty. The gap between rich and poor across countries tends to be greater than it is within most nations. Should anything be done to change this outcome? Is

[39]World Bank, "Principles of Successful Workfare Programs," *World Development Report 2000/2001: Attacking Poverty* (Washington, DC: World Bank, 2001), Box 8.9.

reducing world poverty a global goal? Is there a role for actions that go beyond the nation-state? There is considerable debate over the answers to these questions.

A simple way of portraying global inequality is to divide the world into the high- versus low- and middle-income economies. In 2010, the high-income nations accounted for 16 percent of world population and consumed 55 percent of world output. The low- and middle-income nations represented the rest: 84 percent of world population and 45 percent of world output. This level of global inequality is comparable to that in Brazil and South Africa, two of the world's most unequal nations. This degree of global inequality is not a recent outcome. About 30 years ago the results were similar: in 1980 the high-income nations represented 18 percent of world population and consumed 62 percent of world output.[40] This simple division of the world provides a fairly reliable snapshot of global inequality, but it ignores important differences between countries within each of these groups. In addition, not everyone in a high-income economy is rich, nor is everyone in a low-income nation poor. To resolve these problems, three measures often are used in debates over whether the world is becoming a more equal or unequal place.

One approach is to define **international inequality** by comparing average incomes across countries. For some questions, it is appropriate to rely on this method. For example, the convergence debate, discussed in Chapter 4, asks whether there is a tendency for the income levels across nations to converge over time. Comparing mean incomes is warranted here. This approach treats each of the world's nations equally, whether the Caribbean nation of Dominica, with a population of about 75,000, or China with more than 1.3 billion people. But for a discussion of human welfare, treating each nation the same without regard to its population seems less warranted. An alternative index of international inequality is to weight each nation's mean income by its population. A population-weighted measure of international inequality is better suited for some questions but still leaves a key issue unresolved. By multiplying the average level of income by population, no account is taken of the domestic distribution of income; everyone in China or Dominica is assumed to have the same income as everyone else in the nation. To avoid this problem, one needs a measure of global inequality that compares the income or consumption of each individual regardless of where that person lives. Such a measure describes inequality among all individuals, not just among all nations. It is not surprising that the level of inequality and its trend varies according to which of these definitions is employed.

Figure 6–9 compares two estimates of international inequality, one weighted and the other unweighted by population, from 1961 to 2008. Gini coefficients are

[40]The share of world output refers to gross national income (GNI) measured in terms of PPP. If GNI is measured at market exchange rates, the high-income nations' share of GNI was closer to 80 percent in 1980 and 70 percent in 2010.

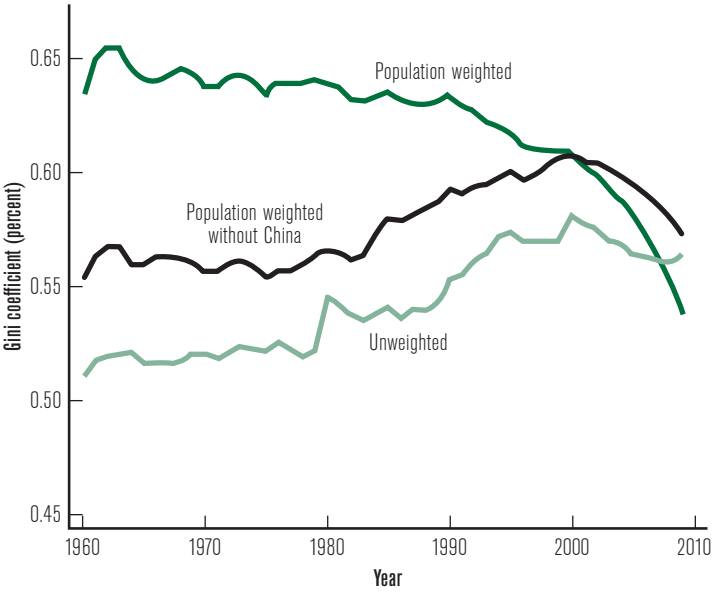


FIGURE 6–9 Trends in International Inequality, 1961–2008

Source: Figure 1 from The International Bank for Reconstruction and Development. The World Bank: *Policy Research Working Paper 5061* by Branko Milanovic, 2009. Reprinted with permission.

estimated annually for each of these series.[41] The unweighted trend (counting each country the same) shows little systematic change until the early 1980s, after which there is an almost continuous rise in inequality until 2000, when it begins to decline. The weighted series (giving larger countries more influence) shows something different. First, it suggests a higher level of inequality until the very end of the series. Second, the time trend never systematically rises and starts to decrease earlier and more rapidly, beginning around 1990.

The differences in trends between the two series are easy to explain. The rise in the unweighted measure of international inequality between 1980 and 2000 reflects the regional growth trends we first encountered in Chapter 2. The 1980s were a "lost decade" for Latin America, with debt crises followed by economic stagnation and decline. In the 1990s, the transition experience in eastern Europe and central Asia was accompanied by steep declines in income. These reversals of fortune contrast with positive growth in the high-income economies and parts of Asia and account for rising international inequality. Decreasing international inequality in the 2000s is evidence of faster growth in most low- and middle-income economies.

[41]Much of this discussion of trends in global inequality draws from Branko Milanovic, *Worlds Apart: Measuring International and Global Inequality* (Princeton, NJ: Princeton University Press, 2005); and Branko Milanovic, "Global Inequality Recalculated: The Effect of New 2005 PPP Estimates on Global Inequality," Policy Research Working Paper 5061 World Bank, Washington, DC, September 2009.

The population-weighted measure of international inequality shows a different trend. For four decades, the measure declined, picking up speed after 1990. Given the construction of this measure, what happened in heavily populated nations drives the outcome. China's experience explains much of the observed trend. This is evident when one compares the population-weighted trend in Figure 6–9 that includes China versus the one that excludes it. China's post-1978 economic transformation quickly moved the nation from low- to middle-income status. This closed some of the gap between China and the high-income nations, especially the heavily populated United States, but increased the income gap between China and the remaining low-income nations. On net, the Chinese population's move toward the middle of the world distribution of income resulted in a significant decline in the population-weighted Gini ratio. Rapid growth since 2000 in India, another heavily populated nation, accounts for the decrease in the population-weighted trend in international inequality excluding China. Rapid income growth in China and India helps explain why by 2008 the population-weighted measure finally falls below the unweighted measure.

Population-weighted international inequality assigns all citizens within a nation the same level of per capita income or consumption—that is, it assumes "perfect income equality" within every nation. This is a poor assumption for measuring inequality for the world's population. Correcting this problem requires more than data on GDP per capita and population size. Household-level data are needed to assign a more precise estimate of income levels to every person in the world and to compute what might be called global inequality. There is considerable variance in the level of global inequality predicted by various studies and no consensus on trends in recent decades. But what all these studies show is that global inequality is significantly higher than population-weighted international inequality. Estimates of global Gini coefficients range from about 0.60 to 0.80.[42] China, India, and the United States, the three most populous nations in the world, contribute significantly to this level of inequality. In recent decades, all three have had economic growth accompanied by rising domestic Gini coefficients.

Measured this way, global inequality at the beginning of the twenty-first century exceeds the level of inequality of almost any nation in the world. It should come as no surprise that the world is a very unequal place, but the *degree* of inequality is striking. Your response to the evidence on the degree of global inequality may be, "Something should be done about this!" If it is, it is important to be clear on the nature of the problems created by global inequality that you wish to resolve.

Consider the following: Global inequality would be lower if, all else equal, economic growth in high-income economies were slower. Some may be in favor of slower economic growth in the United States and elsewhere, but their reasons

[42]Reviews of alternative studies can be found in Branko Milanovic, *Worlds Apart: Measuring International and Global Inequality* (Princeton, NJ: Princeton University Press, 2005), 119–27; and in World Bank, *World Development Report 2006: Equity and Development* (Washington, DC: World Bank, 2006), chap. 9.

may have little to do with global inequality or with the well-being of people in poor nations. You may reject the relentless consumerism of the United States. Americans buy ever-larger houses, fill them with a seemingly endless supply of goods, and drive bigger vehicles on increasingly congested roads. There is much in the lifestyles and consumption habits of high-income economies to criticize. But slower economic growth in rich nations, even if it led to lower global inequality, might have an adverse impact on low- and middle-income nations. If rich nations grew more slowly, their demand for goods from other nations would fall, affecting these nations' growth rates. Global inequality brought about by slower growth in high-income economies could prove a detriment to progress in reducing global poverty. There is no reason to believe that if the United States grew more slowly, Africa or some other poor region would grow more quickly. In all likelihood, the opposite would be true.

There are many ways in which global inequality could be lower today. What if income inequality in China had not risen over the past 20 years? Remember that China has had an exceptionally high growth rate for more than two decades and has been able to lift over 400 million of its citizens above the $1.25-a-day poverty line. China's economic miracle also contributed to reducing global inequality, but the effect would have been even greater if income inequality within China had not risen. Could China have grown rapidly, dramatically reduced poverty, *and* kept its level of inequality from rising? We really do not know because we cannot observe the counterfactual. But there is some chance that lower inequality within China might have required slower economic growth.

Calls for reducing global inequality may be misplaced unless proper consideration is taken of how greater equality might be achieved. Even before considering how to achieve greater global equality, the question of why to do so remains. We have encountered some of these arguments before on a national level and some are also applicable on a global level. First, on the grounds of economic efficiency, the concentration of world incomes may lower productivity growth. Market failures limit poor people from borrowing to finance worthwhile projects, and some redistribution of income could increase the return on total global investments. Second, it may be in the self-interest of higher-income households, wherever they may live, to support redistribution. Health risks, for example, cross borders faster than in the past. The rapid transmission of HIV/AIDS across continents is one example. A redistribution of world resources might mitigate the health risks facing both the poor and the rich. Another reason why it may be in the interest of high-income households to favor greater global equality is that, in an information age, people around the world are aware of the gap between rich and poor. This knowledge and frustration over feeling left out may play a destabilizing role in international affairs, affecting the interests of the rich as well as the poor. Third, a more equitable distribution may be the right goal to strive for on moral grounds. Philosophers since Plato have been writing about this subject. Especially given the size of the income gap between those on the top versus

those on the bottom, global inequality should be seen as an opportunity to address the absolute poverty that has been a central focus of this chapter.

If those in higher income quintiles are averse to high levels of inequality and willing to move toward a world of less poverty and greater equality of incomes and opportunities, what might they do? Columbia University economist Jeffrey Sachs offers both a diagnosis and a blueprint. In his book *The End of Poverty*, he writes,

> The greatest tragedy of our time is that about one sixth of humanity is not even on the development ladder. A large number of the extreme poor are caught in a poverty trap, unable on their own to escape from extreme material deprivation. They are trapped by disease, physical isolation, climate stress, environmental degradation, and by extreme poverty itself. Even though life-saving solutions exist to increase their chances for survival—whether in the form of new farming techniques, or essential medicines, or bed nets that can limit the transmission of malaria—these families and their governments simply lack the financial means to make these critical investments. The world's poor know about the development ladder: they are tantalized by images of affluence from halfway around the world. But they are not able to get a first foothold on the ladder, and so cannot even begin the climb out of poverty.[43]

Sachs proposes a global compact to end absolute poverty by 2025. As with any compact there are at least two parties: poor countries and rich ones. Poor countries are to be held accountable for their efforts to reduce poverty. Corrupt regimes and those that pursue war rather than development cannot be part of this compact. Sachs makes it clear, however, that even among low- and middle-income nations that sign onto the compact, their actions alone may not be enough to end poverty. Conditions in parts of Africa are so dire that many of the poor, lacking the basics of food, clean water, medicines, and healthcare facilities, are, in Sachs's words, "too poor to stay alive."

Given the degree of absolute poverty in the world, there is a need for rich nations to do more to alleviate poverty. Some of Sachs's suggestions meet with wide support among development economists: the need for rich countries to keep their markets open to exports from poor nations and the need for rich nations to invest in global public goods, such as basic science on combating tropical diseases and improving agricultural yields. Other elements of his global compact are more controversial: the need for better environmental stewardship by the rich nations to minimize the impact of climate change on poor nations, the need for debt forgiveness of the accumulated international debts of poor nations owed to multilateral institutions like the IMF and World Bank, and the need for a significant increase in the foreign aid from rich nations to poor ones. We explore these ideas in later chapters. Despite

[43]Jeffrey Sachs, *The End of Poverty: Economic Possibilities for Our Time* (New York: Penguin Press, 2005), p. 19–20.

disagreement about elements of this global compact to end poverty, it is hard to argue against one of its central ideas: Rich nations have a critical role to play in reducing global poverty.

SUMMARY

- The number of people in poverty in a given country depends on both the level of per capita income or consumption and its distribution. Distributional outcomes are described using size distributions, which report inequality in terms of shares going to each population quintile ranked from poorest to richest; the Lorenz curve, which offers a geometric portrait of inequality; and the Gini coefficient, which provides a single summary statistic.
- Income and consumption inequality exhibits significant regional variation. Latin America and sub-Saharan Africa have relatively high levels of inequality; eastern Europe and central Asia, South Asia, and the high-income economies generally have low levels; and inequality in East Asia generally falls in the medium range. These regional patterns seem to have less to do with the *level* of per capita income than with underlying historical and political determinants as well as factor endowments.
- According to World Bank studies, in 2005 nearly 1.4 billion people, representing 25 percent of the developing world's population, lived below the international poverty line of $1.25 a day, sometimes referred to as *absolute poverty* or *extreme poverty*. The number of people living in absolute poverty declined by 520 million between 1981 and 2005. Most of this decline happened in East Asia, especially in China. With generally strong economic growth in many low- and middle-income economies since 2005, absolute poverty probably has declined since then.
- Because world population grew significantly from 1981 to 2005, success at poverty reduction can also be measured in relative terms. Estimates of the head-count index suggest that the incidence of poverty fell significantly in East and South Asia. In sub-Saharan Africa, the level remained at over 50 percent. Most of the world's poor still live in East and South Asia, although both the poverty rate (head-count index) and the severity of poverty (poverty gap) are highest in sub-Saharan Africa.
- Poverty tends to be greater in rural than in urban areas. Racial and ethnic minorities often face higher poverty rates. Women are discriminated against throughout the world. They earn lower wages on average, often are denied inheritance or the right to own land, and traditionally have received less education than men. However, because household resources tend to

be shared as a result of intra-household distribution decisions, there is less evidence that income poverty rates for women are higher than for men.

- Empirical studies confirm that economic growth tends to be good for the poor. Across countries, the incomes of the bottom 20 percent grew at the same rate as GDP per capita. This is an average tendency, and in some cases and in some time periods, the average income of the poorest quintile *fell* despite increases in GDP per capita.
- *Pro-poor growth* refers to a development strategy that combines more rapid economic growth with increased opportunities for the poor to participate in the economy. Many economists believe that market-friendly policies, including maintaining fiscal discipline, reducing price inflation, and increasing economic openness, serve the interests of the poor. But these interventions alone are not enough; some of the poor will benefit and others will be hurt.
- Growth strategies designed to include the poor may have the greatest potential to promote pro-poor growth. Because poverty tends to be disproportionately rural, growth strategies that include agricultural development may be particularly important to pursue.
- Governments must also invest in the education and health of the poor and in the infrastructure on which the poor rely. Conditional cash transfer programs, such as Bolsa Familia in Brazil and Oportunidades in Mexico, have proven cost-effective in reducing current poverty and making investments in children that should reduce future poverty. Social safety nets, such as guaranteed employment schemes, are designed for those temporarily unable to take advantage of the opportunities markets provide.
- Just as we can measure inequality within nations, it is possible to estimate the degree of global inequality. The most comprehensive measure of global inequality compares the income (or consumption) of each individual regardless of where that person lives. Estimates of this measure report Gini coefficients of global inequality of 0.60 to 0.80, values higher than for almost any single nation. Given both the level of absolute poverty in the world and the degree of global income inequality, it is important to consider the steps rich nations can take to help poor nations.



Population

In 1973, marking the completion of his first term as president of the World Bank, Robert McNamara wrote *One Hundred Countries, Two Billion People,* which brings together his basic views on economic development. There is no ambiguity about McNamara's beliefs. He writes, "The greatest single obstacle to the economic and social advancement of the majority of the peoples in the underdeveloped world is rampant population growth."[1]

Before coming to the World Bank, McNamara was a professor at the Harvard Business School, president of the Ford Motor Company, and secretary of defense in the Kennedy and Johnson administrations. A man of incredible intellectual reach, McNamara understood much about the role of capital accumulation and technological change and their contributions to economic growth. But he maintained the view that rapid population growth was a threat that would have "catastrophic consequences" unless dealt with. "The underdeveloped world needs investment capital for a whole gamut of productive projects, but nothing would be more unwise than to allow these projects to fail because they are finally overwhelmed by a tidal wave of population."[2]

The "Two Billion People" in the title of McNamara's book refers to the early 1960s when the world's population was close to 3 billion, with 2 billion, or about two-thirds of the total, living in the developing nations. Since then the world has grown,

[1]Robert McNamara, *One Hundred Countries, Two Billion People* (New York: Praeger, 1973), p. 31.
[2]MacNamara, *One Hundred Countries,* p. 31.

reaching 7 billion in 2011. Over 5.8 billion of these people reside in low- and middle-income economies.

Despite this massive increase in numbers of people, the views of James Wolfensohn, World Bank president from 1995 to 2005, could not be more different from those of his predecessor. Wolfensohn's speeches and publications reveal only limited reference to population growth. Most often these references indicate how population growth exacerbates another problem, whether improving access to clean water or making a dent in the numbers suffering from absolute poverty. The same is true for Wolfensohn's successor, Robert Zoellick. After his first 100 days in office, Zoellick delivered a speech identifying six strategic themes to meet global challenges. Reducing population was not one of them; in fact, "population growth" is not mentioned even once.[3]

Was McNamara unnecessarily alarmist? Or did Wolfensohn and Zoellick make irreversible mistakes by not focusing on rising population numbers? Was the United Nations similarly wrong in not including reductions in population growth as one of the original millennium development goals (MDGs; introduced in Chapter 2)? We begin to address these questions by reviewing the world's population history and exploring the demographic transition that has characterized today's high-income nations and increasingly the low- and middle-income nations as well. We also consider population projections for the future before turning our attention to the complex relationship between population growth and economic development. This relationship is viewed both at an aggregate level and at the level of individual families making decisions about how many children to have. We conclude by reviewing the options nations face in pursuing policies to limit the size of their populations.

A BRIEF HISTORY OF WORLD POPULATION

Anthropologists debate when our first ancestors appeared. For our purposes, we do not have to go that far back in time. We might begin at the end of the last Ice Age, about 13,000 years ago, when humans on all continents were still living as hunter–gatherers, or 12,000 years ago with the first signs of agricultural settlements, or 7,000 years ago with the first indications of urbanization. For most of the thousands of years since then, population growth has been close to zero, with annual births roughly offsetting annual deaths. Opportunities for human survival slowly improved, and by 1 C.E. the world's population is estimated at about 230 million.[4] To gain some perspective on this number, today Indonesia by itself has a population that size.

[3]Robert Zoellick, "An Inclusive & Sustainable Globalization," speech to the National Press Club, Washington, DC, October 10, 2007.

[4]Angus Maddison, *The World Economy: Historical Statistics* (Paris: Development Centre of the Organisation for Economic Co-operation and Development, 2003).

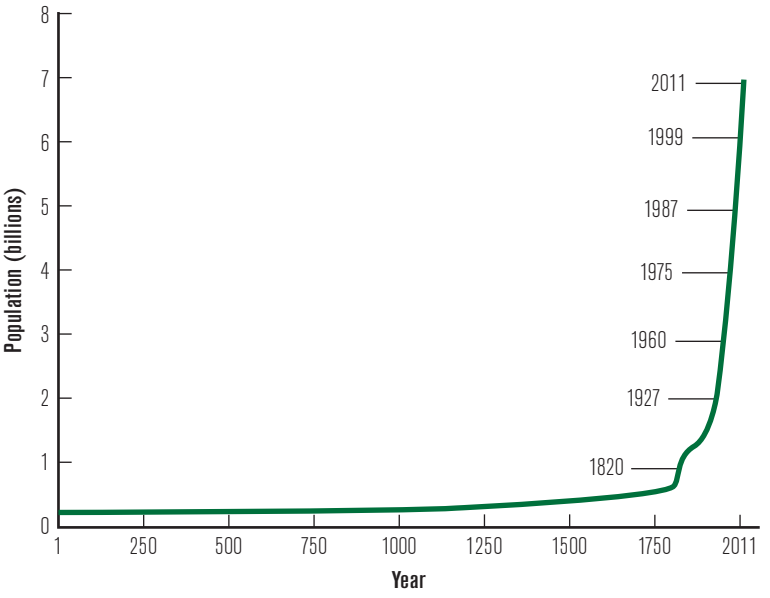


FIGURE 7–1 World Population Growth through History: Years Needed to Add 1 Billion More People

Source: Angus Maddison, "Statistics on World Population, GDP, and Per Capita GDP, 1–2008 AD," www.ggdc.net/MADDISON/Historical_Statistics/vertical-file_02-2010.xls.

Figure 7–1 charts world population from 1 C.E. through 2011. The figure makes it abundantly clear that population growth is a recent and unprecedented event in human history. From the beginning of human settlements, it took more than 10,000 years for the world's population to reach 1 billion in 1820. But the next billion were added in only about 110 years and, for the last four decades, the world added 1 billion people every 12 to 15 years. No wonder McNamara was so worried.

We have a broad understanding of how world population went from a period of relatively small and stable population numbers to the 7 billion (and still growing) that inhabit the planet today. The introduction of settled agriculture revolutionized the earth's capacity to sustain human life. During the years leading up to the Industrial Revolution of the late eighteenth and early nineteenth centuries, the food supply grew and became more reliable. The death rate fell, life expectancy increased, and population growth gradually accelerated. This growth, however, was set back at intervals by famines, plagues, and wars, any of which could wipe out as much as half of the population in a given area. As late as the fourteenth century, the black death (bubonic plague) killed one-third of the population of Europe. Despite these catastrophic events, by 1800, the world's population had grown to almost 1 billion, implying an annual growth rate of a mere 0.08 percent between 1 C.E. and 1800.

The Industrial Revolution, which marked the start of modern economic growth, further expanded the earth's **population-carrying capacity**. Innovations in agriculture matched innovations in industry, permitting labor to be transferred to industry while the productivity of the remaining agricultural laborers rose quickly enough to feed the growing urban population. Transcontinental railroads and fast, reliable ocean shipping further boosted world food output in the late nineteenth century, making it possible to grow more basic foodstuffs in the areas best suited for this activity and get supplies to food-deficit areas in record time in emergencies. Famines, especially in what are now the high-income nations, decreased in frequency and severity. Food prices fell. Meanwhile, modern medicine, sanitation, and pharmaceutical production began to develop. All these factors helped reduce the death rate and accelerate population growth. By 1945, the population of the world was slightly less than 2.5 billion, meaning that global population grew by 0.6 percent per year between 1800 and 1945.

After World War II, there were further dramatic improvements in food supply and disease control. Techniques introduced in the developed countries during the preceding era spread throughout the globe. The result was a veritable revolution in falling death rates and rising life expectancy in both the developed and the developing world. Plummeting death rates in many areas raised population growth rates to levels the world had never known before. Reference to a worldwide **population explosion**, unthinkable for most of human history, became commonplace in the 1960s and 1970s as the world's population grew to 3 and then 4 billion. World population passed 5 billion in 1987, 6 billion in 1999, and 7 billion early this decade; almost all this growth has occurred in the developing countries. Between 1945 and 2009, the growth in world population averaged a historically unprecedented rate of 1.6 percent per annum, twenty times faster than the estimated 0.08 percent between 1 C.E. and 1800.

THE DEMOGRAPHIC TRANSITION

The relationship between annual births and deaths determines population growth. Figure 7–2 depicts the basic stages in the **demographic transition** in a now-developed nation, Finland. Finland's experience is typical of what most high-income countries have experienced over the past several centuries. The top line in Panel a refers to the **crude birth rate**, the number of live births per year per 1,000 people. The lower line refers to the **crude death rate**, the number of deaths per year also per 1,000 people. The difference between these two rates, the excess of births over deaths, is the rate of **natural increase** in the population, which often is expressed as a percentage. For the world as a whole, the natural increase in the population equals the population growth rate; for an individual country, population growth is the difference between the rate of natural increase and net migration.

For hundreds of years before the eighteenth century, Finland was in the first stage of the classic demographic transition. High birth rates were matched by high death rates and the natural increase in population was close to zero. In some years,

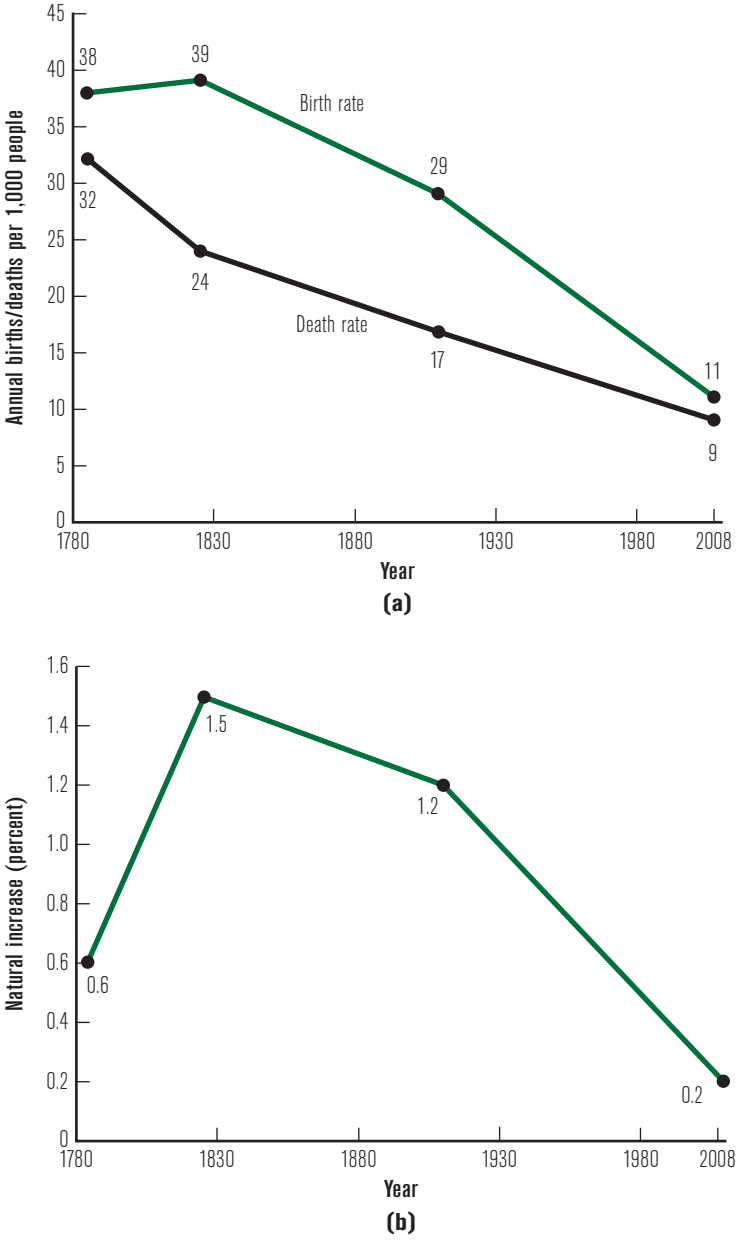


FIGURE 7–2 Demographic Transition for Finland (1785–2008)

Source: Arthur Haupt, Thomas Kane, and Carl Haub, *PRB's Population Handbook* 6th edition (Washington, DC: Population Reference Bureau, 2011).

births exceeded deaths; in other years, the opposite occurred. By 1785, before the start of industrialization and the first year for which data are available, Finland still had high birth and death rates, but the gap between the two resulted in a 0.6 percent rate of natural increase. If this rate had been maintained, it would have taken 117 years for Finland's population to double.

By the early 1800s, Finland was in the second stage of the demographic transition, characterized by death rates declining more quickly than birth rates. This was before the advent of modern medicine and public health interventions; so much of this decline can be attributed to improved nutrition as food supplies became more abundant and less variable. With no change in the birth rate and a fall in deaths, the rate of natural increase in the population reached 1.5 percent by 1830. At this rate, the population would double in just under 50 years. After 1830, death rates continued to fall but not as quickly. At the same time, birth rates started to drop as Finnish families decided to have fewer children. Note that they did so in the absence of modern forms of contraception. Finland was now in the third stage of the demographic transition, with birth rates falling more quickly than death rates. By 1915, Finland's natural increase in population had slowed to 1.2 percent.

By 2008, Finland reached the fourth stage of its demographic transition in which its population growth had fallen close to zero, this time with low birth and death rates. The nation had gone from relatively high birth and death rates producing a low rate of population growth before 1785, through a period of more rapid natural increase in the nineteenth and early twentieth centuries, back down to a very low rate of population growth characterized by both low birth and death rates. Today, Finland has a population of 5 million, growing at 0.2 percent per annum. Abstracting from changes due to migration, at this rate Finland's population would require 350 years to double in size. But Finland's population probably will never double again. As has happened elsewhere in Europe, birth rates may fall even further and the crude death rate will rise as the population ages. In coming decades, Finland may even experience negative population growth, just as Germany and Italy already are.

Based on the few observations available, Finland's rate of natural increase never exceeded 1.5 percent. Compare this to the experience of the developing countries where, as a group, the rate of natural increase reached 2.5 percent during the 1960s. Even today, Pakistan's population is growing by 2.1 percent, Jordan's by 3.2 percent, and Madagascar's by 2.9 percent. Why draw so much attention to a 1- or 1.5-percentage point difference in population growth rates? Because doubling time drops from around 50 years at a growth rate of 1.5 percent to only 28 years at 2.5 percent to 22 years at 3.2 percent. Seemingly small differences in growth rates provide societies with even less time to achieve a demographic transition before population levels increase by substantial amounts.

Clearly, there are major differences between the historical experience of today's industrialized nations and the contemporary demographic transition of the developing nations. Figure 7–3 portrays the movement of crude birth and death rate for the less-developed nations as a group from 1950 to 2005. Panel b charts the corresponding rate of natural increase over the same years. By 1950, the rate of natural increase in the developing nations was 2.1 percent, far larger than the rate ever experienced by today's industrialized countries. The higher rate was the result of somewhat higher birth rates than historically had been the case in today's high-income nations due,

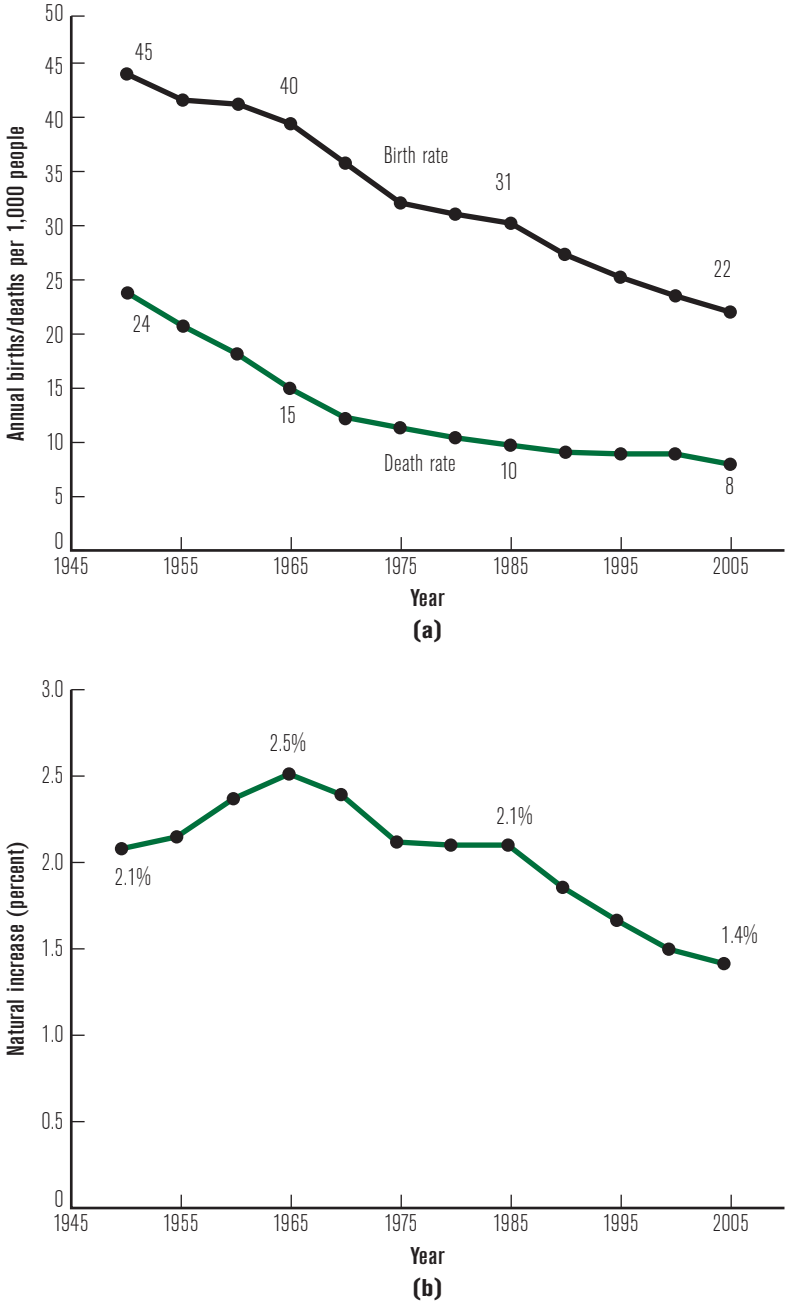


FIGURE 7–3 Demographic Transition for Less-Developed Regions (1950–2008)

Source: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, *World Population Prospects: The 2002 Revision* and *World Population Prospects: The 2008 Revision.*

in part, to earlier age at marriage. Of even greater significance, the decline in death rates in the developing nations began at much lower levels of per capita income and fell much faster than it had historically in the now-developed countries. The transfer of public health interventions and other medical advances, including vaccines and improvements in water and sanitation services, from the developed to the developing world contributed to this outcome.

Between 1950 and 1970, death rates in the developing world continued to drop faster than birth rates, and the natural increase in populations rose. It peaked at around 2.5 percent in the late 1960s, shortly before McNamara warned of the catastrophic consequences of rapid population growth. With little evidence of falling birth rates, it is not surprising that McNamara was so alarmist about population trends. But a demographic transition already was under way. Between 1950 and 2005, the crude birth rate fell from 45 to 22 per 1,000, and even though death rates also fell, from 24 to 8 per 1,000,[5] the natural increase of the developing nations slowed to 1.4 percent, still high by historical standards but with an unmistakable trend. In all regions of the world and in almost every nation, women started having fewer children and **total fertility rates (TFRs)** fell (Box 7–1). These trends help explain why the World Bank and other multilateral organizations appear less concerned about population growth today. The world population is still growing and will continue to do so for several more generations, but the *rate of growth* has slowed and population *explosion* no longer seems an appropriate metaphor.

THE DEMOGRAPHIC SITUATION TODAY

In 2009, world population exceeded 6.7 billion. Only 16 percent resided in the high-income countries, 84 percent lived in the low- and middle-income nations (Table 7–1). Almost 3 out of every 10 inhabitants of the planet lived in East Asia. China alone accounted for just under 20 percent of world population. South Asia, dominated by India's more than 1 billion inhabitants, represented another 23 percent of the world's people. Sub-Saharan Africa accounted for 12 percent of world population, but this share is expected to rise in the future. At 2.4 percent, sub-Saharan Africa had the most rapid rate of natural increase in population of any region.

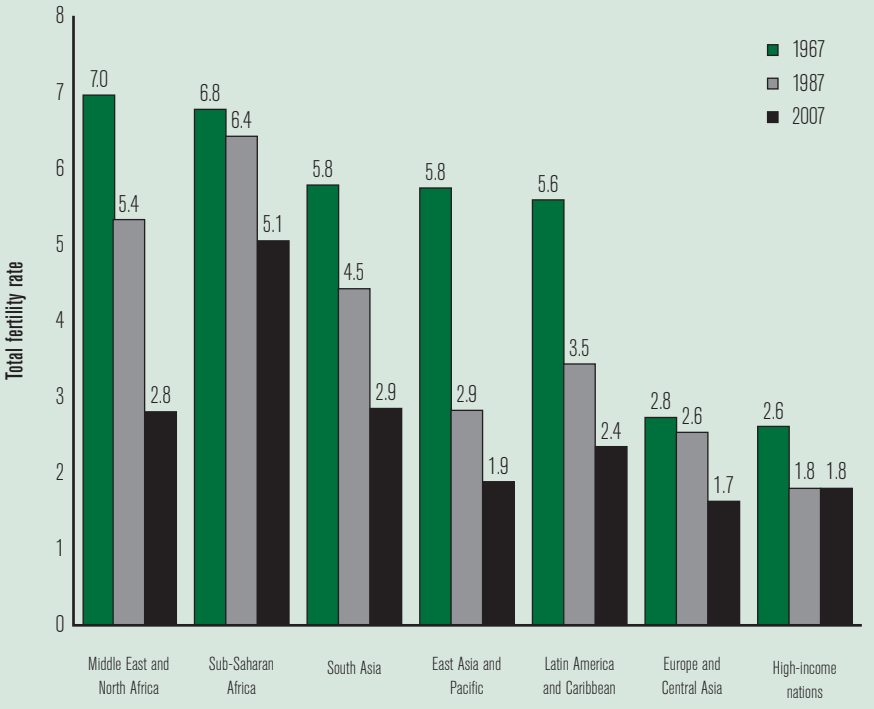
The situation in the high-income countries could not be more different from that in sub-Saharan Africa. With aging populations, crude death rates are likely to rise, and with declining fertility, crude birth rates will fall. Rates of natural increase already have slowed to 0.4 percent. Almost all *population growth* in the high-income countries is attributable to immigration, with openness to immigrants at a higher

[5]Careful inspection of Figures 7–2 and 7–3 reveals that the death rate in 2008 in Finland, at 9 per 1,000 people, is almost identical to the death rate in 2005 in the less-developed regions. This does not mean that life expectancy and the overall survival chances in the two places are the same. If one compares age-specific mortality rates, survival probabilities in Finland exceed those of the developing countries for every age group, especially for infants. But, in the aggregate, because Finland has a much older population than the developing world, the crude death rates, coincidentally, average to almost the same number.

BOX 7-1 TOTAL FERTILITY RATES

Another way to assess population trends is to examine fertility behavior. One of the most common measures of fertility is the total fertility rate (TFR). TFR sometimes is thought of as a measure of the average number of children a woman will bear, but this is not entirely correct. TFR is a *synthetic* measure. It sums the age-specific fertility rates of women in a given year, where age-specific fertility rates refer to the average number of children born to women of a specific age (usually, 15–19, 20–24, . . . , 40–44). In other words, TFR is the number of children the average woman would have in her lifetime if age-specific fertility rates remained constant. But these rates change over time as younger women delay having their children and as women have fewer children over all. TFR is a reliable indicator of the number of children women *currently* are having, and trends in TFR reveal a great deal about the world's demographic transition.

The figure below charts TFRs over the past 40 years for the World Bank's seven geographical regions. In 1967 in Africa, Asia, and Latin America, TFR was 5.5 children or more. Only the high-income economies and the nations of Europe and central Asia had reached low TFRs of fewer than 3 children. Over the next four decades, and at varying speeds, fertility rates fell in every region. In East and



South Asia and in Latin America and the Middle East, TFR fell by half or more. Sub-Saharan Africa's demographic transition took longer to begin, but even in this poorest part of the world, TFR, now at 5.1 children per woman, has fallen by more than one child since 1987. Most demographers expect the decline in Africa's TFR to follow that of other regions and the rate to continue to fall.

TABLE 7-1 Levels and Trends in the World Population, 2009

	TOTAL POPULATION		POPULATION GROWTH		
	NUMBER (MILLIONS)	PERCENT OF TOTAL	BIRTH RATE (PER 1,000)	DEATH RATE (PER 1,000)	NATURAL INCREASE (PERCENT)
World	6,775	100	20	8	1.2
Income category					
Low income	846	13	34	11	2.3
Middle income	4,813	71	19	8	1.1
High income	1,117	16	12	8	0.4
Region					
East Asia	1,944	29	14	7	0.7
South Asia	1,568	23	24	7	1.7
Sub-Saharan Africa	840	12	38	14	2.4
Latin America	572	8	18	6	1.2
Europe and Central Asia	404	6	15	11	0.4
Middle East and North Africa	331	5	24	6	1.8

Source: World Bank, "World Development Indicators Online," http://databank.worldbank.org.

level in Australia, Canada, New Zealand, and the United States than in Europe and Japan. Total fertility rates in many of the high-income economies already have dropped below replacement levels. Women must have at least two children each during their childbearing years for the population to replace itself. Allowing for infant and child mortality, demographers estimate that the replacement level of total fertility is around 2.1. In countries such as Germany, Italy, and Japan, where the TFR is below the replacement level and immigration rates also are low, policy makers are concerned about a population implosion rather than a population explosion. These societies can anticipate smaller populations overall as well as an increasing number of elderly people becoming dependent on a shrinking labor force.

Below replacement level, TFR also characterizes eastern Europe and central Asia, where the rate of natural increase is only 0.4 percent. East Asia's TFR has dipped

below replacement levels (1.9 in 2007). By contrast, neither Africa, Latin America, nor the Middle East has progressed as far along its demographic transitions, implying faster growth in population and an increasing share of the world's total.

THE DEMOGRAPHIC FUTURE

When extrapolated into the future, even modest population growth rates generate projected total populations that seem unthinkable. Continued growth at the world's 2010 rate of natural increase, 1.2 percent, would bring population to 11 billion by 2050 and 26 billion by 2100. This type of projection, beloved by some popular writers, is frightening to many. It is hard to imagine life in a world with three to four times as many people as there are today. How will this expanded population live? How will the globe's finite supplies of space and natural resources be affected?

While a significant increase in world population can be expected during your lifetime, it is unlikely to double. Linear extrapolations of current trends are badly misleading because after accelerating for more than two centuries, world population growth has been slowing down for the past four decades. This trend is likely to continue, but no one can be certain how quickly the transition to lower fertility levels will be or whether there will be any reversals in recent trends.

Every few years the United Nations produces forecasts of future population growth.[6] These projections offer a variety of scenarios based on a similar set of assumptions about declining mortality but using different assumptions about the amount and speed of fertility declines. Figure 7–4 combines estimates of the world population from 1820 to 2000 with the United Nations' low-, medium-, and high-population projections for the twenty-first century. The high scenario assumes the world TFR does not reach replacement levels, world population reaches 10.8 billion people by 2050, and continues to expand to over 15 billion by 2100. The medium scenario has the world reaching slightly less than replacement levels by 2050. Under the medium variant, world population rises to over 9 billion by 2050. Population growth then continues to slow throughout the century, leveling off at around 10 billion in 2100. The low scenario assumes a future TFR well below replacement levels. Under these assumptions world population hits a maximum of 8.1 billion shortly before 2050 and declines to 6 billion by the end of the century. The United Nations offers the three variants as alternative projections without assigning probabilities to which outcomes are more likely.

What all three scenarios share in common is the projection that world population will continue to grow over the next 50 years, in other words, over the adult lifetime of a university student today. The scenarios differ in the size of the projected

[6]In addition to the United Nations, the International Institute for Applied Systems Analysis in Austria, the U.S. Census Bureau, and the World Bank produce independent population projections. A useful introduction to how such projections are made appears in Brian O'Neill and Deborah Balk, "World Population Futures," *Population Bulletin* (Population Reference Bureau) 56, no. 3 (September 2001).

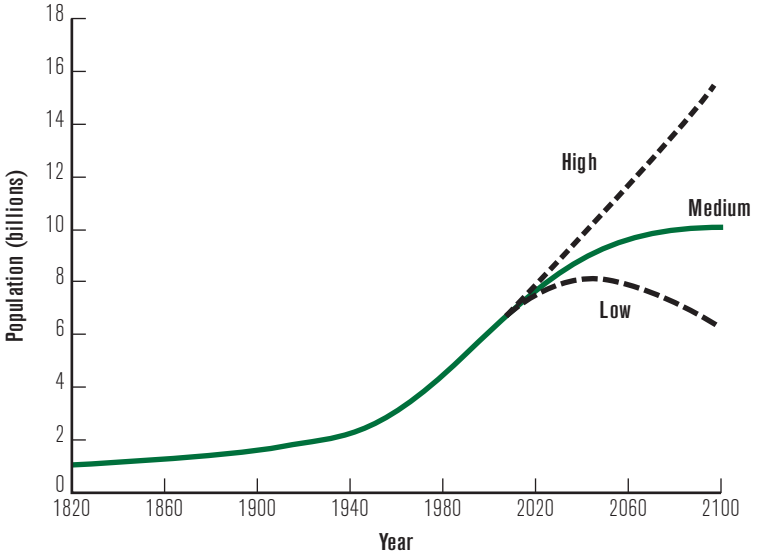


FIGURE 7–4 World Population Historical Trends and Projections

UN population projections for 2000–2100. The scenarios presented here represent different future levels of fertility.

Source: Angus Maddison, "Statistics on World Population, GDP, and Per Capita GDP, 1–2008 AD," www.ggdc.net/MADDISON/Historical_Statistics/vertical-file_02-2010.xls. U.N. Department of Economic and Social Affairs, Population Division, Population Estimates and Projection Section, *World Population Prospects, The 2010 Revision*, October 20, 2011, available at http://esa.un.org/unpd/wpp/index.htm; accessed February 2012.

growth within a broad range: from 1 to almost 4 billion. There are three basic reasons for this projected continued increase. The first two explain why the TFR may remain above replacement levels: (1) a desire for large families and (2) a failure to achieve the desired number of children (so-called unwanted children). The third reason is **population momentum**, a demographic concept requiring some explanation.

Populations that have been growing rapidly, as they have been especially in the developing world, have large numbers of people in, or about to enter, the most fertile age brackets. Even if all the world's couples today were to start having just enough children to replace themselves, the total population would continue to grow for many more decades before eventually leveling off. This is because today's children, who outnumber their parents, will become tomorrow's parents. Stated differently, past fertility decisions echo far into the future (Box 7–2). Population momentum alone could add several billion more people to the world's population in the twenty-first century.[7]

Another way of appreciating the demographic future is to focus on specific countries, especially those projected to be the world's most populous. Table 7–2 lists the

[7]John Bongaarts estimated that half of the projected increase in the world's population during the twenty-first century could be attributed to population momentum alone. John Bongaarts, "Population Policy Options in the Developing World," *Science* 263, no. 5148 (February 11, 1994).

BOX 7-2 POPULATION MOMENTUM

The impact of population momentum on future population levels can be illustrated by the following example of two families. The first generation of each family consists of one man and one woman. Each woman has four children over her reproductive life. This second generation includes four females and four males.

PERIOD ONE
Fertility above Replacement Level; Total Population, 12

First generation:

Second generation:



PERIOD TWO
Fertility at Replacement Level; Total Population, 16

First generation dies:

Second generation marries:

Third generation is born:



PERIOD THREE
Fertility at Replacement Level; Total Population, 16

Second generation dies:

Third generation marries:

Fourth generation is born:



In period two, the first generation dies and everyone in the second generation marries. Each woman in the four resulting couples has two children, producing a third generation of four males and four females. Even though the second generation reaches replacement levels of fertility, population momentum causes a 33 percent increase in total population, from 12 to 16, between period one and period two.

The process is repeated in period three. The second generation dies; the third generation marries and produces a fourth generation. If the third generation remains at replacement levels of fertility, the total population stabilizes. The size of the steady-state population in the final period is the result not only of the fertility decisions of the second and third generations, but of those of the first generation as well.

Population momentum also can work in reverse. If subsequent generations have lower than replacement fertility, the decline in total population could be rapid, reflecting the fertility decisions of both current and previous generations.

TABLE 7-2 Projections: The World's 10 Most Populous Nations in 2050*

	2050		2010		POPULATION INCREASE	
COUNTRY	POPULATION (MILLIONS)	RANK	POPULATION (MILLIONS)	RANK	MILLIONS	PERCENT
India	1,748	1	1,189	2	559	47
China	1,437	2	1,338	1	99	7
United States	423	3	310	3	113	36
Pakistan	335	4	185	6	150	81
Nigeria	326	5	158	8	168	106
Indonesia	309	6	235	4	74	31
Bangladesh	222	7	164	7	58	35
Brazil	215	8	193	5	22	11
Ethiopia	174	9	85	13	89	105
Congo	166	10	68	19	98	144
Total (share of world's population)	5,355 (0.56)		3,925 (0.57)		1,430 (0.55)†	

*Projections are based on the "United Nations" medium variant.

†The absolute increase in the population of the 10 most populous nations in 2050 divided by the change in total world population between 2010 and 2050.

Source: Population Reference Bureau, *2010 World Population Data Sheet,* www.prb.org/pdf10/10wpds_eng.pdf.

10 most populous nations in 2050 according to the United Nations' medium-variant projections. The table also lists the population of those countries in 2010. These 10 nations are projected to account for 5.4 billion people in 2050, over half of the projected world population at that time.

China, India, and the United States will continue to be the three most populated nations, although India will replace China in the top spot. Increases in China's population, which already is well below TFR replacement levels, will be due exclusively to population momentum; in the United States, immigration will be the primary factor for increasing numbers; and in India, both population momentum and fertility above replacement levels will contribute to a projected increase of over 550 million *more* inhabitants than the 1.1 billion who live in India today.

Just as striking are the expected population gains elsewhere in South Asia. Pakistan is projected to increase its population by 81 percent and Bangladesh by 35 percent. In sub-Saharan Africa, the Democratic Republic of the Congo, Ethiopia, and Nigeria all are projected to experience even greater percentage increases in their populations. Projections for the Congo and Ethiopia are so large as to catapult them into the top-10 category in 2050, replacing Japan and Russia, which were among the top 10 in 2010.

It must be remembered that all these estimates of population trends are only projections, which become less and less reliable the further out in time one goes, and much could happen to change these outcomes. The impact on future population size of infectious diseases, some unknown today, is difficult to predict. Civil strife and migration will play a role as well. But what these projections make clear is that we should expect significant increases in world population, almost exclusively in today's low- and middle-income nations, over the next 50 years. Given the current difficulties many of these nations face today, the challenge of achieving economic development for ever-larger numbers of people seems daunting.

THE CAUSES OF POPULATION GROWTH

Discussions of the world's population future often turn acrimonious. Questions of whether and whose populations are to be limited and by what means are sensitive. Before we can confront such issues intelligently, we must consider what is known about both the causes and the effects of rapid population growth. We must concern ourselves with the two-way relationship between the growth of population and economic development—that is, with how population growth affects economic development and how economic development affects population growth. We deal with the latter relationship first.

THOMAS MALTHUS, POPULATION PESSIMIST

The most famous and influential demographic theorist of all time was **Thomas Malthus** (1766–1834). Malthus believed that "the passion between the sexes" would cause population to expand as long and far as food supplies permitted. People generally would not limit procreation below the biological maximum. Should wages somehow rise above the subsistence level, workers would marry younger and have more children, more of whom would survive. But this situation could be only temporary. In time, the rise in population would create an increase in labor supply, which would press against fixed land resources. Eventually, through diminishing returns, food prices would rise and real wages would fall back to the subsistence level. If this process went too far, famines and rising deaths would result. Malthus did not think that the growth of the food supply could stay ahead of population growth in the long run. In a famous example, he argued that food supplies grow according to an arithmetic (additive) progression, whereas population follows an explosive geometric (multiplicative) progression.

In the grim Malthusian world, population growth is limited primarily by factors working through the death rate, what he called *positive checks*. In this deceptively mild phrase, Malthus included all the disasters that exterminate people in large numbers: epidemics, famines, and wars. These phenomena, he believed, generally constitute the operative limitation on population. Only in later editions of his famous *Essay on the Principle of Population* did he concede the possibility of a second, less drastic, category of limiting factors: "preventive checks" that work through the birth rate. Here, Malthus had in mind primarily measures of "restraint," such as a later age of marriage. Unlike latter-day Malthusians, he did not advocate birth control, which as a minister he considered immoral. Although he grudgingly admitted that humanity might voluntarily control its own numerical growth, Malthus invested little hope in the possibility.

The gloominess of the Malthusian theory is understandable when one considers that its author lived during the early years of the Industrial Revolution. In all prior history, population had tended to expand in response to economic gains. With unprecedented economic growth under way in the world he knew, what could Malthus expect except an acceleration of natural increase as death rates fell? That indeed was happening during his lifetime.

Malthus did not live to witness the next stage in the European demographic transition. The early decline in death rates was followed, with a lag, by a fall in fertility; beginning in the middle of the nineteenth century, wages began to increase dramatically in contrast to Malthus's prediction. Why did all this happen? Wages rose, despite accelerating population growth, because capital accumulation and technical change offset any tendency for the marginal product of labor to decline. Death rates fell through a combination of higher incomes (better nutrition and living conditions) and better preventive and curative health measures. The fall in the birth rate is

harder to understand. There are both biological and economic reasons to expect, as Malthus did, that fertility would rise, not fall, as income went up. Healthier, better-fed women have a greater biological capacity to conceive, carry a child full term, and give birth to a healthy infant. Also, people might marry earlier when times are good, and better-off families have the financial capacity to support more children. Why, then, do increases in income seem to lead to declines in fertility? An answer to this question must be sought in post-Malthusian demographic theory.

WHY BIRTH RATES DECLINE

Two kinds of demographic change affect the crude birth rate. The first is change in the population shares of different age groups. A rise in the share of people of reproductive age (roughly 15 to 45) increases the birth rate, as we saw earlier in the discussion of population momentum. Conversely, if the proportion of older people in the population goes up, as is happening in many industrial countries today, the birth rate drops. The second factor is fertility, the rate at which women have children.

Why, then, do people have as many children as they do? Is it because they are moved by Malthus's "passion between the sexes" and do not know how to prevent the resulting births? Or is it because they are tradition bound and custom ridden? Or is it rational in some economic and social settings to have large families? All three positions have some merit. The case for the first one, implicitly, is what those who recommend providing birth control as a response to rapid population growth probably have in mind. A Latin American doctor at an international conference put it bluntly. "People don't really want children," he said. "They want sex and don't know how to avoid the births that result." This viewpoint captures the element of spontaneity inevitably present in the reproductive process. Yet the evidence suggests that all societies consciously control human fertility. Rarely does the number of children that the average woman has over her childbearing years even approach her biological capacity to bear children. One exceptional case is that of the Hutterites, a communal sect that left Russia and settled in Canada and the northern Great Plains in the late 1800s. The group's religious beliefs and lifestyle encouraged maximum fertility. Hutterite women, on average, bore over 10 children, a fertility rate no contemporary low- or middle-income nation comes close to approaching. All societies practice some method of controlling their numbers whether by aborting pregnancies, disposing of unwanted infants, or inhibiting conception. (Recall our earlier discussion of fertility decline in nineteenth-century Finland.)

As for the second proposition, it has been said that many children are the social norm in traditional societies, that society looks askance at couples who have no or few children, that a man who lacks wealth at least can have children, and that a woman's principal socially recognized function in a traditional society is to bear and rear children. Such norms and attitudes are important, but they probably are not the

decisive factors in human fertility. Fertility is determined by a complex combination of forces. Social scientists in recent years have given increasing credence to elements of individual rationality in the process. Simply stated, they believe that most families in traditional societies have many children because it is in their interest to do so.

This brings us to the third explanation. Although some might regard it as a cold, inhumane way of looking at the matter, it is nevertheless true that children impose certain costs on their parents and confer certain benefits. In some low-income settings, especially in rural areas, children may supplement family earnings by working. On family farms and in other household enterprises, there usually is something that even a young child can do to increase production. In many poor societies, children work outside the home. In the longer run, children also provide a form of social security, which is important in societies that lack institutional programs to assist the elderly. In some cultures, it is considered especially important to have a son who survives to adulthood; if infant and child mortality is high, this can motivate couples to keep having children until two or three sons have been born, just to be safe. In addition to these economic benefits, which are probably more important in a low-income society than a more-affluent one, children also can yield psychic benefits (and costs!), as all parents know.

The economic costs of children can be divided into explicit and implicit costs. Children entail cash outlays for food, clothing, shelter, and often for education. Implicit costs arise when childcare by a member of the family, usually the mother, involves a loss of earning time. Some of the costs felt by parents parallel the costs of population growth experienced at the national level. For example, more children in a family may mean smaller inheritances of agricultural land, an example of a natural resource constraint operating at the family level. Similarly, it may be harder to send all the children in a larger family to school; this reflects the pressures on social investment felt when population growth is rapid.

Viewing childbearing as an economic decision has several important implications:

1. Fertility should be higher when children earn income or contribute to household enterprises than when they do not.
2. Reducing infant deaths should lower fertility because fewer births are needed to produce a given desired number of surviving children.
3. The introduction of an institutionalized social security system should lower fertility by reducing the need for parents to depend on their children for support in their old age.
4. Fertility should fall when there is an increase in opportunities for women to work in jobs that are relatively incompatible with childbearing, essentially work outside the home.
5. Fertility should be higher when income is higher because the explicit costs are more easily borne.

The first four predictions have received substantial support from empirical studies. The fifth, however, conflicts with observation. Fertility usually is negatively related to income. The negative relationship shows up both in time-series data (that is, fertility usually declines over time as income rises) and in cross-section data (fertility generally is higher in poor countries than in rich ones; also, in most societies, middle- and upper-income families have fewer children than poor families).

Several theorists have wrestled with the seeming anomaly that household demand for children falls as income rises; in other words, children are "inferior goods." Economist Gary Becker of the University of Chicago, a pioneer of *the new household economics*, analyzes children as a kind of consumer durable that yields benefits over time.[8] Couples maximize a joint (expected) utility function in which the "goods" they can "buy" are (1) number of living children, (2) child quality (a vector of characteristics including education and health), and (3) conventional goods and services. The constraints faced by parents in Becker's model are (1) their time and (2) the cost of purchased goods and services.

Becker explains the fall in fertility as income rises over time by saying that the cost of children tends to rise, especially because the opportunity cost of the parents' time, particularly the mother's, goes up. The familiar income and substitution effects that result from any price change are at work here. As a woman's market wage rises, her income also rises, leading to an increase in her demand for any normal good, presumably including the demand for more children. But as a woman's wage rises, the opportunity cost of her time also goes up, so the price of raising children rises. As a result, her demand for any activity that is intensive in the use of her time decreases, including child rearing. If the substitution effect is greater than the income effect, couples may decide to have fewer offspring. Becker also argues that given the rising cost of child *quantity,* many parents opt to invest in child *quality* and spend more time and money on a smaller number of children. Wanting "higher-quality" children as income rises also reverts the demand for children back to the demand for a "normal good."

Much of Becker's work on the family was directed at understanding declining fertility levels in the United States and other high-income economies. But the application of the new household economics to the circumstances of developing nations also offers keen insights. Replacing Malthus's assumption that people are driven solely by their passions, fertility now is seen as the outcome of a rational process. This does not mean that every child born, in rich and poor nations alike, is the result of a conscious cost–benefit calculation. Economics rarely can explain individual outcomes; its power is in explaining average tendencies. What the new household economics suggests is that rich and poor people alike weigh the consequences of their actions, and to understand behavior, we may gain more insights by assuming people act in what they perceive as their own best interests. Having large numbers of children then may be seen not as an irrational act but as the result of the difficult choices poor people face. For poor families, having many children may make economic sense.

[8]Gary Becker, *A Treatise on the Family* (Cambridge, MA: Harvard University Press, 1981).

The new household economics provides insights into the fertility decisions of families in the developing world and, in so doing, raises an apparent contradiction. If individual decisions about fertility are rational and in the perceived best interest of the family, why did McNamara, and many others, identify population growth as "the greatest single obstacle" to economic development? How could rational decisions at the household level result in so dire a society wide outcome?

POPULATION GROWTH AND ECONOMIC DEVELOPMENT

Even if population growth is not the greatest obstacle, has population growth been a major impediment to achieving economic development and is it still one today? To a casual observer who has spent time in Cairo or Calcutta, Manila or Mexico City, or in hundreds of other cities or villages in the developing world, the answer might seem obvious. Looking at the masses of poor people in these places, it would be hard not to conclude that the situation could be measurably improved if only there were fewer people. (A casual observer might draw the same conclusion in New York or Tokyo.)

But more than casual observation is needed to understand the complex relationship between population growth and economic development, especially if one is going to consider policy interventions to reduce population growth. First, we must be clear on the question we are asking. We cannot ask whether people who have been born would have been better off having never been born. This is a question economics cannot possibly answer. Our question can only be whether per capita income, life expectancy, educational attainment, or any other indicator of economic development would be higher if a nation's population had grown more slowly. Given the centrality of economic growth to the development process, much of our attention is on the relationship between population growth and economic growth.

Simon Kuznets, a Nobel Prize winner in economics and a pioneer in the use of data to examine trends in economic growth, reported on a lack of correlation between population growth and per capita output growth for 40 less-developed nations between the early 1950s and 1964.[9] Similar studies have been conducted since Kuznets's work, many of which also report a low correlation between the two growth rates over various time periods.[10] Starting in the 1980s, it was more common

[9]For the 40 less-developed countries in Kuznets's sample, the rank correlation between rates of population growth and growth in per capita GDP was 0.111. From Simon Kuznets, "Population and Economic Growth," *Proceedings of the American Philosophical Society* 3 (June 1967), as reported in United Nations, *The Determinants and Consequences of Population Trends,* Department of Social Affairs, Population Division, Population Studies No. 17 (New York: United Nations, 1973).

[10]Discussion of these studies can be found in Allen Kelley and Robert Schmidt, "Economic and Demographic Change: A Synthesis of Models, Findings and Perspectives," in Nancy Birdsall, Allen Kelley, and Steven Sinding, eds., *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World* (Oxford: Oxford University Press, 2001).

to find a negative correlation between them. In growth regressions, which include many possible determinants of economic growth, the significance of population growth varies depending on what other variables are included.

The lack of one pattern relating population growth to economic growth should not be misinterpreted. It does not mean that population has no systematic effect on economic growth; it simply means that the relationship is complex. We need to ask, How does population growth affect the determinants of economic growth? What direction of causality is there between population growth and economic growth? And how important are the individual components of population growth—fertility, mortality, and migration—to gross domestic product (GDP) growth (and vice-versa)?

We begin by challenging a *naïve view* of how population growth and economic growth are related. The growth rate in GDP per capita equals the growth rate of GDP minus the growth rate in population.[11] The naïve view then holds that, for any given rate of GDP growth, per capita growth would be faster the slower the growth in population. This conclusion is true, but its implicit assumption is false. Growth in output is not independent of growth in the population. The simplest reason is that children born today are the labor force of the future, and labor, along with land and capital, is one of the main factors of production for any economy. Stated differently, population growth affects both the level of GDP and how many people must share in that GDP. Once one recognizes these connections, the critical question becomes how population growth affects the core determinants of economic growth: the accumulation of factors of production and the productivity of those factors.

POPULATION AND ACCUMULATION

A pioneering model of population's effect on material welfare was written in 1958 by demographer Ansley Coale and economist Edgar Hoover.[12] Their work falls squarely into the camp of **population pessimists**, or anti-natalists, those who perceive population growth as harmful to economic development. Coale and Hoover argue that a reduction in the birth rate could raise per capita income in three important ways. First, lower fertility levels would slow future labor force growth. The amount of investment then needed to provide a constant amount of capital per worker for a growing number of workers (*capital widening*) would go down and permit more investment to be used to increase capital per worker (*capital deepening*).[13] Second, with lower

[11]The growth rate of any fraction, A/B, is equal to the growth of the numerator, A, minus the growth of the denominator, B. This is because the growth rate of any variable, X, can be found by taking the derivative of the log of the variable with respect to time: $d(\ln X)/dt = (dX/dt)/X = g(X)$, where g is the rate of growth. If $X = (A/B)$, then $\ln(A/B) = \ln A - \ln B$, and $d[\ln(A/B)]/dt = d(\ln A)/dt - d(\ln B)/dt$, or $g(A/B) = g(A) - g(B)$. In the case of GDP per capita, $g(\text{GDP/population}) = g(\text{GDP}) - g(\text{population})$.

[12]Ainsley Coale and Edgar Hoover, *Population Growth and Economic Development in Low-Income Countries: A Case Study of India's Prospects* (Princeton, NJ: Princeton University Press, 1958).

[13]The implications of capital widening and capital deepening for economic growth are developed more fully in the discussion of the Solow growth model in Chapter 4.

fertility and fewer children, public funds could be diverted away from education and health expenditures and toward physical capital, which Coale and Hoover assumed would be a more productive use of government spending. Third, slower population growth would lower the **dependency ratio**, the ratio of the non-working-age population (usually 0 to 14 years and 65 and over) to the working age population (Box 7–3). A lower dependency ratio, in turn, would reduce consumption and increase saving at any given level of income, permitting a higher rate of asset accumulation. Taken together, these three benefits from slower population growth could raise income levels at both the household and aggregate levels.

Coale and Hoover's conclusions, generally, have not always been supported by later research.[14] For a given amount of investment, a larger labor force results in less capital per worker, but the quantitative significance of this effect on output often appears small. Coale and Hoover's concern about the diversion of resources toward education also has not held up to later research. Demographic factors do not exert much of an independent effect on the share of GDP allocated to education and other social welfare programs. In addition, education and health increasingly are seen, not as consumption expenditures, but as investments in human capital that may have returns equal to or higher than the returns on investments in physical capital.

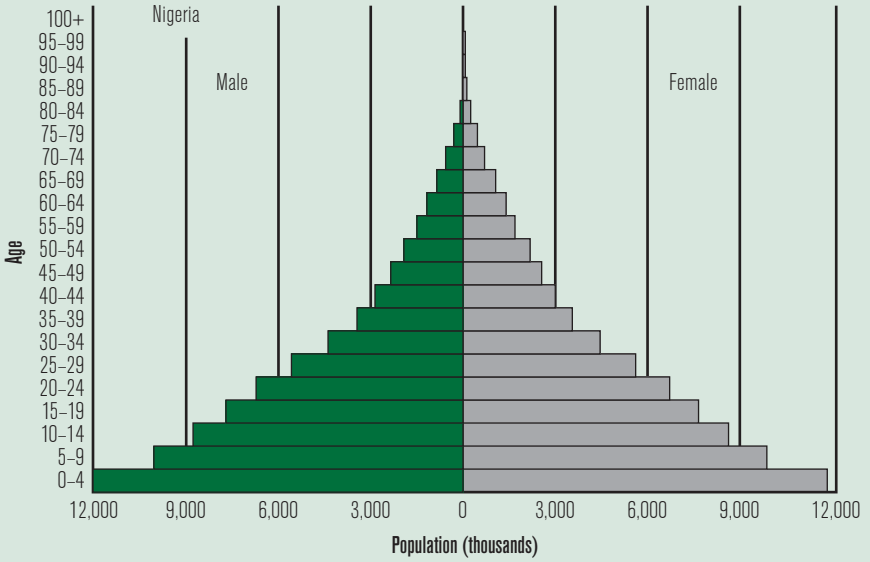
The third relationship highlighted by Coale and Hoover, among population growth, dependency ratios, and saving, has attracted a great deal of analysis. Some studies find little correlation. This may be because business and government savings are relatively independent of demographic change. Even household saving may be fairly insensitive to changes in population growth if most saving comes from wealthy families. Such households' saving behavior would be affected by neither their own fertility decisions nor those of the poor. On the other hand, there is cross-country evidence that, as the share rises of the population 15 years of age and younger, the aggregate saving rate falls.

This last insight is part of a more recent analysis often referred to as the **demographic dividend** or **bonus**. The dividend refers to the opportunity created when a nation with rapid population growth experiences a fall in fertility. As fertility falls, population growth falls too, but for a period of time the working age population (15- to 65-year-olds) grows *more rapidly* than the youth population (0 to 14). This is because the growth of the working age population depends on earlier and higher birth rates. The age pyramids in Box 7–3 experience a bulge in the number of young adults and a shrinking of the number of children. Not only are there now more potential workers per child, but more women join the labor force as fewer years are spent on childbearing. With relatively more workers, the productive potential of the economy expands. During this phase of the demographic transition, the society still has a relatively small older age cohort so the problem of the dependency of the elderly is not yet a concern. It will eventually become one, as it has in an increasing number of developed and developing nations today, making the demographic dividend a temporary opportunity.

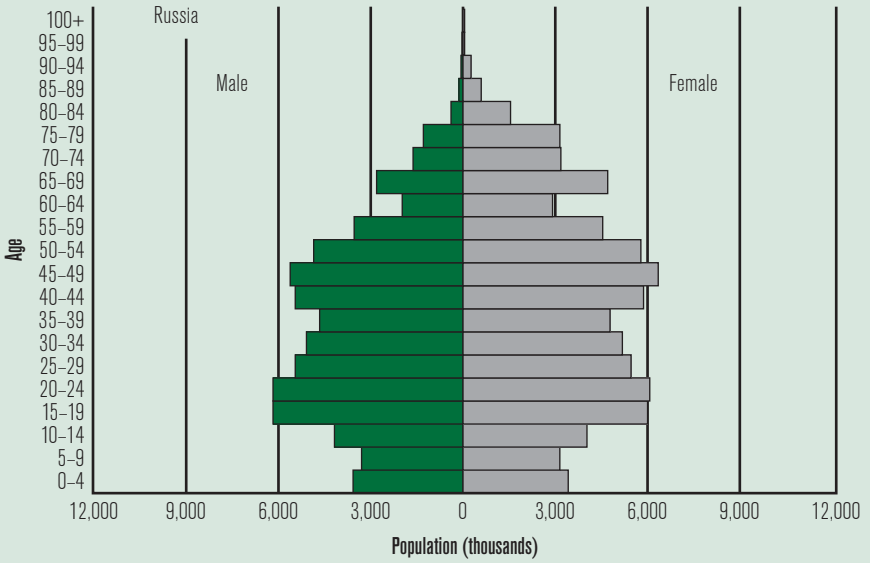
[14]Nancy Birdsall, "Economic Analysis of Rapid Population Growth," *World Bank Research Observer* 4, no. 1 (January 1989).

BOX 7-3 POPULATION GROWTH, AGE STRUCTURE, AND DEPENDENCY RATIOS

In 2009, Nigeria had a population of 153 million people compared to Russia with 142 million. Although similar in population size, the two nations are at very different points in their demographic transitions. Nigeria has a total fertility rate



(a)



(b)

of 5.7; Russia's is 1.5. Nigeria's population growth rate this past decade was +2.1 percent per year; Russia's was −0.35 percent. Only recently has Nigeria reached the stage of the demographic transition where birth rates have started to fall. Russia passed through this phase decades ago.

Differences in crude birth and death rates, in addition to their impact on population growth, have a large effect on a nation's age structure. Population pyramids (on previous page) illustrate these effects. Each bar of the pyramid refers to the number of males and females, respectively, in a given age cohort. Nigeria has the distinctive population pyramid of a nation in which population growth is rapid. Each age cohort is larger than the one born before it, the result of both high fertility and population momentum. In contrast, Russia's population pyramid is more rectangular, the result of low population growth. The base of Russia's pyramid is narrowing because fertility has fallen below replacement levels. (Differences among population cohorts also reflect historical events. Russia's 60- to 64-year-old cohort is especially small because this group was born between 1944 and 1948, during and just after World War II. Low fertility and high infant mortality during those years explains the shortfall in this age cohort today.)

Nigeria's youth dependency ratio, the number of children (ages 0 to 14 years) divided by the number of people of working ages (15 to 64 years), obviously will be higher than Russia's. The ratio equals 0.87 for Nigeria and only 0.21 for Russia. But Russia has a higher elderly (older than 64) dependency ratio of 0.14 than Nigeria's (0.03). Differences in age structure have important implications for every aspect of a society, from what it consumes to how it votes to the culture it creates. What implications differences in dependency ratios have for overall economic growth and development is considered in the text.

Source: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, *World Population Prospects: The 2008 Revision*, available at www.un.org/esa/population/publications/wpp2008/wpp2008_highlights.pdf.

The decline in fertility not only increases the size of the working age population relative to the youth population. It may also increase savings and capital accumulation as more of the population is now in the working and savings part of the life cycle. But the demographic dividend is not a sure thing; it is only an opportunity. If other economic policies are not in place, the potential will not be realized. Comparisons of Latin America with East Asia bear this out. Both regions experienced similar demographic transitions, with falling birth rates and increasing

working age cohorts. But the similarity ends there. East Asia boomed; Latin America did not.

Some researchers argue that East Asia's demographic transition contributed substantially to the region's rapid economic growth. Using cross-country growth regressions, they attribute one-third or more of the region's growth to the rapid increase in its working age population after the decline in fertility. Others think the effect on growth was far smaller, finding little empirical evidence at the household level to support the life-cycle hypothesis of saving central to the demographic dividend.[15] The relationship between population growth and saving remains unresolved.

POPULATION AND PRODUCTIVITY

In addition to its potential effects on the accumulation process, population growth also can influence the other important determinant of economic growth, the productivity of assets. Concern over these impacts dates back to Thomas Malthus. Malthus saw population growth leading to an inevitable decline in agricultural productivity and having devastating consequences for humanity. What Malthus did not envisage is the revolution in agricultural technology that would raise, not lower, the productivity of land. Malthus never would have believed that, in a world of 7 billion people, individuals, on average, would eat better than they did in his day, when world population was closer to 1 billion. Economic historian Robert Fogel cites evidence that, in the late eighteenth century, about the time Malthus wrote his famous essay, the average daily caloric consumption in France was 1,753 and in England in the neighborhood of 2,100. Contrary to Malthus's predictions, in the developing world today the average is much higher, exceeding 2,600 calories.[16]

[15]The significance of the demographic dividend is presented in David Bloom and Jeffrey Williamson, "Demographic Transitions and Economic Miracles in Emerging Asia," *World Bank Economic Review* 12, no. 3 (1998); David Bloom, David Canning, and Jaypee Sevilla, *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change* (Santa Monica: Rand, 2003). A dissenting view is presented in Angus Deaton and Christina Paxson, "Growth, Demographic Structure and National Saving in Taiwan," *Population and Development Review* 26 (2000; Suppl.), 141–73.

[16]There is a group of countries that in recent years have consumed fewer than 1,750 calories per day, the level in France in the late eighteenth century. In 1999–2001, this group included Afghanistan, Burundi, the Congo, Eritrea, Somalia, and Tajikistan. Although all but Tajikistan had extraordinarily high TFRs of 6.0 or more at the time, what also distinguishes this group is that they were all failed states. It is likely that their malnutrition was caused less by overpopulation than by the disruption of agriculture due to massive political instability. Historical data on caloric intake are from Robert Fogel, "The Relevance of Malthus for the Study of Mortality Today: Long-Run Influences on Health, Mortality, Labor Force Participation, and Population Growth" (231–84), in Kerstin Lindahl-Kiessling and Hans Landberg, eds., *Population, Economic Development, and the Environment* (Oxford: Oxford University Press, 1994). More recent data are from United Nations Food and Agriculture Organization, *The State of Food Insecurity in the World* (Rome: FAO, 2003).

Despite these improvements, neo-Malthusian ideas still are popular. Concern over an imbalance between population and food supplies continues to be voiced, as is concern over how population pressures will affect the availability of fresh water and energy, the spread of infectious diseases, the degree of biodiversity and climate change, and the overall sustainability of the environment. Neo-Malthusians, like their intellectual predecessor, believe that there are limits to the carrying capacity of the planet and that science and technology cannot resolve fundamental problems of diminishing returns.[17]

Malthus's thinking placed him squarely in the camp of population pessimists. But some later economists, **population optimists**, would view population growth as having the potential to increase factor productivity. Several reasons for such a relationship have been proposed. First, a larger population, the result of more rapid population growth, can yield economies of scale in production and consumption. Roads are a good example. The return on an investment in a rural road is higher, up to a point, the more people there are to use the road. Other forms of infrastructure as well as public services in health and education exhibit similar scale effects. Second, there is some evidence that population pressures can induce technological change. Increasing population density in previously underpopulated regions can encourage more labor-intensive farming systems, increasing the return to land and, perhaps, to other inputs. Third, economist Julian Simon, perhaps the greatest population optimist of all, argued that a larger population contains more entrepreneurs and other creators, who can make major contributions to solving the problems of humanity. Simon called human ingenuity the "ultimate resource" that can overcome any depletion of other resources.[18]

Just as is true for the population pessimists, empirical support for the positions of the population optimists is spotty. Scale effects exist but often are realized at population densities that already have been achieved; a possible exception is some sparsely settled regions of Africa. Diseconomies of scale also exist and city sizes in parts of the world may be approaching those levels. Finally, there is no simple one-for-one correspondence between population pressures and technological change. Population is only one of many factors that affect the nature and quality of the institutional environment that conditions the introduction of new technologies and methods of pro-

[17]The neo-Malthusian perspective is presented in Lester Brown, Gary Gardner, and Brian Halweil, *Beyond Malthus: Nineteen Dimensions of the Population Challenge* (New York: Norton, 1999). The opposing view, that the planet is not overreaching its carrying capacity, is presented in Bjorn Lomborg, *The Skeptical Environmentalist: Measuring the Real State of the World* (Cambridge: Cambridge University Press, 2001).

[18]The seminal work on population pressures inducing changes in agricultural production is Ester Boserup, *The Conditions of Agricultural Growth* (Chicago: Aldine, 1965.) On human ingenuity, see Julian Simon, *The Ultimate Resource* (Princeton, NJ: Princeton University Press, 1981).

duction. Green revolution technologies, which can dramatically increase agricultural output for specific crops, such as rice and wheat, have been adopted in some densely populated areas but not in others. This suggests that population pressures alone have not been the deciding factor.

POPULATION AND MARKET FAILURES

In addition to the population pessimists and the population optimists who present such conflicting views, there is a third school of thought on population and development: the **population revisionists** (referred to by some as population neutralists). The revisionists situate themselves between the two extreme camps, arguing that there is no one size that fits all on population matters. Because of the varied influences of population growth on economic and social variables, growing populations may or may not be detrimental to economic development, depending on time, place, and circumstances.

The revisionists also bring together micro models of fertility behavior with macro assessments of the consequences of population growth. This merging of micro and macro is critical. According to the prevailing microeconomic model of fertility, individuals have control over their fertility decisions and behave rationally. In other words, couples, on average, make decisions on the number of children they want in their own best interest. If this insight is correct, how is it possible that actually tens of millions of rational individual decisions concerning the number of children to have could result in detrimental outcomes for nations or even for the planet as whole? The revisionist answer is a familiar one to economists: **market failures**, situations in which the costs or benefits of reproductive behavior by individuals (households) are not fully borne by them.

Revisionists agree with some neo-Malthusians that rapid population growth may hasten depletion of natural resources or harm the environment. But unlike the neo-Malthusians, revisionists argue the fundamental problem is not too many people, but the lack of well-defined property rights. As in the classic "Tragedy of the Commons," population growth can destroy a common resource—for example, common grazing lands or a fishery—because no one family takes into consideration the impact of its use on others. A larger family might help that family gain greater benefits from the common resource (at least in the short run) but, at the same time, help speed its destruction, to the detriment of all.[19] A similar argument can be made about many government services, whether in education, health, sanitation, or transport. Each family may be acting rationally, but if the population grows too quickly, government services may not expand quickly enough. The resulting congestion of government services may produce lower quality of life for all. The root cause of the problem is not

[19]Garret Hardin, "The Tragedy of the Commons," *Science* 162, no. 1 (1968), 243–48.

population growth but the inability of government to finance the increased demand for publicly provided goods and services. In both these examples, population growth may exacerbate an existing market failure.

Revisionists also call attention to a failure in the market for contraception. If there are poor, incomplete, or imperfect markets either for information on contraception or for the contraceptives themselves, women have more children than they want and population growth is higher. One study reports that in Haiti 40 percent of women 15 to 49 years old preferred to avoid a pregnancy but were not using contraception.[20] To the extent that this was the result of a lack of information about or access to birth control, there is a market failure for contraceptives. In such circumstances, fertility levels do not fully reflect individual preferences and are too high. The ultimate source of the market failure is less apparent. Is it due to lack of information, government restrictions on the sale of birth control devices, or something else? The price of birth control is far less than the price of raising a child. Free markets deliver Coca-Cola worldwide, why not contraceptives?[21]

In addition to drawing attention to market failures, population revisionists focus on the impact of fertility decisions on dimensions of human welfare other than growth in per capita income, such as income distribution. At a household level, high fertility may benefit parents, in terms of earnings from child labor and old age support. But these same advantages for parents may work to the disadvantage of children if they result in fewer resources available per child for human capital investments. Higher-birth-order children may be at a particular disadvantage. Looking across households, because higher fertility is inversely correlated with household income level, more rapid population growth may increase income inequality and worsen poverty outcomes.

Population revisionists have contributed to the otherwise polarized debate on population between the pessimists and optimists. The revisionists emphasize that rapid population growth is unlikely to be the primary impediment to economic development, focusing instead on how population growth can exacerbate the failings in other markets or particular government policies. Even these negative effects are likely to be mitigated over the long run as households adjust to changing circumstances. Despite their more nuanced approach, the population revisionists seem unable to identify the specific circumstances and country settings in which market failures are large and policies to limit births are warranted. Their perspectives, however, help frame the debate over appropriate population policies.

[20]Sara Maki, "Unmet Need for Family Planning Persists in Developing Countries" Population Reference Bureau (October 2007), www.prb.org/Articles/2007/UnmetNeed.aspx accessed March 2012.

[21]Adapted from William Easterly, *The Elusive Quest for Growth* (Cambridge: MIT Press, 2001), p. 89–90.

POPULATION POLICY

While economic analysis cannot provide a definitive answer as to where, or even whether, intervention is warranted to slow population growth, most governments in developing nations and most donors favor slower population growth. In pursuit of this objective, governments implement a variety of population policies. Economic reasoning can help evaluate the relative merits of specific policies when the stated goal is to lower population growth rates.

The growth of a population can be reduced either by lowering the birth rate or raising the death rate. In practice, the only acceptable policy solution is reducing the birth rate. No one would advocate increasing the death rate as a way to slow population growth, but nations and donors do not always follow a strictly anti-mortality approach. The initially slow donor response to the HIV/AIDS crisis in sub-Saharan Africa is a case in point.

If slowing the rate of population growth is the goal, then reducing the birth rate is the solution. This can be achieved by reducing the share of the population of reproductive age or the fertility rate. Raising the average age of women at childbearing slows down population momentum and, in time, leads to a reduction in the share of the population of reproductive age. Providing women with more education and better employment opportunities can achieve these results. These interventions will also reduce fertility levels. And reduction in fertility, whether by direct or indirect means, is the primary target of population policies.

Two controversies surround population policies aimed at reducing fertility. One is the debate over the significance of family planning efforts versus broad-based socioeconomic development. The other is the debate over the use of relatively authoritarian tactics, including elements of China's 1979 one child campaign, to achieve population goals. Before getting to either of these controversies, we can identify some noncontroversial elements of population policy.

Educating girls is one of the least controversial and most effective means of achieving a long-term decline in fertility. As predicted by the new household economics, women with more education have a higher opportunity cost of time and choose to have fewer children. This result has been born out empirically by numerous studies. (Figure 7–5 provides evidence from four nations.) Because educating girls is valuable in its own right, all schools of thought on population agree to support improvements in girls' education. Any subsequent impact on fertility is as much a consequence as a motivation for such policies. The same can be said for interventions that reduce infant and child mortality. Such interventions, usually involving prenatal care, better birthing practices, and health initiatives directed at the young, can significantly improve survival chances. These interventions need not be motivated by a desire to lower fertility, even though they have this effect.

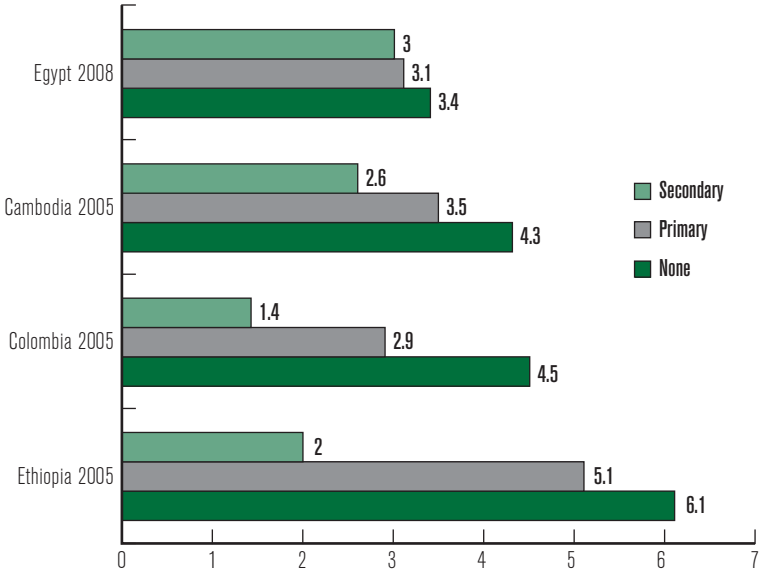


FIGURE 7–5 Total Fertility Rates by Mother's Education

Source: Demographic and Health Surveys, available at www.measuredhs.com.

All policies that promote economic development and lead to more education, better health, higher per capita incomes, and less poverty are associated with declining fertility, even if this is not their primary purpose. At the first United Nations World Population Conference, held in Bucharest in 1974, a popular slogan was, "Take care of the people and the population will take care of itself." Some development specialists go so far as to conclude that economic development is the best contraceptive. But others disagree and argue that more direct interventions to control fertility outcomes are needed.

FAMILY PLANNING

Family planning refers to a range of reproductive health information and services. It includes information on contraception and provision of contraceptives of various types, often highly subsidized or delivered at no cost to the user. Without government family planning services, couples could purchase contraceptives available at market prices or could control their fertility using traditional methods (abstinence, rhythm method, withdrawal). More modern forms of contraception (condoms; hormone therapies including birth control pills, patches, and injections; IUDs; male and female sterilization), however, are more reliable and associated with a greater probability of achieving desired fertility. One of the most important contributions of family planning services, whether provided publicly or through the market, is to help families reach their desired fertility levels. As levels of desired fertility fall, modern forms of contraception provide the means for achieving these goals.

Some population specialists also draw attention to the significant proportion of births in the developing world that are unwanted. Information on unwanted births is obtained from household survey data in which women report the number of births they have had as well as the number of children they would like to have. Using data from the 1970s and 1980s, demographer John Bongaarts estimated that approximately one out of every four children born in the developing world (excluding China) was unwanted.[22] Estimates for the past decade suggest much smaller numbers; for a sample of developing nations representing about two-thirds of the developing world excluding China, the total "not wanted" was only 8 percent, with another 9 percent "wanted later."[23] Reducing the number of unwanted births is another way of reducing population growth.

Whether family planning programs are effective at reducing unwanted births is a matter of some debate. This may sound surprising: Isn't the lack of contraception the reason women have births they do not want? Some further explanation is needed. When women reduce their desired level of fertility, contraception plays an important role in achieving these lower levels. Desired levels of fertility and contraceptive prevalence are well correlated, with causality most likely running from the former to the latter. When women desire small families, contraceptive prevalence is high. However, contraceptive use is less well correlated with reported differences between actual and desired fertility—that is, with the number of unwanted births. Using data from the 1970s and 1980s, development economist Lant Pritchett found that after controlling for desired fertility there was little correlation between contraceptive prevalence and excess births.[24] Women may have more births than they want but the availability of contraception may not explain why.

Family planning programs may influence individual preferences about how many children to have. This dimension of family planning efforts involves providing information, counseling, and various forms of "persuasion" concerning the number and spacing of children. Public campaigns to discourage high fertility are employed by many countries. In the 1980s, the slogan of India's national family planning program was *"Hum Do, Hamaray Do,"* which when translated means "Us Two, Our Two," or for two parents, two kids—a model family. The slogan was displayed widely on the sides of buses, billboards, and in TV advertisements and came with a logo, a silhouette of two parents holding the hands of two children, a boy in shorts and a pigtailed girl in skirts. The slogan on India's Ministry of Health and Family Welfare website recently flashed, "Control Population! Have fun with one!" In Indonesia, the *"Dua cukup"* program recommended "two is enough"; in the Philippines, public ads showed two happy children and advised, "If you love them, plan for them;" and in Zimbabwe, a family planning poster suggested,

[22]Bongaarts, "Population Policy Options," p. 771–76.

[23]Population Reference Bureau, *Family Planning Worldwide: 2008 Data Sheet,* www.prb.org/Publications/Datasheets/2008/familyplanningworldwide.aspx.

[24]Lant Pritchett, "Desired Fertility and the Impact of Population Policies," *Population and Development Review* 20, no. 1 (March 1994), 1–55.

"A small family brings a better life." These campaigns can affect social norms and the desire to have many children.

Family planning information also can be more personalized. One widely studied program of this type began in the late 1970s in the Matlab region of Bangladesh. It included fortnightly visits to each married woman in half the villages in the region. The visit was from a project employee, who discussed family planning needs and provided contraceptives. As a result of these visits, fertility rates fell dramatically by 1.6 births relative to the half of villages in the region that did not receive such visits. One evaluation of the program in Matlab considered its long-term consequences. After roughly 20 years, families who lived in the program villages not only had fewer children with greater spacing between them but were healthier and more prosperous than residents of similar villages that had not been served.[25] An important lesson of the Matlab experiment is that family planning can be effective at reducing fertility and providing other benefits, but it can entail substantial costs. Fortnightly field visits, free contraceptives, and administrative expenses per woman amounted to 10 percent of per capita income, sums that could not be sustained at a national level.[26]

Family planning programs need not be as expensive as in the Matlab program to be effective, but the general point remains. If the goal is to reduce fertility levels, is it better to devote scarce resources to expand family planning programs, or might it be more effective to devote these same resources to general improvements in socioeconomic status, which, in turn, could result in an equal or greater decline in fertility? In their evaluations of Indonesia's rapid decline in fertility rates, economists Paul Gertler and John Molyneaux determined that heavy subsidization of contraceptives, representing about half of the family planning program's expenditures, had a marginal impact on reducing fertility. Expansion of the distribution network into rural areas had a more significant effect. But most of the increased use of contraceptives was induced through economic development and the improved education and opportunities facing Indonesian women.[27]

Family planning may not play a large independent role in reducing fertility but this does not suggest that governments should abandon these programs or limit the availability of contraceptives. Modern forms of contraception enable families to better control their fertility. In this sense, family planning and development are complements, not substitutes. Family planning services also help people lead healthier and more satisfying lives. By being better able to time pregnancies, women can better space their children. Children born three to five years apart have higher rates of survival than children born within two years of one another. Avoiding unintended pregnancies can also reduce the need for abortions. In many low- and middle-income

[25]James Gribble and Maj-Lis Voss, "Family Planning and Economic Well-Being: New Evidence from Bangladesh," *Policy Brief* (Population Reference Bureau, May 2008).

[26]Pritchett, "Desired Fertility and the Impact of Population Policies," p. 37.

[27]Paul Gertler and John Molyneaux, "How Economic Development and Family Planning Programs Combined to Reduce Indonesian Fertility," *Demography* 31, no. 1 (February 1994), 33–63; John Molyneaux and Paul Gertler, "The Impact of Targeted Family Planning Programs in Indonesia," *Population and Development Review* 26 (2000; Suppl.), 61–85.

nations, abortions are illegal and take place under unsafe circumstances, contributing to maternal mortality. Increased use of contraceptives can help reduce the amount of premature deaths of both women and children in poor nations.

The 1994 Cairo International Conference on Population and Development endorsed the idea that family planning be replaced by a broader program of reproductive health. The World Health Organization defines the goals of this new approach as follows:

> Reproductive health implies that people are able to have a responsible, satisfying, and safe sex life and that they have the capability to reproduce and the freedom to decide if, when, and how often they do so. Implicit in this last condition are the right of men and women to have access to safe, affordable, and acceptable methods of fertility regulation of their choice, and the right of access to appropriate health care services that will enable women to go safely through pregnancy and childbirth and provide couples with the best chance of having a healthy infant.[28]

In the same spirit, the United Nations in 2007 added achieving universal access to reproductive health as a target to the MDG to Improve Maternal Health.[29]

AUTHORITARIAN APPROACHES

Some countries, concerned over the growth of their populations, have resorted to relatively authoritarian approaches to reduce their birth rates. In India, during the 1970s, the government added male sterilization to its list of promoted family planning methods. It did so because the government felt unable to control population growth by conventional means. Incentives were offered to men who agreed to be sterilized, and quotas were assigned to officials charged with carrying out the program. Problems arose when force allegedly was used against low-status individuals by officials anxious to fill their quotas. The sterilization campaign provoked an adverse political reaction because the methods of population control were regarded as excessively zealous and callous in their disregard of individual rights. Indira Gandhi's surprising defeat in India's 1977 general election was attributed in part to the population policy of her emergency government.

China's experience has been different. Chinese leaders have expressed concern about the size of the nation's population since the census of 1953 revealed a population of almost 600 million. Modest attempts to influence fertility outcomes followed, but not until 1971 was a more aggressive stance adopted. In that year the "planned births" campaign was initiated and established three reproductive norms, *wan, xi, shao* ("later, longer, fewer"): later marriage, longer spacing between births,

[28]Cited in World Bank, *Population and Development* (Washington, DC: World Bank, 1994), p. 81.

[29]Specific goals and strategies for improving reproductive health are presented in, World Bank, "The World Bank's Reproductive Health Action Plan, 2010–2015" (April 2010), http://siteresources.worldbank.org/INTPRH/Resources/376374-1261312056980/RHActionPlanFinalMay112010.pdf.

and fewer children. To implement these norms, birth targets were set for administrative units at all levels throughout China. The responsibility for achieving these targets was placed in the hands of officials heading units ranging from provinces of 2 to 90 million people down to production brigades of 250 to 800. The national government held information and motivation campaigns to persuade people to have fewer children, but it was left to local officials to fill out many details of the program and finance much of its cost. In China's villages, where much of the population lived, IUD insertions, abortions, and, later, sterilizations were offered. A wider range of contraceptives was provided in the cities. National spokespeople maintained that participation in the program was voluntary, but local officials with targets to fulfill often applied pressure. At the production brigade level, birth planning became intensely personal, as couples were required to seek approval to have a child in a particular year. (The application might be accepted or they might be asked to wait a year or two.)

The *wan xi shao* campaign began by discouraging couples from having more than two children; by the end of the 1970s, couples were encouraged to have only one. During the 1970s, total fertility rates in China plummeted from around six children per woman to less than three, a rate of fertility decline without precedent in world history. But, despite lowered fertility, population projections continued to cause concern, and in 1979, the **one child campaign** was promulgated. Couples, initially, were required to apply for official approval before conceiving the one allowed child. Those who complied could receive income supplements, extra maternity leave, and preferential treatment when applying for public housing. Couples bearing more than one child might be fined or lose access to education or other privileges. Early in the campaign, the government employed harsh methods to enforce its population goals. A system of incentives and penalties, plus other forms of social pressure, moved women to undergo sterilization after two births. China's one child approach has gone through various phases since its inception, with periods of stricter and looser interpretations. Since the late 1980s, rural couples in most provinces were allowed to have two children if their first was a daughter. Chinese couples still are obliged strictly to limit the number of children they have but elements of the earlier coercion to force such outcomes have waned.[30]

Several lessons can be learned from China's experience. First, the dramatic decline in fertility that took place in China is the result of more than the government's population policies. At the same time that these programs were implemented, economic reforms ushered in a period of rapid economic growth and urbanization, which decreased desired fertility. Even without the one child campaign, Chinese fer-

[30]For a survey of demographic trends and policies in the People's Republic of China, see Nancy Riley, "China's Population: New Trends and Challenges," Population Reference Bureau *Population Bulletin* 59, no. 2 (June 2004). The human face of China's policy is captured in the multimedia presentation by American Public Media, *Marketplace*, "China's One Child Policy" (June 2010), available at http://marketplace.publicradio.org/projects/project_display.php?proj_identifier=2010/06/21/china-one-child-policy; accessed February 2012.

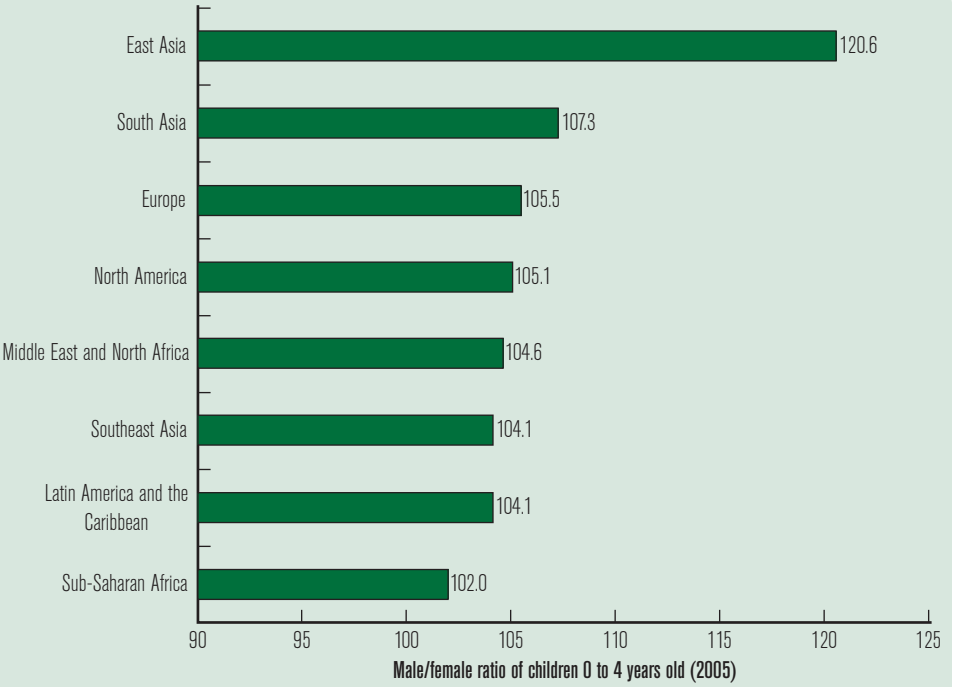
tility levels would have fallen but probably not as far. Today, many Chinese couples would likely choose on their own to have only one child or even none. Second, the sharp fall in births may have longer-term adverse consequences. China will face a rapidly aging population in the decades ahead, which may stress both private and public systems of old-age support. China also faces a growing imbalance between the number of males and females (Box 7–4). This is the result of a traditional preference

BOX 7–4 MISSING GIRLS, MISSING WOMEN

Most parents believe that the chances of having a daughter or a son are fifty–fifty. But this is not true. For most human populations, the expected ratio of boys to girls at birth is closer to 105:100. Scientists are not entirely sure why this is the case, but believe it compensates for biological weaknesses in males resulting in higher mortality rates for boy versus girl infants. Because more than biology determines survival, social scientists draw attention to sex ratios, at birth and throughout the life cycle, when they deviate from what is expected in a gender-neutral environment.

The chart on the next page compares the sex ratio of children 0 to 4 years old in various regions. Europe and North America have ratios close to 105:100. The lower ratios in some regions may be due to a tendency for populations of African descent to record male to female ratios at birth closer to 102:100 or to the higher rates of male infant and child mortality. The most striking result is the pattern in East Asia, where in 2005 there were close to 121 boys for every 100 girls. This is neither a recent outcome nor is it biologically driven. The shortage of girls 0 to 4 years old is the result of long-standing son preference in the region. In previous periods, the shortage of young females was the result of deliberate female infanticide and neglect. Today, sex-selective abortion is another mechanism causing male children to outnumber female children, one facilitated by the increasing availability of ultrasound and other prenatal screening devices.

Sex-selective abortions, abandonment of female infants, and differential neglect of girls, especially in medical care, are all means of realizing son preference. They have been common in many parts of East Asia but none more so than in China. Although son preference has a long history in China, the one-child campaign and the overall move to low fertility rates among Chinese households have contributed to these outcomes. In 1982, at the beginning of the one-child policy, the sex ratio of first births was 107; second births, 105; and all subsequent births, 110. The change by 2000 was dramatic. The sex ratio of first births in China was still 107, but for second births it rose to 152 and for any subsequent



Source: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, *World Population Prospects: The 2008 Revision*, available at www.un.org/esa/population/publications/wpp2008/wpp2008_highlights.pdf.

births it was almost 160. Second and higher-order births represented only one-third of all births in 2000, but sex selection for these births resulted in the high and grossly imbalanced sex ratio for all Chinese births.

China is not alone in this extreme form of discrimination against females, which has been referred to as *gendercide*. Amartya Sen examined the issue of differential mortality of males versus females and found that survival chances vary across countries and over the life cycle. In India, the differential used to show up less at birth and more during childhood and the late teen years. Today, sex-selective abortions are on the rise in India. Sen and other researchers concluded in 1990 that more than *100 million women were missing* worldwide, 80 million from China and India combined, as a result of sex-selective abortions and the relative neglect of girls and women. Today the numbers may well be higher.

Sources: Nancy E. Riley, "China's Population: New Trends and Challenges," Population Reference Bureau, *Population Bulletin* 59, no. 2 (June 2004); Amartya Sen, "More Than 100 Million Women Are Missing," *New York Review of Books* (December 20, 1990); Stephan Klasen and Claudia Wink, "'Missing Women': Revisiting the Debate," *Feminist Economics* 9, nos. 2–3 (November 2003). The cover story of the March 6, 2010, issue of *The Economist* was on gendercide.

for sons combined with policies that limited the number of children a couple could have. This most severe form of discrimination against females has human rights implications and may result in a host of social problems for future generations of Chinese women and men.

Finally, nations that today are worried about the growth in their populations are unlikely to be able to follow the Chinese approach. Few states can exercise as much social control as Chinese governments have in the recent past. India's experience in the 1970s suggests that a democracy is unlikely to accept the authoritarian approaches employed in China. But, even if a nation could do so, should they? Amartya Sen argues that, if freedom is valued at all, depriving people of the freedom to regulate their own fertility, one of life's most personal decisions, must represent a significant social loss.[31] Such a loss can be justified only if the social cost of not intervening is high. As this chapter has shown, it is hard to make the case that individuals systematically make poor decisions when it comes to childbearing or that the negative externalities of these decisions are large. If this is true, what justification is there for coercion to regulate fertility? Noncoercive approaches, including broad-based socioeconomic improvements, can achieve similar ends and should be pursued instead.

POPULATION ISSUES FOR THE TWENTY-FIRST CENTURY

In retrospect, the fears of Robert McNamara and many others about a population explosion no longer seem warranted. Population growth has slowed dramatically, and the expectation is no longer that world population will increase indefinitely. By mid-century, the world's population is expected to peak, admittedly with billions more people then when McNamara first expressed his fears.

But the drop in population growth rates does not mean that demographic concerns are a thing of the past. The world economy, if the United Nation's medium term projection comes to pass, still must accommodate another 2 billion more people over the next 40 years, almost all of whom will be born in today's developing nations. A second major demographic development will be the change in the age distribution of the world's population. There will be fewer children and many more elderly; not only are the developed nations aging but so are many of the developing ones. According to United Nations' projections, those 65 and older made up 6.7 percent of Asia's population in 2010; in 2050 they will make up 17.3 percent. In China alone, in part due to the one child policy, the percentage of seniors is projected to rise from

[31]Amartya Sen, "Population: Delusion and Reality," *The New York Review of Books* 41, no. 15 (September 22, 1994).

8.2 to 23.3 percent of the population. In Latin America the change will be similar to that in Asia. Only in sub-Saharan Africa, where fertility rates remain high, will the share of the elderly remain in single digits.

How will the needs of the elderly be met? As households grow smaller in size, traditional systems of family support are likely to breakdown. Financing old-age pensions, a huge problem now in developed nations, will increasingly confront emerging economies. Chile's pension system is often cited as a model for dealing with old age security. It no longer relies on a traditional pay-as-you-go (PAYGO) system similar to Social Security in the United States, where workers pay taxes each year that finance the pensions of those who already are retired. Starting in 1981, Chile developed a national system of mandatory personal retirement accounts, similar to 401K plans in the United States. This system of prefunded retirement savings relieves some of the burden of publicly financed old-age support. But it is not a complete solution, as many Chilean workers are self-employed or work in the informal sector and do not participate. About half of Chilean women were not covered because they did not work outside of their homes. Low-income workers may not earn enough for their accounts to lift them above poverty after their working years. Recent reforms, which include more public funding of old-age security, address some of these limitations but continue to build on the personal retirement account system.[32]

A changing age structure will challenge some economies; others will continue to face high fertility and the burdens (and opportunities) of a young population. This will be the case for many nations in sub-Saharan Africa. If this region follows what has happened elsewhere, economic growth and development will reduce the demand for children and accelerate a demographic transition to lower birth and death rates, and a slower rate of population growth.

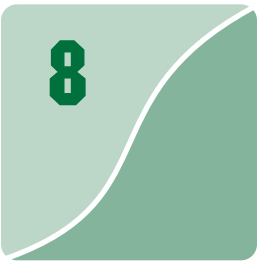
SUMMARY

- The twentieth century will be remembered for populating the planet. In 1900, the world population stood at about 1.5 billion. By 2000, there were over 6 billion people; 5 billion living in developing nations. In no prior century did the world population increase by even 1 billion. World population growth has slowed since the 1960s, and current projections do not envision that the twenty-first century will see increases as large as those of the last century.
- The driving force behind the population increases of the last 100 years is the demographic transition, especially in the developing world. Improvements

[32]Estelle James, Alejandra Cox Edwards, and Augusto Iglesias, "Chile's New Pension Reforms," Policy Report No. 326, National Center for Policy Analysis, Dallas, TX, March 2010.

in nutrition, medical knowledge, and public health led to a rapid decline in death rates. Birth rates took longer to fall, and in the interim, the rate of natural increase in the populations of low- and middle-income nations soared to unprecedented levels. Today, most regions of the world are experiencing a decline in fertility and are approaching replacement levels. Sub-Saharan Africa is the one exception, where total fertility remains high, but even in this region, some decline is evident.

- There is considerable debate over the consequences of population growth on economic development. This is because of the multitude of ways population growth affects the determinants of growth; population growth, for example, may inhibit capital deepening, but it also may encourage institutional and technological innovations that increase factor productivity.
- In trying to understand the impact of population growth on economic development it is important to reconcile micro behavior with macro outcomes. If individual couples are making rational choices over the number of children they have, then at a macro level, why might a problem arise? The answer is market failures—for example, if the social costs of raising children exceed the private costs. In practice, it has proven difficult to identify the precise circumstances in which this happens and how large the resulting costs are.
- The debate over whether population growth is an obstacle to economic development is reflected in the choice of population policies governments pursue. If household behavior is rational and market failures are small, then development may be the best contraceptive. With better education, especially for girls; better employment opportunities, especially for women; and improved healthcare, especially for the young; couples desire fewer children. Family planning programs also may influence social norms about family size and assist individuals in attaining desired fertility levels in a safe and reliable manner.
- Some governments, especially in China, decided that the macro consequences of population growth are so negative that authoritarian steps must be taken to limit fertility. Such approaches have proven effective at reducing fertility levels and rates of population growth. But their development benefits are less obvious. Coercive population policies entail high social costs and loss of individual freedom. It is difficult to justify these actions with such limited evidence that reduced fertility contributes to faster economic growth and development.
- Despite the slowing of world population growth, the twenty-first century will face demographic challenges of its own, including a rising share of the elderly. With fewer workers per nonworking senior, nations will need to craft fiscally sustainable systems of old age support.



Education

Belief in the importance of education for economic development is not a new idea. The Chinese philosopher Guan Zhong, writing in the seventh century B.C.E., advised, "If you plan for a year, plant a seed. If for ten years, plant a tree. If for a hundred years, teach the people. When you sow a seed once, you will reap a single harvest. When you teach the people, you will reap a hundred harvests."

Contemporary scholars voice similar perspectives. Nobel laureate Theodore Schultz, in his 1960 presidential address to the American Economic Association, highlighted the importance of investing in people. Schultz referred to the acquisition of skills and knowledge as investments in human capital. Economists consider resources spent on physical capital as an investment that yields a future return rather than as consumption. Schultz argued that expenditures on people should be viewed in a similar way. He believed that most of the "impressive rise" in the earnings of workers in the industrialized countries was due to the growth in human capital, and a limiting factor in the advance of poor countries was insufficient investment in people.[1]

Among the major categories of human capital investments, Schultz included formal education at the elementary, secondary, and higher levels; health services; on-the-job training; and private and government sponsored training programs. In this chapter, we focus on formal education, specifically on schooling, and its implications for economic growth and development. In the next chapter, we consider health. On-the-job training includes everything from a parent teaching a child to

[1]Theodore W. Schultz, "Investment in Human Capital," *American Economic Review* 51, 1 (March 1961), 1–17.

farm to an apprentice learning a craft from a master tradesman to a multinational corporation offering courses to newly hired or experienced workers. All these forms of on-the-job training are important sources of human capital investment, but they ordinarily take place with limited amounts of policy intervention and public spending, especially when compared to education and healthcare. Schooling also provides *general* human capital, which is a prerequisite to the *specific* human capital more often associated with different forms of training. For these reasons, we restrict our attention in this chapter to schooling.

Over the past 40 years, there has been tremendous growth in schooling in all regions of the world. It is estimated that in 1960, 55 percent of the population age 15 and over in the developing world had never attended school. By 2010, this percentage had fallen to only 17 percent. It will fall even further in the decades ahead. In many countries expansion in schooling has contributed significantly to the human capital of the labor force and subsequent economic growth and development. But in other settings, schooling has not yet delivered what Guan Zhong and Theodore Schultz predicted in terms of growth and development. It is not that the Chinese philosopher and the American Nobel laureate were wrong. Instead, education, like so many other prescriptions for economic development, is not a panacea. It is not the sole determinant of economic success. If economic conditions do not encourage productive economic activity, the demand for educated workers will be weak, and even those with schooling may struggle to generate income for themselves and their families.

After considering recent trends in schooling and learning, this chapter examines the evidence on schooling as an investment. Schooling generally displays attractive rates of return, but even when this is the case, spending on schooling often delivers less than it might. Countries often underinvest in schooling overall, misallocate resources across the different levels of schooling, and inefficiently use resources within schools themselves. In some cases, the problem is insufficient or misallocated resources; in others, it is a lack of accountability among teachers, principals, and others responsible for the education of the young. The seventh-century Chinese philosopher Guan Zhong may have been right in his overall view of education but he did not spell out the details: How does one induce parents to send and keep their children in school; how does one combat rampant absenteeism among teachers; and, even more fundamental, how does one increase the amount of learning students achieve by attending school?

TRENDS AND PATTERNS

If you are reading this book for a course in development economics, you are among a privileged group in the world who has the opportunity to attend university. You may not feel part of an especially select group as you make your way through the

crowded hallways of your college or university, but you are part of one. Worldwide, gross enrollment rates in any type of tertiary education (schooling beyond high school) amount to about one out of every four members of your age group. In the high-income countries, the rate exceeded 67 percent in 2008; in the middle-income nations, it reached 25 percent; and in the low-income countries, it was only 7.5 percent. Just as striking is a comparison with the generations that came before you. In the United States today, about one out of every three adults over the age of 25 received schooling beyond the secondary level. In France, the ratio is closer to one in five. In the developing nations it is often much lower: around 6 percent for both China and India; 7 percent in Brazil; and between 2 and 3 percent in Ghana and Kenya.[2] The experience of your age cohort, as compared to the schooling experience of earlier generations, tells us a lot about the stocks and flows of education in the world today. **Stocks** refer to the amount of schooling embodied in a population; **flows**, to the net change in those stocks as the result of school enrollment patterns.

STOCKS AND FLOWS

The scarcity of education in the developing nations is evident in the data on the levels of schooling completed among today's adult population (Figure 8–1). In the United States, an advanced economy, only 3 percent of adults did not complete six

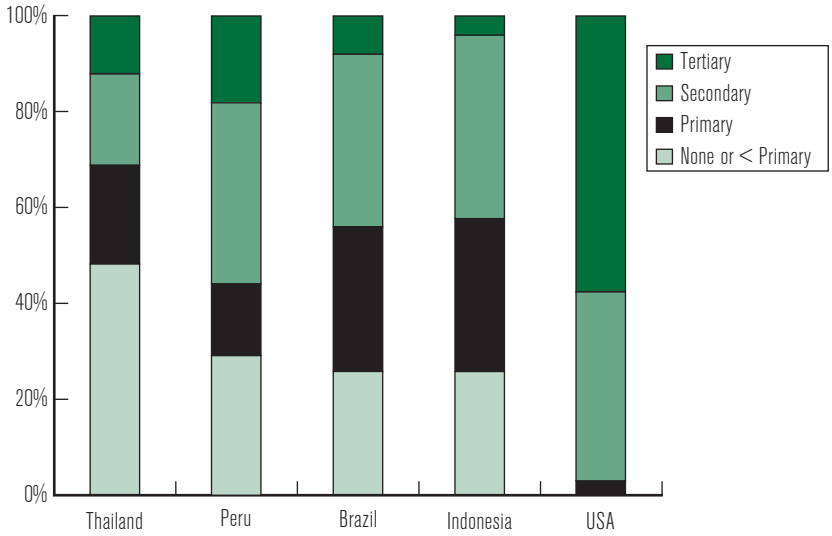


FIGURE 8–1 Educational Attainment of the Adult Population (ages 25 to 64)

Sources: Data for Thailand, Peru, Brazil, and Indonesia are from United Nations Educational, Scientific and Cultural Organization (UNESCO). Data for the United States are available from the U.S. Census Bureau.

[2]Estimates from the Barro-Lee Educational Attainment Dataset (2010), available at www.barrolee.com, accessed February 2012.

years of schooling (that is, did not complete primary school). The situation in low- and middle-income nations could not be more different. In Brazil about one out of every four adults did not attend or complete primary school; in Thailand almost one out of every two adults. Household data from the two Indian states of Uttar Pradesh and Bihar tell a similar story. By the late 1990s these two states had a combined population of 250 million people and were among the poorest states in India and hence among the poorest people in the world. Over two-thirds of their adult population did not complete primary school, most of them having never attended school.

Minimal levels of education characterize today's labor force in many low- and middle-income countries, but the situation is changing. Throughout the world, parents and governments, as well as bilateral donors and international agencies, are advocating that current and future generations receive more schooling. This trend is evident in a comparison of school enrollment rates by region over the past three decades (Table 8–1). At the primary school level, **gross enrollment rates** either have been high for decades (East Asia, Europe and central Asia, Latin America) or have expanded rapidly (Middle East and North Africa, South Asia and sub-Saharan Africa). Consistent with the millennium development goal (MDG) on achieving universal primary education, regional trends indicate real progress in increasing the amount of schooling children receive. But there is more progress to be made; in 2008, close to 70 million children were not in school.

Gross enrollment rates refer to the total number of children enrolled in a given school category divided by the number of children of the age group that officially corresponds to that level of schooling. In the case of primary school, the relevant age group usually is 6 to 11 years old. If older children (or younger ones) enroll in primary school, the gross enrollment ratio can exceed 100, as it does in several regions. Gross

TABLE 8–1 Changes in Schooling, Gross Enrollment Rates by Region, 1975–2005

REGION*	PRIMARY		SECONDARY		TERTIARY	
	1975	2005	1975	2005	1980	2005
East Asia and the Pacific	112.8	111.4	36.5	72.5	2.5	20.1
Europe and central Asia	96.9	96.9	84.7	87.9	32.9	54.7
Latin America and the Caribbean	109.1	117.7	38.3	89.7	13.1	29.6
Middle East and North Africa	81.5	106.3	31.2	75.6	10.5	26.1
South Asia	76.7	109.7	26.5	49.5	4.5	10.4
Sub-Saharan Africa	58.8	92.0	11.3	31.8	1.2	5.3
High-income economies	98.4	99.8	78.6	97.7	34.3	62.3

*Data for each region are aggregated from country-level data. Countries in the region were not included if data were unavailable.

Source: United Nations Educational, Scientific and Cultural Organization (UNESCO), Institute of Statistics, available at http://stats.uis.unesco.org/unesco/tableviewer/document.aspx?ReportId=143.

enrollment ratios that exceed 100 often are a mixed blessing. The good news is that they indicate lots of children are enrolled in school; the bad news is that they usually imply that many children are repeating grades. If repetition rates are high, as they often are in poor countries, gross enrollment rates can be high even though a substantial number of children never attend school.[3]

Alternative measures of school attendance take account of this problem. **Net enrollment rates** refer to enrollments of only those of the relevant age group. **Grade survival rates** estimate how many children actually complete a certain grade level. Estimates of these measures confirm the substantial progress nations have made in expanding primary education but also highlight that universal *completion* of primary school remains a challenge. In 2008, net enrollment ratios in low- and middle-income countries were 81 and 88 percent, respectively; in sub-Saharan Africa they were 75 percent. Estimates of grade survival rates suggest that 1 out of 5 children in some of the poorest nations did not complete five years of schooling; in some middle-income nations, more than 1 in 10 failed to do so. These failures are most pronounced among the poorest children, those with families in the bottom quintile of the income distribution. Because of these tendencies, the MDG of achieving universal primary education for all children by 2015 is unlikely to be met.

Achieving universal primary education remains an important goal. Another challenge is to expand secondary and tertiary enrollments. Huge enrollment gains in secondary education have been realized in virtually all regions (Table 8–1). Latin America's gains are the greatest, with gross secondary school enrollment expanding from approximately 38 to 90 percent between 1975 and 2005. The poorest regions, sub-Saharan Africa and South Asia, still lag the other regions, but much progress has been made in 30 years. It is not surprising that both the absolute level and expansion of tertiary enrollments remain low in almost all developing regions relative to the high income group. Even the current star performers in terms of economic growth, East and South Asia, do not stand out in terms of the level of tertiary enrollment rates achieved to date. Faced with limited resources, governments must decide whether to allocate additional resources to achieving universal primary education or to expanding opportunities at either the secondary or tertiary levels. We return to this controversial issue later in the chapter.

Trends in school enrollment rates across regions, combined with the age structure of today's developing world, tell us how human capital endowments will change in the decades ahead. Subsequent generations will be more numerous and have

[3]Paul Glewwe and Michael Kremer offer the following hypothetical example, "In a school system with 6 years of primary education, a 100 percent gross enrollment rate is consistent with 75 percent of children taking 8 years to complete primary school (because each child repeats two grades) and 25 percent of children never attending school." From "Schools, Teachers, and Education Outcomes in Developing Countries," in E. Hanushek and F. Welch, eds., *Handbook on the Economics of Education* (Boston: Elsevier, 2006).

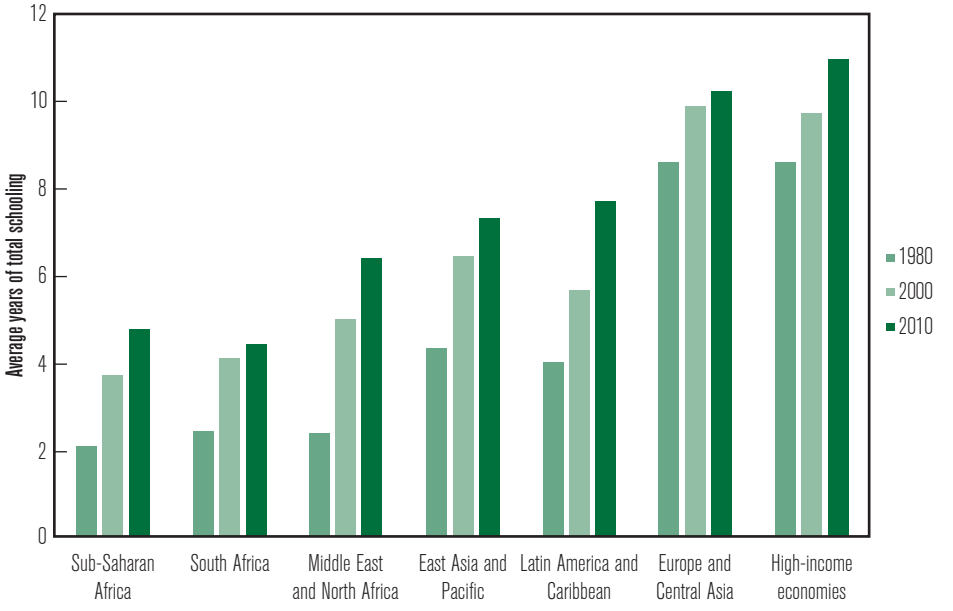


FIGURE 8–2 Education Attainment, Adult Population 25 and over

Source: Robert Barro and Jong-Wha Lee, "A New Data Set of Educational Attainment in the World, 1950–2010." NBER Working Paper No. 15902, (Cambridge, MA: NBER, April 2010). Data available at www.barrolee.com/data/dataexp.htm.

more schooling than their parents. This trend already is evident. Figure 8–2 provides estimates of the increase in the average years of schooling completed by adults 25 and over for all regions between 1980 and 2010.[4] As expected, average years of schooling have gone up everywhere. What is striking about these numbers is the extent of this increase. The educational attainment of the potential workforce in East Asia has gone from a predicted mean of 4.4 years to 7.4 years of schooling; in South Asia from 2.5 to 4.5 years, and in the Middle East and North Africa from 2.5 years to 6.5 years. The average level of educational attainment may still seem low, especially relative to the high-income economies (11 years in 2010), but remember these estimates are population means. In the developing nations, people in their 50s and 60s with no or little schooling are averaged together with people in their 20s and 30s whose educational attainment exceeds the mean. The trend is unmistakable; younger workers are entering the labor force with more years of schooling and are enriching the

[4]Calculating average years of schooling for a nation would seem a straightforward task. It is not, in part, because census and survey data in many developing nations are not conducted on a regular basis. Economists estimate educational attainment using a perpetual inventory method, which combines available census and survey data reflecting the stock of human capital with information on enrollment and repetition rates, which reflects the flow of schooling. Economists Robert Barro and Jong-Wha Lee provide the most-often-cited estimates and projections of educational attainment. Their most recent estimates are reported in Robert Barro and Jong-Wha Lee, "A New Data Set of Educational Attainment in the World, 1950–2010," NBER Working Paper No. 15902, (Cambridge, MA: NBER, April 2010).

human capital of the developing countries. It is fair to say that since 1980 there has been a global revolution in the provision of schooling.

BOYS VERSUS GIRLS

Progress has been achieved not only in the number of years of schooling received but also in the distribution of who receives this schooling. There has been considerable progress in spreading the opportunity to attend school to girls as well as boys, although girls' education still lags. Favoring the education of boys over girls is a long-standing and pervasive practice throughout much of the world, but this practice is changing. The MDGs called for eliminating gender disparity in primary and secondary education by 2005 and at all levels by 2015.

Estimates of average years of schooling for men and women over the age of 15 reveal a gender gap that has declined significantly over time. For all developing nations, the mean number of years of schooling for men in 1960 was only 3.1 years and for women even less, 2.0 years. Taking the ratio of the two, women completed only 62 percent the schooling of men. By 1980, this gap had narrowed a bit with the gender ratio at 74 percent. Much more progress has been made since 1980. For the developing world as a whole in 2010, women were estimated to have completed an average of 6.5 years of schooling as compared to 7.6 years for adult men, meaning a gender ratio of 86 percent.

Regional patterns show considerable variation (Table 8–2) around these trends. All regions have narrowed the gender gap in schooling, with Europe and central Asia and Latin America having almost achieved gender parity. South Asia, by comparison, has a long way to go. Estimates for 1980 show that adult women had received virtually

TABLE 8–2 Estimates of School Attainment for Adults 25 and over by Gender and Region, 1980 and 2010

	1980			2010		
REGION	WOMEN (YEARS)	MEN (YEARS)	RATIO (%)	WOMEN (YEARS)	MEN (YEARS)	RATIO (%)
East Asia and Pacific	3.0	4.7	65	6.8	8.0	85
Europe and central Asia	6.9	8.2	84	10.2	10.4	98
Latin America and the Caribbean	3.9	4.3	89	7.7	7.9	97
Middle East and North Africa	1.4	3.2	43	5.5	7.5	74
South Asia	0.9	2.8	33	3.4	5.6	60
Sub-Saharan Africa	1.5	2.9	53	4.0	5.7	70
High-income economies	8.1	8.9	91	10.9	11.2	97

Source: Robert Barro and Jong-Wha Lee, "A New Data Set of Educational Attainment in the World, 1950–2010." NBER Working Paper No. 15902, (Cambridge, MA: NBER, April 2010). Data available at www.barrolee.com/data/dataexp.htm.

no schooling (0.9 years), only one third of what adult men received (2.8 years). By 2010, outcomes had improved. Women averaged 3.4 years of schooling, now about 60 percent of what men completed. South Asia's gender gap is expected to continue to narrow. Today in South Asia, gross primary school enrollment rates for boys and girls both exceed 100 percent, with the rate for girls equal to 95 percent those of boys; the secondary school enrollment rate is 56 percent for boys and 48 percent for girls. When these young people move on to replace their parents and grandparents in the labor force, the region's stock of human capital will be larger and its distribution between men and women relatively more equal.

SCHOOLING VERSUS EDUCATION

Years of schooling and the gender distribution of that schooling are two ways of charting the accumulation of human capital. But schooling is only a means to an end; the real goal is education—that is, the skills individuals acquire from time spent studying and learning. Students can sit through many years of schooling and learn very little. Measuring learning outcomes is notoriously difficult within one nation let alone across many. But there have been attempts to apply basic tests of core competencies in reading and mathematics across nations. The results are sobering because of both the gap between rich and poor nations and the implications for the quality of schooling available in many low- and middle-income nations.

One source of cross-country information on learning outcomes is the Programme for International Student Assessment (PISA). Initially, PISA administered examinations only among Organisation for Economic Co-Operation and Development (OECD) members—that is, mostly high-income economies. But interest in their work led to an expansion of testing to some non-OECD and mostly middle-income settings. PISA's goal is to assess how well 15-year-olds, approaching the end of their compulsory schooling years, are prepared "to meet the challenges of today's societies." Representative samples of 15-year-olds attending school are selected from participating nations, with the total number of students tested ranging from 4,500 to 10,000 per country. Average scores on the 2009 reading test for a subset of the more than 60 participating nations are presented in Figure 8–3. Korea had the highest mean score, with Korean students performing better than many nations with considerably higher incomes. Finland was 2nd; the United States ranked 16th (tied with Poland).

The middle-income nations in the table have scores that are almost always lower than the high-income nations. This is not surprising. We expect these scores to be income elastic. At higher incomes, economies have more to spend on schooling and we believe there is some association between school expenditures and student learning. At higher incomes, children may be healthier and more able to learn. They also are more likely to have parents who attended school, another predictor of student performance. Of course, income is not all that matters. Korea and Poland's superior performance relative to much richer nations cannot be due to income.

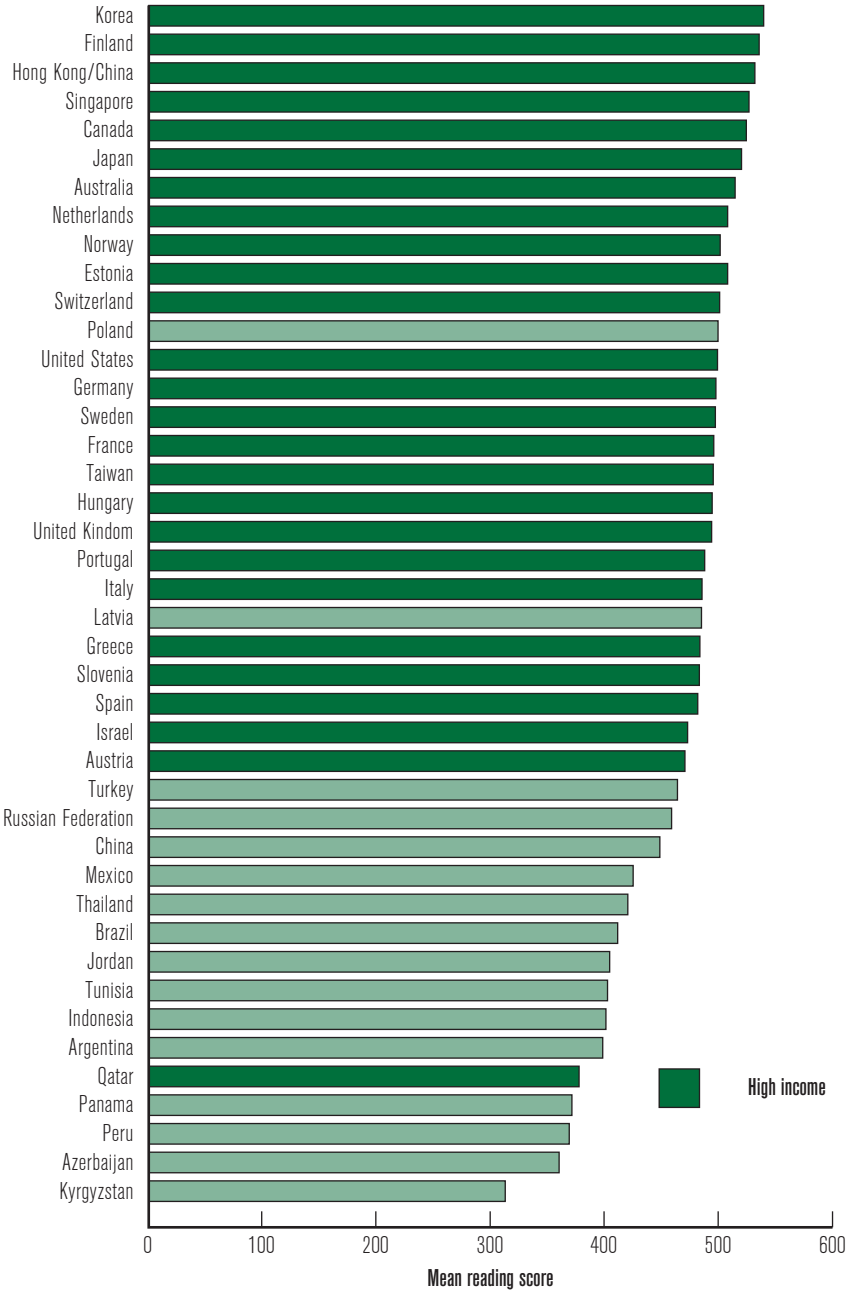


FIGURE 8–3 Learning Outcomes: Reading Achievement of 15-Year-Olds, 2009

Source: Organisation for Economic Co-Operation and Development (OECD), *PISA: 2009 Results: Executive Summary*, available at www.pisa.oecd.org/dataoecd/34/60/46619703.pdf, accessed February 2012.

What is alarming about the findings is not that middle-income nations have lower scores, that was expected, but how much lower these scores are relative to high-income economies. On the reading achievement test, 19 percent of all OECD students failed to achieve proficiency Level 2: a level low enough that students "lack

the essential skills needed to participate effectively and productively in society."[5] Raising the percentage of 15-year-olds in OECD countries to Level 2 proficiency is a formidable challenge. But the challenge is so much greater in the middle-income nations. Among the 26 middle-income nations in the PISA data set, the median percentage of 15-year-old boys who did not achieve Level 2 proficiency in reading was 50 percent; for girls it was 33 percent. For boys in Indonesia it was 65 percent; in Peru, 70 percent. In the only low-income nation in the sample, the Kyrgyz Republic, it was 88 percent. Clearly, the education a 15-year-old boy received in the Kyrgyz Republic had little in common with a 15-year-old boy in Korea.

Results on mathematical skills show an even larger gap in learning between students in OECD countries and those outside of the OECD group. Because most middle-income nations have not achieved universal secondary school enrollments, PISA results *overestimate* the measured abilities of all their 15-year-olds. Those who are not in school are unlikely to know as much as those attending school; and the PISA exam was given to only those in school! If PISA surveyed more low-income economies, the difference in learning outcomes with the advanced economies would likely be even starker.

It is possible that the disparity in results is due to cultural bias, because any test is likely to suffer to some degree from this problem, but the results on student performance across countries reported by the PISA study are similar to those from other evaluations. One study from Bangladesh found that 58 percent of a sample of rural children age 11 and older failed to identify seven out of eight letters; a similar percentage had difficulty identifying numbers and geometric shapes. A 2007 nationally representative survey of third-graders in India found that half could not read a simple first-grade paragraph. One of the questions from the Third International Mathematics and Science Study asks the following:

Three-fifths of the students in a class are girls. If 5 girls and 5 boys are added, which statement is true? (a) There are more girls than boys; (b) There are the same number of girls as boys; (c) There are more boys than girls; (d) You cannot tell whether there are more girls or boys from the information provided.

Approximatley 82 percent of Japanese and 62 percent of U.S. eighth-graders provided the correct answer (a). But only 31 percent of Colombian and South African eighth-graders identified the correct answer. With four choices available, random guessing would result in the correct score 25 percent of the time; the schoolchildren in Colombia and South Africa performed only marginally better.[6]

[5]OECD, *PISA at a Glance* (Paris: OECD, 2010), p. 12.

[6]The Bangladesh results are from Vincent Greaney, Shahidur Khandker, and Mahmudul Alam, *Bangladesh: Assessing Basic Learning Skills* (Dhaka: University Press, 1999) as reported by Glewwe and Kremer, *Handbook on the Economics of Education.* The Trends in International Mathematics and Science Study results are cited in Lant Pritchett, "Access to Education," in Bjorn Lomborg, ed., *Global Crises, Global Solutions* (London: Cambridge University Press, 2004).

That learning outcomes in developing nations, on average, are below the levels prevailing in developed nations is not a surprising finding. Rich nations have many advantages over poorer ones. Given the rapid expansion in school enrollments over the past few decades, many developing nations had no choice but to focus on expanding quantity, not on improving quality. There simply could not have been, for example, enough experienced teachers and principals to handle the primary and secondary schooling needs resulting from the increase in school enrollments. Anyone with firsthand experience of schools in poor countries, especially in rural areas, is familiar with the problems: too many unqualified teachers, endemic absenteeism among students *and* teachers, and too few books and other teaching materials. In such environments, it is no wonder that learning outcomes are so poor.

EDUCATION AS AN INVESTMENT

Despite problems of low quality, the demand for education remains high in most countries. Parents want their children to have better lives and often see getting an education as a way to achieve them. In terms of material well-being, what parents perceive is borne out by the data. On average, people with more schooling earn more than people with less schooling. Few economic outcomes are as robust as the relationship between earnings and schooling. Primary school graduates tend to earn more than those with no schooling, secondary school graduates earn more than their counterparts who completed only primary school, and those with tertiary education tend to earn the most. Of course, these are average tendencies. A professor with many years of schooling may earn less than a plumber who went to school for fewer years, but on average, the relationship holds. It holds for both men and women, it holds in fast growing economies and stagnant ones, and it holds across income levels and regions.

Figure 8–4 illustrates the relationship between schooling and earnings for one country, Nicaragua. Using household data, Figure 8–4 charts **age-earnings profiles** for men and women separately, by level of education attained. Workers at all schooling levels tend to see their earnings rise quickly in their early years, reach a plateau, and sometimes fall off in later years. The more educated the worker, the higher the age-earnings profile tends to be. In Nicaragua, the earnings premiums for more educated men are unambiguous. As level of schooling rises, each age-earnings profile lies above the one that preceded it. For Nicaraguan women, the trends are not as definitive due, in part, to the relatively small number of observations on working women in this data set. A larger sample would most likely produce the same non-overlapping earnings profiles that characterize Nicaraguan men. Comparison of the earnings of Nicaraguan men and women reveals one other relationship. At every education level women tend to earn less than their male

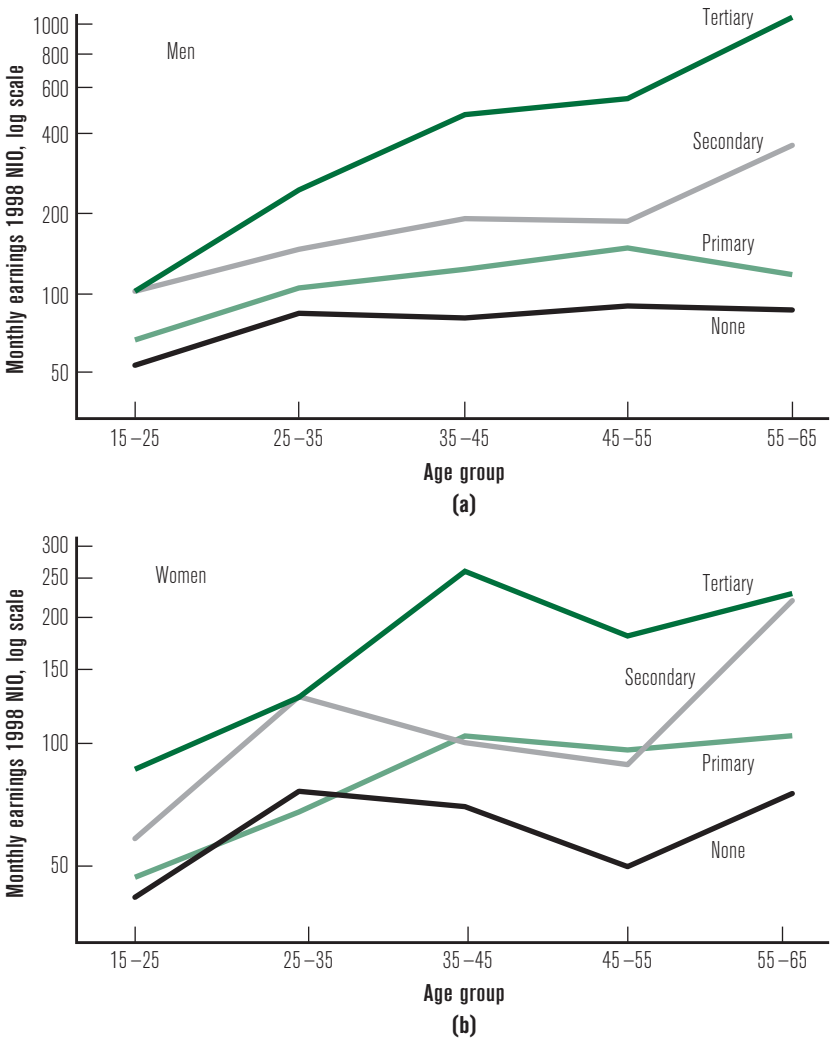


FIGURE 8–4 Earnings by Age, Education, and Gender, Nicaragua, 1998

Source: World Bank, Nicaragua Living Standards Measurement Study Survey, Encuesta Nacional de Hogares sobre Medición de Niveles de Vida (ENMV), 1998, available via www.worldbank.org/lsms/country/ni98/ni98docs.html#English.

counterparts, a worldwide result almost as common as the tendency for earnings to rise with education levels.

If Figure 8–4 were reproduced for any country the results would be similar: People with more years of schooling tend to earn more than people with fewer years. This suggests thinking about education as an investment. Economists define investment as the flow of resources into the production of new capital. In the case of schooling, resources by both governments and households are spent in the expectation that schooling produces human capital. The expectation is also that these investments will yield a positive return. By attending school, an individual hopes to acquire

human capital, which makes that individual more productive and, therefore, better compensated. This relationship holds whether the individual becomes a farmer or trader, laborer or artisan, professional or government official.

There are several benefits to thinking of education as an investment. Schooling competes with many other activities on which governments and households spend their scarce resources. How does a government decide how much to devote to schooling? Should more resources be spent on schools as opposed to health clinics, tube wells for clean water, or rural roads? Even within the education sector, how should a government allocate its spending: Is achieving universal primary education a better use of resources than expanding colleges and universities? Governments must make these decisions and thinking about education as an investment can help. Households make similar decisions. When education is viewed as an investment it becomes easier to understand why some families send their children to work and not to school, while other families do everything in their power to make sure a child receives an education.

THE RATE OF RETURN TO SCHOOLING

Figure 8–5 illustrates the factors that determine the **private return** on schooling. The figure presents two stylized streams of future earnings, one if the individuals complete only primary school and the other if they complete a secondary education. All individuals are assumed to complete primary school at age 11. If they

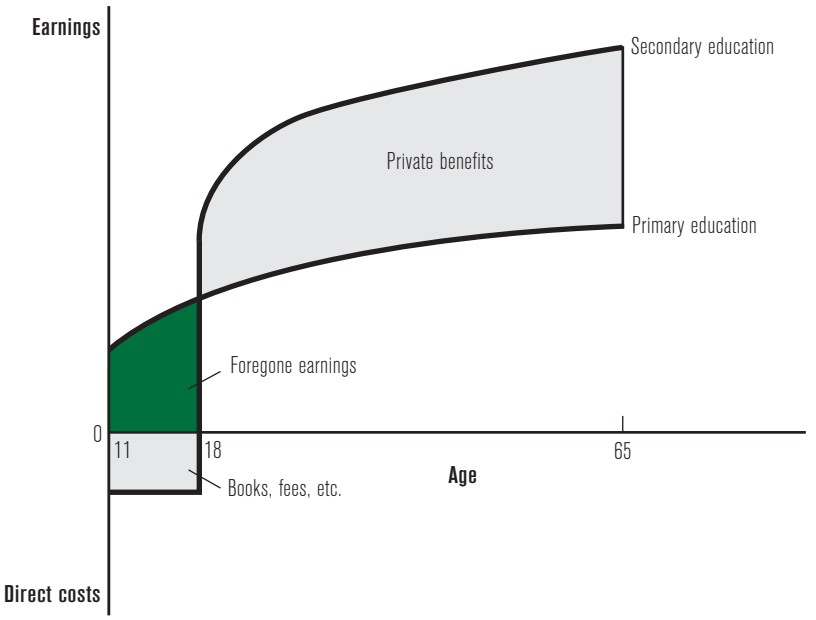


FIGURE 8–5 Determinants of the Private Returns to Schooling

do not continue with their education, they begin to work and follow the earnings profile marked as primary education. Alternatively, individuals might complete secondary education and remain in school until age 18. For ease of exposition, we assume that individuals do not work while in school and, therefore, have zero earnings while in school. Once secondary school is completed, these individuals follow the higher of the two earnings profiles. The private pecuniary benefits of a secondary education, as we saw in Figure 8–4, are the earnings the individual expects to receive beyond what they would have earned with only a primary school education.

But costs also are associated with the decision to remain in school. These costs include any foregone earnings, sacrificed because an individual is in school and not working. These costs may seem low for a young person considering attending primary or secondary school. But for the majority of households in the developing nations, the opportunity costs of sending a child to school are real. In rural households, for example, children work in the fields, take care of animals, or watch siblings, permitting others, including adults, to be engaged in farming or other productive activities. In addition to foregone earnings, households face direct costs in sending a child to school. These costs may include school or uniform fees, payments for books and other materials, transportation costs, or other "unofficial" fees to ensure a child gets the attention of a teacher. Although the sums may seem small, to a poor family they may be overwhelming.

To think of education as an investment, we need a systematic way to compare the costs and benefits of schooling depicted in Figure 8–5. Economists argue that one cannot simply add up the costs and compare them to the benefits because money received in the future is worth less than money that can be spent today. Such positive time preference is the result of both uncertainty about the future and the opportunity cost of resources—namely, the returns an alternative investment might yield. Future benefits need to be **discounted** to compare them to current costs. One way of doing so is to compute the **present value (PV)** of all costs and benefits.[7]

Completing secondary school results in earnings whose PV can be computed as

$$\mathrm{PV}_B = \sum_{t=1}^{n} B_t/(1 + i)^t \qquad [8\text{–}1]$$

[7]The need to discount future benefits can be motivated by the following example. If you are asked to trade $100 today in return for $100 one year from now, you likely will not accept the offer. Beyond any uncertainty you may have about being paid back in 12 months, you could do better by depositing your $100 in a relatively safe alternative, like a bank account. If that account offers you a 5 percent return, your $100 today is worth $105 one year from now, $110.25 in two years, and so forth. This is the familiar process of compounding. Discounting future benefits employs the same logic but working in reverse. The present value of some future benefit is the amount of money that, if available today, by virtue of compounding, would equal that amount. If the opportunity cost of your capital is 5 percent, the present value of $105 received one year from now equals [$105 ÷ (1 + 0.05)] = $100; similarly, the present value of $110.25 received in two years also is $100 since [$110.25 ÷ $(1 + 0.05)^2$] = $100.

where PV_B equals the sum of the present value of all future private benefits (the earnings differential received with more schooling) over an individual's n years of working; B_t equals the benefits in year t; and i is the interest rate (the household's opportunity cost of capital), which discounts future benefits. A similar calculation can be made for the private costs of schooling (or for any other investment option):

$$PV_C = \sum_{t=1}^{n} C_t/(1 + i)^t \qquad [8–2]$$

where PV_C now refers to the sum of the present value of all anticipated private costs and C_t equals the costs, including foregone earnings while attending school plus any direct costs paid by the household, incurred in each of the t years.

There are several ways one can then determine whether a particular investment is worthwhile. In the analysis of education, a common way is to estimate an internal rate of return for the investment. The **internal rate of return** is a derived interest rate, r, which equates PV_B to PV_C. In other words, the level of r such that

$$\sum_{t=1}^{n} (B_t - C_t)/(1 + r)^t = 0 \qquad [8–3]$$

Even though the dollar value of the private benefits of an education is likely to far exceed the dollar value of the private costs, a value for r can be found because future benefits, occurring many years from now, are much more heavily discounted than are costs incurred more immediately. Once the private rate of return to schooling is estimated, it can be compared to the rates of return available to other household investments. This helps rank the economic benefits of education as compared to alternative uses of household resources.

Schooling represents not only a private investment; it also is a social one. We, therefore, can define a **social return** to schooling. Figure 8–6 presents a schematic representation. On the cost side, schooling entails more than the foregone earnings of the individual and the payments of households. Primary school is "free" in most parts of the world, yet someone pays for teachers, the construction of schools, and the like. When evaluating education as an investment from society's perspective, these costs must be taken into consideration. The social return includes all the costs entailed in the provision of schooling. On the benefit side, schooling benefits the individual through higher earnings but schooling may also produce a positive **externality**: benefits that accrue to members of society above and beyond the benefits to the individual who receives the education. Potential positive externalities from school include health spillovers (the children of educated mothers tend to be healthier, which benefits children in other families because it reduces the transmission rate of certain diseases), reductions in crime, and more-informed political participation and decisions. More schooling, especially higher education, may also lead to technological progress that is not fully captured by private returns. To calculate a social rate of return to education use equation 8–3, but all the benefits of schooling, not just private earnings, and all the costs, private and public, should be included.

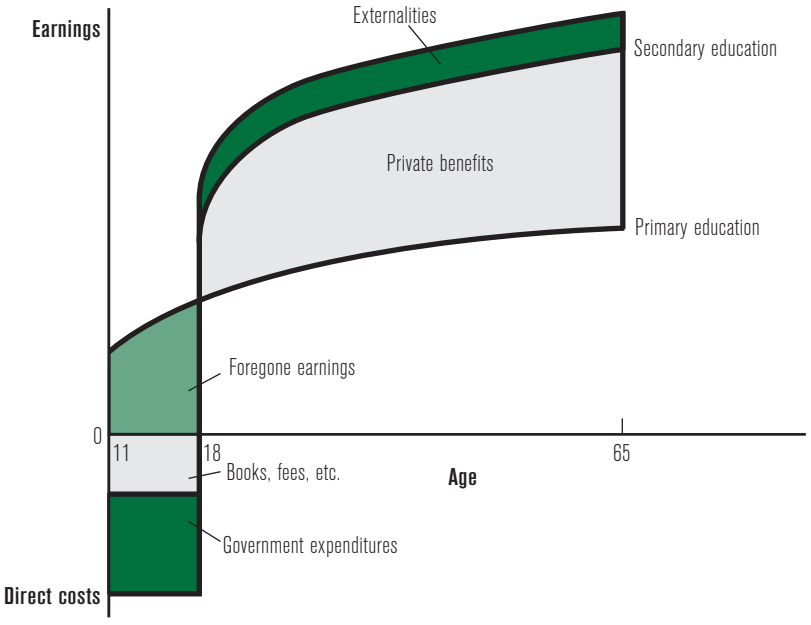


FIGURE 8-6 Determinants of the Social Returns to Schooling

Note: For ease of exposition, the figure is drawn without any positive externalities arising from primary education.

ESTIMATED RATES OF RETURN

Moving from the conceptual to the empirical, a large number of rate-of-return estimates for schooling are available for both developed and developing nations. To interpret these estimates it is important to be familiar with many of their implicit assumptions. First, there is no way of knowing what anyone's future earnings will be. Economists use the current earnings of individuals at various ages as an estimate of the expected age-earnings profile of someone with a particular level of education. Substituting such cross-sectional findings (comparing wages for different ages at one point in time) for what may happen over time to the wages of individuals as they grow older is a common empirical technique in economics; however, especially in a rapidly changing economy, estimates based on cross-sectional findings may not be a good proxy for events that will occur in the future.

Second, estimates of the return on education require some measure of individual earnings. In low-income nations and in many middle-income nations, the vast majority of workers do not work for wages. They are self-employed, often subsistence farmers or unpaid family workers engaged in petty trading or other informal sector activity. Such workers do not report the kind of earnings data included in standard rate of return on schooling estimations. By relying only on those who receive earnings, most rate-of-return estimates are based on a fraction of a nation's labor force.

Third, problems of **identification** confront most empirical studies. Returns to schooling are no exception. If individuals who have more years of schooling earn more than those with fewer, how do we know schooling caused the higher earnings? Perhaps people who are innately more talented both earn more and receive more schooling. Is the increase in earnings a result of their schooling or their innate abilities? If we do not control for innate ability—something that is not easy to measure—and for many other variables that might influence future earnings, how can we be certain that our estimates of the rate of return to schooling are correct?

FIRST-GENERATION ESTIMATES

Economists have developed different techniques for estimating rates of return to education, and it is instructive to see how methodological advances have replaced earlier approaches and dealt with problems of identification. Some of the first studies have been compiled in meta-analyses by World Bank economists George Psacharopoulos and Harry Patrinos. Table 8–3 is from their work and includes calculations of internal rates-of-return for over 75 nations. Some of these studies refer to outcomes as far back as the late 1950s, whereas others refer to the 1990s.[8] Rates of return estimated for individual countries are averaged together and presented as a mean value for different income categories and regions. These are simple cross-country averages and are not weighted by population size. They are *real* not nominal rates of return because they are based on cross-sectional estimates that measure earnings in the same current dollars.

A number of patterns can be observed in these data. Returns on all levels of schooling and in all income categories are almost always in double digits. Returns

TABLE 8–3 Returns to Schooling by Level and Country Income Group

INCOME	PRIVATE RATE OF RETURN			"SOCIAL" RATE OF RETURN*		
CATEGORY	PRIMARY	SECONDARY	TERTIARY	PRIMARY	SECONDARY	TERTIARY
Low	25.8	19.9	26.0	21.3	15.7	11.2
Middle	27.4	18.0	19.3	18.8	12.9	11.3
High	25.6	12.2	12.4	13.4	10.3	9.5

*The estimates of social rates of return account for only government expenditures on schooling and do not include estimates of any positive externalities resulting from schooling.

Source: G. Psacharopoulus and H. Patrinos, "Returns to Investment in Education: A Further Update," *Education Economics* 12, no. 2 (August 2004).

[8]George Psacharopoulos and Harry Patrinos, "Returns to Investment in Education: A Further Update," *Education Economics* 12, no. 2 (August 2004).

to schooling also tend to be greater for developing versus developed economies. This may seem surprising given the much higher wages and salaries workers earn in high-income nations. But, remember, rates of return measure something different from the level of earnings. Take the case of graduates of tertiary education. Given the relative scarcity of students with a tertiary education in poorer nations, the pay premium to having such an education may be greater than the relative pay gap between university and high school graduates in richer nations. It is the relative scarcity of labor skills, a combination of the strength of labor demand and the extent of labor supply, that contributes to the attractiveness of schooling as an investment. In poorer nations, educated workers are relatively scarce, often making schooling (at the primary, secondary, and tertiary levels) an investment with a higher rate of return than in advanced economies, where educated workers are far more abundant.

Looking across levels of schooling, returns are highest for primary schooling. This is probably because the opportunity cost of attending school in terms of foregone earnings is lowest when attending primary school and rises thereafter. Returns to secondary school are below or close to those of a tertiary education. This may be due to much larger pay differentials between tertiary and secondary school graduates versus between secondary and primary school graduates. Or it may reflect opportunities tertiary school students have to earn some income while studying (so there is less cost from foregone earnings), including receipt of government stipends for attending a school, a common practice in many developing nations, especially during the time period of the studies represented in the table.

Estimated "social" returns in Table 8–3 are lower than private ones at every level of schooling. Theoretically this need not be the case (see Figure 8–6). If positive externalities are sufficiently large, social returns could exceed the private rate of return even if social costs are greater than private costs. However, no direct means are available for estimating the positive externalities of schooling; the average "social" rates of return to schooling reported in Table 8–3 account for only social costs, using private not social benefits; hence our use of quotation marks around the word *social*. By incorporating the full cost of schooling but not any positive externalities, "social" returns for a given level of schooling must be less than the corresponding private return. Note also that the relative gap between private and "social" returns is especially large for tertiary schooling in low- and middle-income nations. This reflects the relatively high per pupil cost and high degree of state subsidy for tertiary education, often including free tuition. Even though the "social" returns are below private returns, the rates are sufficiently high to make schooling an attractive public investment compared to many alternatives, including investments in physical capital.

Another approach to estimating the rate of return to education is to estimate wage equations by which regressions are run to determine the impact on earnings of schooling adjusting for age and other demographic characteristics. The coefficient on schooling can be interpreted as an estimate of the private return to an additional

year of schooling (Box 8–1). The return is based on all years of schooling rather than distinguishing between primary, secondary, and tertiary. Psacharopoulos and Patrinos report on these estimates as well and find the range for low- and middle-income nations by region is 7 to 12 percent per additional year of schooling. The attractiveness of such returns can be illustrated by comparing them to estimates of the return on other investment opportunities. For example, U.S. government securities, one of the world's safest investments, have yielded an average real (inflation-adjusted) return of about 1 percent per year over the last 50 years; an index of the U.S. stock market, a riskier investment, over the same period yields an annual real return of 2 to 3 percent. Of course, most families do not choose between investing in U.S. stocks or government securities versus sending a child to primary school; a comparison of these rates of return is intended only to put into perspective the relative magnitudes and to highlight the attractiveness of human capital as an investment.

BOX 8-1 ESTIMATING RATES OF RETURN FROM WAGE EQUATIONS

Conceptually, the notion of rates of return to schooling is relatively straightforward. Estimating such returns is more complex. One of the most common approaches is to estimate the impact on earnings of schooling, age, and other demographic characteristics. Data on these variables can be obtained from either household or firm surveys and used to estimate a human capital earnings function or wage equation like the following:

$$\ln E_i = \alpha + \beta_1 S_i + \beta_2 \exp_i + \beta_3 \exp_i^2 + \varepsilon_i$$

where E_i refers to the earnings or wages of each individual; S_i, the years of schooling completed; and $\exp_i$ and $\exp_i^2$, the years of work experience of the individual and its square, often approximated by the individual's age and its square. The reason for including the squared term is to capture the expected nonlinearity in age-earnings profiles because earnings tend to rise at a decreasing rate over a worker's lifetime. The regression equation also includes an error term, ε_i.

In this specification of the wage equation, the coefficient on schooling (β_1), where $\beta_1 = (\partial E/E)/\partial S$, provides an estimate of the average percent increase in earnings received by workers per additional year of schooling. It is an estimate of how wages in an economy vary by education for the year in which the data are obtained.

The term β_1 is also interpreted as the average annual private rate of return to one additional year of schooling, regardless of the level of schooling already attained. This interpretation requires a number of assumptions, including the

same assumption used in the internal rate-of-return calculation: that earnings differentials by education *in a cross-section* are a good approximation of what will happen to pay differentials *over time* as workers age. Another assumption (one not required by the internal rate-of-return approach) requires that foregone earnings represent the only private cost of schooling. In some instances these are reasonable assumptions, in others not.

Estimates of wage equations tend to find a higher return to schooling for women than men.[9] This does not mean that women earn more than men; we already noted that for a given level of education men, on average, always earn more than women. Once again, the higher return indicates that the *relative benefit*, in terms of earnings, of an extra year of schooling is higher for women than men. There is further evidence to believe that the social return to educating women may exceed that of men because of the positive health and fertility externalities that are generated. This is because educating women reduces child mortality more than educating men does; educating women reduces fertility and, in so doing, reduces maternal mortality; and educating women is more effective than educating men in reducing the spread of HIV/AIDS. A 1992 study concluded that providing 1,000 girls in India with an extra year of primary school would avert 2 maternal deaths, 43 infant deaths, and 300 births. The cost of providing this education was about 60 percent of the estimated discounted social benefit, suggesting a huge social rate of return on sending girls to school.[10] This is powerful evidence in support of the MDG of reducing gender disparities in schooling.

SECOND-GENERATION ESTIMATES

The estimates of internal rates of return to education presented in Table 8–3 are subject to some debate. We already pointed out the limitations of using cross-section results to approximate a time series event, especially in environments in which labor market outcomes are changing rapidly. Another concern is the quality of studies included in the Psacharopoulos and Patrinos compilations. One critical review of rate-of-return estimates for Africa found many of the original studies were based on

[9]In the survey by Psacharopoulos and Patrinos, the average return to a year of schooling for women was 9.8 percent for women and 8.7 for men. In a different study, Peter Orazem and Elizabeth King use harmonized household data sets from 48 developing countries to estimate wage equations separately for women and men. The average return for women was 9.7 percent, for men 6.7 percent. Even more telling than the average, out of 71 cases the return was higher for women 59 times and for men only 5 times. "Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy," in T. Paul Schultz and John Strauss, eds., *Handbook of Development Economics*, vol. 4 (Amsterdam: Elsevier, 2008).

[10]Lawrence Summers, "Investing in All the People: Educating Women in Developing Countries," EDI Seminar Paper No. 45 (Washington, DC: World Bank, 1994).

guesstimates of prevailing wages by consultants rather than on data from surveys. Another problem was limiting observations to those workers earning wages—that is, to those employed in the formal sector.[11] This is especially tricky when many formal sector jobs are in the public sector where wage setting may not reflect productivity criteria as is more likely in the private sector. Private rates of return to primary education for Africa in the Psacharopoulos and Patrinos sample range from 7.9 percent in Tanzania in 1991 to 99 percent in Botswana and Liberia in 1983. This large a range in estimates strains the creditability of the reported findings. Better data and more careful analysis might result in smaller estimates of the average returns to schooling, especially for low-income nations.

Another criticism of studies that form the basis of the results reported in Table 8–3 is the failure to account for all the school costs facing families. In Honduras, for example, the private return to primary school is estimated at 21 percent, if it is assumed that there are no direct costs to the family of attending school and if the child is considered to have foregone earnings only during the last two years of primary school, at ages 11 and 12. If the actual direct costs for uniforms, school supplies, transportation, and other parental contributions are included, the private returns drop to 16 percent; if foregone earnings are added for age 10, the returns drop to below 15 percent.[12] These criticisms challenge the magnitude of some of the reported rates of return to schooling but generally do not reverse the conclusion that schooling has a sufficiently high rate of return to make it a good investment for both the individual and society.

An additional criticism concerns the identification problems mentioned earlier that plague many of the studies that contributed to the first-generation estimates contained in Table 8–3 and to many of the estimates derived from the wage equations discussed in Box 8–1. One solution to these identification problems is to identify a **natural experiment** in which a set of circumstances permits the researcher to better isolate the effect of an intervention on one group (the treatment group) versus another (the control group). In a seminal paper, MIT economist Esther Duflo identifies a natural experiment in Indonesia that permits estimation of the return on schooling. Following the oil boom of 1973, the Indonesian government undertook a massive construction program, resulting in over 60,000 new primary schools. It stands as one of the world's fastest school-building programs ever undertaken. The natural experiment involves the subsequent educational attainment and adult wages of those who were 12 to 17 years old in 1974 and finished school just before the program began and were therefore not impacted by it versus those who were 2 to 6 years old in 1974 and immediately benefited from the new schools. Because the location of the new schools was not random, there is another dimension to the natural experiment involving the region where the

[11]Paul Bennell, "Rates of Return to Education: Does the Conventional Pattern Prevail in sub-Saharan Africa?" *World Development* 24, no. 1 (January 1996).

[12]Patrick McEwan, "Private Costs and the Rate-of-Return to Education," *Applied Economic Letters* 6 (1999).

child lived. Using state-of-the-art econometrics, Duflo is able to identify the expansion in the supply of schools as responsible for an increase in the number of years of schooling received by the treatment group and the impact of this increase on their earnings some 20 years later relative to the control group. She estimates a return on schooling ranging from 6.8 to 10.6 percent, results consistent with many earlier estimates.[13]

Results of other studies that grapple with identification problems reach somewhat different conclusions. In a recent review article, University of Pennsylvania economist Jere Behrman draws attention to problems stemming from omitted variables, including measures of innate cognitive skills, early-life nutrition and school quality. Studies that attempt to account for these determinants find that in addition to years of schooling, the quality of the school and the ability of the student have a significant effect on future earnings. When these variables are included in the regressions, the estimated return on *years* of schooling declines, sometimes by a significant amount.[14] Better data and improved econometric techniques are generating refined estimates of the return not only to years spent in school but also to other dimensions of schools and attributes of students. While differences in magnitudes remain, these newer estimates continue to support the notion that education is a worthwhile investment.

PUZZLES

Beyond disagreements about how high the rate of return to schooling is for different levels of schooling and for different countries and regions, several puzzles emerge about the relationship between these returns and both microeconomic and macroeconomic outcomes. One of the microeconomic puzzles concerns schooling and learning. The key issue is why does education have an economic return? The usual answer is that education provides the individual with cognitive skills, skills learned as distinct from innate ability, which makes him or her more productive in the marketplace. Higher productivity, in turn, is associated with higher compensation, a result that holds if one is discussing farm labor or professionals (Box 8–2). If schools produce cognitive skills, then we expect to observe a positive correlation between years of schooling and earnings and a positive rate of return on attending school. The puzzle involves results reported earlier in the chapter that reveal schooling in many developing countries often produces little in the way of learning. If true, why in these settings is there a continued association between schooling and earnings and significant rates of return to education? One possibility is that schools may have been better in the past, accounting for the higher earnings of older workers today. If school quality is more of a recent problem, it is less likely to show up in the cross-section studies

[13]Esther Duflo, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review* 91, 4 (September 2001).

[14]Jere Behrman, "Investment in Education: Inputs and Incentives," in Dani Rodrik and Mark Rosenzweig, eds., *Handbook of Development Economics*, vol. 5 (Amsterdam: Elsevier, 2010).

relating schooling to earnings that combine younger and older workers. And if school quality is a more contemporary problem, as younger workers age they may not realize as much of a return on their education as did today's older workers.

BOX 8-2 RETURNS TO SCHOOLING AND INCOME OPPORTUNITIES

In a series of studies, Andrew Foster and Mark Rosenzweig, economists at Brown and Yale Universities, respectively, examine the interaction between returns to schooling and technological change in agriculture. Their core argument is that schooling has a return primarily when new income-generating opportunities arise. In one study of worker productivity, data were assembled on workers engaged primarily in harvesting in Bukidnon, the Philippines. Controlling for a worker's physical ability, proxied by gender and height, there was no impact of the worker's education on the wage received. In this situation, workers performed a routine task and little was gained from additional schooling.

A different outcome was observed in India. In a study of over 4,000 rural households between the years 1968 and 1981, the authors assessed the impact schooling had on farm profitability. The time period studied covered the introduction of the green revolution to India, when imported high-yield variety (HYV) seeds were introduced that could dramatically increase farm output (usually because it permitted double and even triple cropping during a calendar year). HYV seeds are particularly sensitive to the use of complementary inputs, including irrigation and fertilizer, and it was especially important for farmers to be able to learn how to use these seeds because the required farming techniques were different from traditional methods.

In areas where conditions were suitable for the introduction of the HYV, farmers with primary schooling had higher profits than those without schooling, holding all other inputs constant. Investments in education had a high return to these farmers because of increased returns to learning and new information. In other areas of the country, including the Indian state of Kerala, where farmers had above average years of schooling, conditions were not suitable for the new seeds. These farmers got little return because there were no new opportunities that required education.

Rosenzweig concludes that "schooling returns are high when the returns to learning are also high." This can be because of a new technology, like the green revolution or new opportunities created by changes in market or political regimes. Not only do those with more schooling reap a return in these situations but they also tend to invest more heavily in the schooling of their children. School

enrollment rates increased more rapidly in those Indian communities that benefited from the technological change than in those that did not.

Sources: Andrew Foster and Mark Rosenzweig, "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review* 86, no. 4 (September 1996); Mark Rosenzweig, "Why Are There Returns to Schooling?" *American Economic Review* 85, no. 2 (May 1995).

One of the macroeconomic puzzles results from trying to reconcile another set of empirical observations. Earlier in this chapter, we reported on trends in schooling by region. The evidence is clear. There has been rapid growth in schooling throughout the world over the past decades, resulting in some convergence in years of schooling per worker across countries. Given the rates of return to education reported by Psacharopoulos and Patrinos and others, such an expansion of schooling should have produced more rapid aggregate economic growth in many regions and some convergence in incomes. This is expected because schooling has grown much more rapidly in developing than developed nations and the returns to schooling are higher in low- and middle-income nations than in high-income settings. But this is not what the aggregate data show. Instead, there has been considerable divergence in per capita incomes at the same time there has been some convergence in schooling.

Schooling, of course, is not the only determinant of aggregate growth, and other factors may be more important in determining growth rates of gross domestic product (GDP) per capita. But there remains something disconcerting about the micro evidence on the accumulation of schooling and on the attractive rates of return to these investments and the subsequent lack of evidence that increased schooling results in significantly higher economic growth rates. Exploration of this puzzle led one researcher to ask, "Where has all the education gone?" and then to speculate that, in environments not conducive to growth, those with more education may engage in rent seeking and other unproductive activities that are privately remunerative but "socially dysfunctional."[15] This may be part of the explanation, but a fuller reconciliation of micro and macro evidence on schooling outcomes remains a challenge for future research.

[15]Excellent discussion of the macroeconomic puzzles involving schooling and economic growth can be found in Lant Pritchett, "Where Has All the Education Gone?" *World Bank Economic Review* 15, no. 3 (2001) and "Does Learning to Add Up Add Up? The Returns to Schooling in Aggregate Data," *Handbook on the Economics of Education* (Boston: Elsevier, 2006). A less-technical review of some of these points can be found in William Easterly, "Educated for What?" in *The Elusive Quest for Growth* (Cambridge, MA: MIT Press, 2001).

MAKING SCHOOLING MORE PRODUCTIVE

Despite debate and puzzles over the rate of return to schooling, no one would conclude that parents or governments in developing nations should invest less in schooling. Schooling has benefits that go well beyond narrow economic returns. It often is considered a merit good, a good that a society determines all members should have access to regardless of ability or willingness to pay. There is also evidence that schooling improves health, with educated mothers having fewer and healthier children. The challenge facing policy makers and all those concerned with promoting economic development is to understand how to make schooling a better investment: better for students and their families who devote so much of their time to education, and better for governments and donors who finance much of the direct costs.

Rates of return on schooling depend on what happens both in schools and in the labor market after students graduate. Much of this textbook is devoted to the latter, to understanding how to increase the rate of economic growth in an economy and with it an increase in the demand for labor. We will not go over these elements again here. But it is worth noting that low returns to education often have a lot to do with failures to increase the demand for labor. High unemployment among school leavers, including graduates of universities, often reflects failures in promoting economic growth rather than failures in schools. This has been true in transition economies but also is apparent in countries such as Argentina and Egypt, which have not made good use of the human capital they accumulated. High dropout rates, as well as low and even declining school enrollment rates (the latter occurred throughout sub-Saharan Africa during the 1990s), may have as much to do with low returns in the labor market as with a family's resource constraints or problems with the schools themselves. Students and their families may decide that the benefits of schooling do not justify the costs. And in many instances they may be right. The provision of schooling is no guarantee that getting an education will yield a high return. Problems with labor demand highlight that schooling is not a panacea; it is not *the one solution* to problems of poverty and economic backwardness.

But problems on the demand side do not imply that the supply of education is without flaws. The remainder of this chapter focuses on the supply of education in an attempt to understand how investments in schooling can be made more productive. In many countries schooling is not doing as much as it can to promote development. Some of the reasons for this are that there is underinvestment in schooling overall, that governments misallocate resources across different levels of schooling, and that there are systematic inefficiencies in the use of resources within schools.[16]

[16]Two early reviews of resource use and schooling are World Bank, *Financing Education in Developing Countries* (Washington, DC: World Bank, 1986) and *Priorities and Strategies for Education* (Washington, DC: World Bank, 1995). Recent surveys of improving school outcomes in developing nations are Glewwe and Kremer, "Schools, Teachers, and Education Outcomes" and Pritchett, "Access to Education."

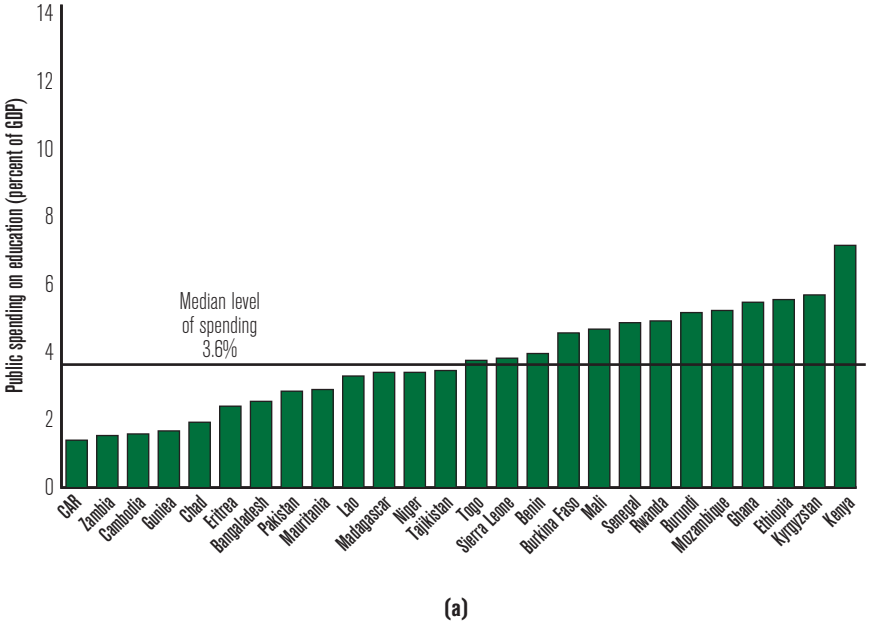
UNDERINVESTMENT

Despite attractive economic returns and other benefits to schooling, many developing countries spend too little on educating their children. Determining the right amount that should be spent is no simple matter. First, estimates of rates of return are imprecise at best; second, spending more money on schooling is no guarantee that the money will be well spent and produce better education outcomes. But money does matter, and cross-country data reveal considerable variance in public spending on schooling (comparable data on private spending are not available). In Figure 8–7, among low- and lower-middle-income nations, the range in public expenditures on education as a share of GDP is from about 1.5 to over 12 percent. We cannot be sure that the average spending share within these income groups is the "right" amount, especially because we do not know the corresponding amount of private spending. But nations such as Cambodia, Chad, and Zambia, where public spending on schooling is under 2 percent of GDP (well below the 3.6 percent median of the low-income group) stand as candidates for increasing the amount their governments spend on education. Similarly, Angola, Georgia, and Peru devote well below the lower-middle-income median of 4.3 percent of GDP to schooling and may be seriously underinvesting in education. To the extent that any of these and other countries also have a high dependency burden of young people, they especially need to devote a larger share of GDP to schooling because those expenditures must be spread among a large number of students.

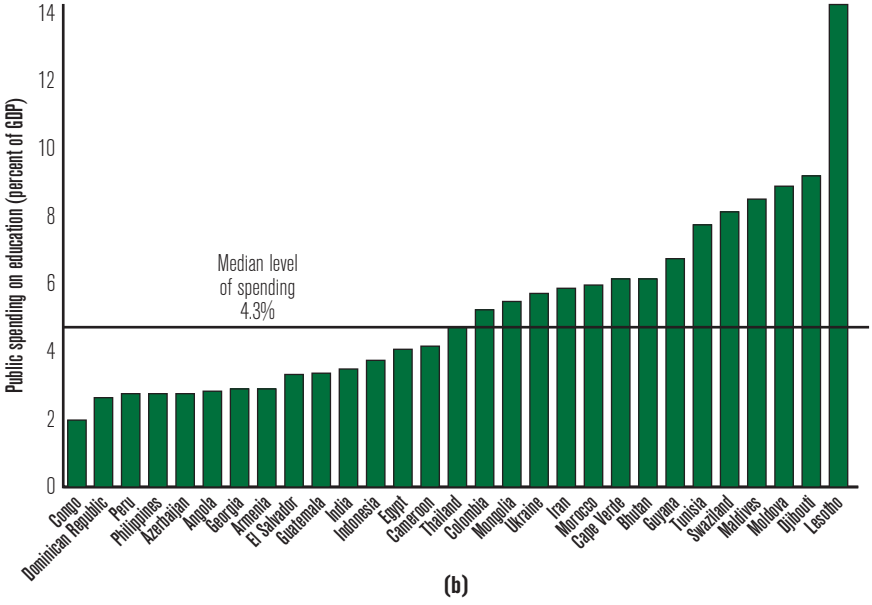
There are many explanations for why nations may underinvest in schooling. One is as a response to fiscal crises. Education and the social sectors in general, as opposed to the military or debt service, often are victims of budget cuts. Economic downturns and negative growth rates have impacts on education spending and can result in an entire cohort of children getting less schooling than would have been the case in a more stable economy. Protecting schooling and the other social sectors is a major challenge nations face whenever government revenues start to fall.

MISALLOCATION

Misallocation of resources between different levels of schooling refers to government decisions on how much to spend on primary, secondary, and tertiary levels. The estimates of rates of return to schooling in Table 8–3 indicate that the social returns are highest for primary school followed by secondary then tertiary schooling. Given the construction of these estimates, the results are not surprising. The direct costs of tertiary schooling (everything from junior colleges through doctoral programs) per student, including expenditures on computers, laboratories, libraries, professors, and the like, are much higher than the per student costs of either primary or secondary education. Because the government often pays for all or most of these costs, social returns to higher education are going to be significantly less than are private returns.



(a)



(b)

FIGURE 8–7 Public Spending on Education, 2005–07

(a) Lower-income economies. (b) Lower-middle-income economies.

GDP, gross domestic product.

Source: World Bank, "World Development Indicators Online," http://databank.worldbank.org.

This, in turn, suggests that governments should initially allocate more resources to primary schools. Once universal primary education is achieved, more resources should be devoted to secondary schools. According to this logic, tertiary education should have the lowest priority.[17]

Such conclusions encounter much opposition. First, the estimates in Table 8–3 do not take account of any (difficult to measure) positive externalities associated with different levels of education. If a nation has a chronic shortage of trained professionals in agronomy, engineering, finance, medicine, and other fields, developing such skills may yield a large social return.[18] Second, the argument that more resources should be devoted to lower rather than higher levels of schooling is reminiscent of the colonial era, when education of native populations was suppressed, leaving many nations with a minuscule number of educated citizens at the time of independence. Rich nations did not wait to achieve universal primary education before expanding their universities; why should poor nations?

While there is a legitimate debate over whether scarce resources should be allocated to one level of schooling versus another, evidence on the potential misallocation of public resources among different levels of schooling is suggested by comparing enrollment rates and expenditure patterns on tertiary education. Among sub-Saharan African countries for which the relevant data are available, the gross enrollment rate for tertiary education is 5.3 percent, while the median share of total government education expenditures devoted to the tertiary sector is 16 percent. In Latin America, tertiary enrollment rates are almost six times bigger, but the expenditure share is the same as in sub-Saharan Africa. In the high-income economies, the gross enrollment rate is over 60 percent, but the median public expenditure share is only 24 percent. United Nations Educational, Scientific, and Cultural Organization (UNESCO) computes a related and telling statistic: public expenditures per tertiary student as a percentage of gross national income (GNI) per capita. In 2007 this ratio was 206 percent in Senegal, 39 percent in Mexico, and 30 percent in France.

Gaps between enrollments and expenditures suggest that a closer look needs to be given to how the money is being spent. Much higher expenditures per student enrolled in tertiary institutions in some countries stem from generous allowances given to those enrolled in university; in essence, paying students to obtain

[17]The theoretical rule that governments should follow is really more subtle. Governments should allocate resources to equate the marginal rate of return on its investment. Internal rates of return refer to average returns (the average for all graduates of primary, secondary, or tertiary schooling) not the marginal return, which refers to the economic rate of return for the last student educated. When a particular level of schooling gets close to universal enrollment and completion, it is likely that the marginal social return will be below the average return, perhaps, well below it.

[18]Arguments in favor of for devoting more resources to higher education in developing nations is presented in World Bank, *Higher Education in Developing Countries: Perils and Promise* (Washington, DC: World Bank, 2000).

an education that is likely to yield them considerably higher earnings in the future. Noneducational expenditures in the form of student allowances and scholarships as well as subsidized housing, healthcare, loans, and so on, accounted for 50 percent of public expenditures on tertiary education among Francophone African countries, 20 percent in Latin America, and only 14 percent in OECD countries. Such allowances are often given in addition to free tuition.[19] With growing numbers of tertiary school students, such expenditures add further stress to tight education budgets. Given the tendency for those of higher socioeconomic status to have the opportunity to attend tertiary school, allowances and free tuition are also regressive uses of public money. A review of enrollment in public and private tertiary schools in Latin America found that 75 percent of the students in Colombia were from the top two quintiles of the income distribution; in Mexico, it was 83 percent, and in Brazil, 92 percent.[20] Many of these families have resources that could be devoted to higher education. When the state finances so much of the education of their sons and daughters, the outcome is a regressive one.

One approach to correcting the misallocation of resources across schooling levels is to have university students pay for some of their education. This can be justified because higher education is likely to have a high private rate of return, even if the student pays (including by taking out loans) for some of his or her education. The use of tuition reduces public contributions and permits some shifting of state spending from tertiary to primary and secondary schooling. The economic logic of charging students, especially those with the ability to pay, for some of the cost of their tertiary schooling may be clear, but instituting such fees may be far from straightforward. In Mexico, the constitution calls for free public education at all levels. Confronted with rising numbers of university students and the rising costs of providing that education, the National Autonomous University of Mexico (UNAM) in 1999 decided to raise student fees from the level last set in 1948. By 1999, these fees, adjusted for price inflation, amounted to only 2 cents ($US) per year. The administration of UNAM proposed new fees of about $150 for those with the economic means to cover these costs. Such fees would have offset less than 10 percent of the government's expenditures on the university. In response to the proposal, many of the school's 270,000 students went on strike and the school remained closed for almost a full year. In the end, the school modified its proposal, making the $150 fee voluntary, and the university's president resigned. Despite the economic logic behind the user fees on the grounds of both economic efficiency and social equity, the politics surrounding the proposal made it unworkable.

[19]World Bank, *Constructing Knowledge Societies: New Challenges for Tertiary Education* (Washington, DC: World Bank, 2002), pp. 51–52.

[20]World Bank, *Constructing Knowledge Societies,* p. 195.

IMPROVING SCHOOLS

In addition to spending more on schooling overall and reallocating some expenditures between levels, the use of resources *within* schools needs to be improved. Parents complain about schools in almost every community in the world, from suburban Boston to rural India. Some of these complaints are the same, "My child is not getting enough attention" or "My child is not learning enough." But some are quite different, "Why are teachers absent so much of the time?" or "Why are there no books for students to read?" In some instances the solutions to these problems require more financial resources; in others, they require spending available resources better; and in still others, the problems are less about resources than about holding teachers and principals more accountable for their actions.

Inefficient use of resources and lack of accountability within schools are not new issues. But recent research has given new insight into these long-standing concerns. Especially important have been the use of **randomized controlled trials (RCTs)**, also referred to as randomized evaluations. Research organizations like the Abdul Latif Jameel Poverty Action Lab (J-PAL) at MIT are pioneering the use of experiments in the field to learn more about what works in schooling, health, microcredit, and other areas. Unlike the natural experiments discussed earlier, RCTs are *actual* experiments that researchers design to determine if a particular intervention shows promise. Similar to natural experiments and to studies in medicine, one group receives a treatment and outcomes are compared to a control group that does not receive a treatment. Experimental design focuses on making sure the treatment and control groups, whether individuals, households, or entire villages, are the same in order to isolate the impact of the treatment on later outcomes.

A well-designed RCT can help resolve problems of causality that confront other types of analysis. In a conventional cross-section regression analysis of reading and mathematic achievement in Ghanaian middle schools, one of the most statistically significant variables in raising student test scores was repairing leaking roofs. Maybe this was because by repairing the leaks, school remained open and students attended school more often, which led to greater learning. Or maybe the underlying relationship was that more motivated teachers, principals, and parents were more likely to keep the building in good order. If the latter explains the statistical finding, then a nationwide initiative to repair leaky school roofs would not yield the hoped-for improvement in children's learning.[21] An RCT that experimented, for example, with building maintenance might resolve this question. Insights obtained from RCTs and other types of research offer suggestions on how to improve the participation and retention of students as well as how to enhance student learning.

[21]Glewwe and Kremer, "Schools, Teachers, and Education Outcomes."

REDUCING THE COSTS OF GOING TO SCHOOL

Earlier we noted that the MDG of achieving universal primary education by 2015 is not likely to be met. Data on grade survival rates suggest that 20 percent of children in low-income nations do not complete even four years of schooling. One reason why children may not be in school is because there is no conveniently located school to attend. This certainly can be a problem in remote regions, and many children continue to travel long distances to get to school. However, research findings suggest lack of attendance is much more often a problem of household demand than of school supply.[22]

On the demand side, households may not send their children to school, including primary school, because poor families cannot afford the fees required for school attendance. These can be user charges to attend school or fees to pay for textbooks or mandatory school uniforms. A school uniform in Kenya costs $6, a nontrivial amount in a nation where one fifth of the population lives below the $1.25-a-day poverty line. Starting in the late 1990s, governments in several East African nations (including Kenya, Tanzania, Uganda, and Zambia) eliminated all user fees for primary schools; in some cases, school uniforms were no longer compulsory. Governments subsequently reported substantial increases in enrollment.

It is not surprising that eliminating school fees would lead to an increase in enrollment, but it comes at a cost. Many families previously had paid the fees and sent their children to school. By eliminating fees, the government lost much-needed revenue to pay for teachers and other school needs. Financing schooling now requires new revenues or reallocating funds from some other use. The critical economic question is whether reducing school fees represents an efficient use of resources. Michael Kremer, a development economist at Harvard University and a leading proponent of the use of RCTs and other experimental approaches, evaluated one program in Kenya where 7 out of 14 poorly performing schools were randomly selected. The treatment schools provided free uniforms and textbooks to their students, paid for by a Dutch NGO, and also received improved classrooms. Dropout rates fell considerably, and many students from nearby schools transferred into the treatment group, raising class size by 50 percent. Kremer and his co-authors concluded that the financial benefit of free uniforms was a significant factor in improving school attendance. (By keeping girls in school longer, free uniforms were also associated with a more than 10 percent reduction in teen childbearing.) The researchers concluded that, because parents accepted larger class sizes in return for these benefits, the government could have reallocated education spending, trading off larger classes for lower fees, and thereby increasing school completion rates.[23]

Another approach to increasing enrollments and school retention is to design means-tested and targeted programs in which families are paid to send their children

[22]Pritchett, "Access to Education."

[23]Glewwe and Kremer, "Schools, Teachers, and Education Outcomes."

to school. Several countries have adopted this approach, referred to as **conditional cash transfers (CCTs)**. Bolsa Familia (family allowance) in Brazil and Oportunidades in Mexico provide modest stipends to millions of poor families as long as their children remain in school. Both have been associated with gains in enrollment and retention (Box 8–3). So has a program in Cambodia financed by the Japan Fund for Poverty Reduction. The 45 neediest girls in each of 93 participating schools were awarded a $45 scholarship, a large sum in this low-income nation, as long as the student enrolled in school, maintained a passing grade, and had a good record of attendance. Evaluation of the program reports large increases in desired outcome measures. The Food for Schooling program in Bangladesh provides a free monthly ration of wheat or other grains to eligible families, where eligibility is means tested. To qualify, households must own less than half an acre of land and the household head must be a woman, a day laborer, or work in a very low-wage activity. The family can consume the grain or sell it and use the cash for other expenses. To maintain eligibility, children must attend 85 percent of classes each month. The program has been successful in increasing primary school enrollment, promoting school attendance, and reducing dropout rates, especially among girls.

BOX 8–3 MEXICO'S PROGRAMA DE EDUCACIÓN, SALUD Y ALIMENTACIÓN (PROGRESA)/OPORTUNIDADES

When Ernesto Zedillo became Mexico's president in 1995, a fifth of the population could not afford the minimum daily nutritional requirements, 10 million Mexicans lacked access to basic health services, more than 1.5 million children were out of school, and student absenteeism and school desertions were three times higher in poor and remote areas than in the rest of the country. His administration decided that a new approach to poverty alleviation was needed. The Education, Health, and Nutrition Programs of Mexico, called PROGRESA, introduced a set of *conditional cash transfers* to poor families—the families would receive cash if their children were enrolled in school and if family members visited health clinics for checkups and nutritional and hygiene information.

The program was intended to remedy several shortcomings of earlier programs. First, it would counter the bias in poor families toward consumption by bolstering investment in human capital. Second, it would recognize the interdependencies among education, health, and nutrition. Third, it would link cash transfers to household behavior, aiming at changing attitudes. Fourth, to reduce political interference, the program's goals, rules, requirements, and evaluation methods would be widely publicized.

Children over the age of seven were eligible for education transfers. Benefits increased by grade (because the opportunity costs of being in school increase with age) and were higher for girls in middle school, to encourage their enrollment. To retain the benefits, children needed to maintain an 85 percent attendance record and could not repeat a grade more than once. Transfers went to mothers, who were thought to be more responsible for caring for children. In 1999, the average monthly transfer was around $24 per family, nearly 20 percent of mean household consumption before the program, with benefits available for up to three years.

By the end of 2002, the program had about 21 million beneficiaries, roughly a fifth of the Mexican population. Almost 60 percent of program transfers went to households in the poorest 20 percent of the national income distribution and more than 80 percent went to the poorest 40 percent.

School enrollment of participating households rose, especially in middle school and more so for girls than boys. Much of the increase came from increases in the transition from primary to middle school. The program worked principally by keeping children in school, not by encouraging those who had dropped out to return. Grade repetition decreased among participating families even for children in grades one and two who were not eligible for program benefits. Perhaps, this was because household income was higher or because older siblings were now spending more time in school. The impact of the program on learning is less clear. Teachers reported improvements, attributing them to better attendance, student interest, and nutrition. But a study conducted one year after the program started found no difference in test scores. Follow-up studies on longer-term effects are planned.

In 2001 and under a new Mexican president, PROGRESA was renamed, *Oportunidades*. The success of the program is reflected not only in schooling and health outcome measures but in its ability to survive changes in political leadership. It also has become a model for similar programs throughout Latin America and the rest of the world.

Source: Adapted from World Bank, "Spotlight on PROGRESA," 2004 World Development Report, "Making Services Work for Poor People" (Washington, DC: World Bank and Oxford University Press, 2004), 30–31. (World Bank, 2009).

CCTs like the ones just described all have multiple aims and can be seen as part of a package to alleviate poverty. If such transfers of cash or grain were going to be made on poverty grounds alone, then making them conditional on sending children to school makes the education component highly cost-effective. If they are intended primarily as interventions to expand enrollment and school retention,

cost-effectiveness depends, in part, on how carefully the targeting is done. If eligibility is too broad, recipients will include families that already were sending their children to school and the cash or food expenditures will yield no marginal improvements in schooling for these households. Despite such concerns over cost-effectiveness, the success of CCTs such as Bolsa Familia and Oportunidades in raising school enrollments has resulted in their adoption in dozens of nations.

INEFFICIENT USE OF RESOURCES

Improving the efficient use of resources within schools also involves making decisions about spending. Should scarce resources be allocated to buildings or teachers, to teachers or school supplies (blackboards, desks, textbooks, etc.), or to any of a number of other alternatives? The problems that schools in poor nations face can be overwhelming. There usually are shortages of everything. There are too few textbooks and sometimes there are none. Roofs leak, and many students are crammed around simple desks if there are desks at all. Teachers often are untrained and children often too sick to learn effectively. Deciding on how to allocate resources in such environments is not easy.

In the 1980s, some of the debate over resource allocation within schools concerned building schools versus expenditures on recurrent costs to run schools. These recurrent costs included teacher and administrative salaries as well as nonwage expenditures on school supplies and equipment. Donors, who often financed much of the spending on schools, especially in low-income nations, had a preference for constructing new school buildings. The donors could point to the tangible product of their aid, something constituents back home might want to see. Donors also expected a partnership with those governments who received aid and assumed that governments would cover the recurrent costs. The result of these arrangements often was the construction of schools, often at higher standards than necessary and sometimes requiring imported building materials, met by an underfunding of recurrent costs. Schools without teachers or classrooms without chalk or writing paper often resulted.

By the 1990s, the approach had changed. Concern over the quantity and availability of schools diminished, and increasing attention was paid to improving the quality of schools. Resources were redirected toward recurrent expenditures, especially nonwage items. But this approach also proved wanting; improvements in learning outcomes remained elusive.

One cost-effective intervention for increasing school attendance (in addition to lowering fees) is not an educational input; it is biannual dosages of albendazole or praziquantel, two medications used to reduce intestinal worms such as hookworms or schistosomes. Such intestinal worms affect a quarter of the world's population and are especially common among children. Severe infections can cause anemia,

protein-energy malnutrition and other health problems. On the basis of an RCT, students in some schools in rural Kenya, where worm infestation was particularly severe, received treatments and others did not. Researchers found that school absenteeism decreased by 25 percent in treatment versus control schools. Some children who previously were weak or listless were able to attend school because of the drugs; in other cases, deworming improved concentration and may have made attending more worthwhile. Beyond their effect on school attendance, these drugs also improved health outcomes.

Deworming proved highly cost-effective because the treatment cost is low, about $0.50 per child per year. Deworming increased school attendance by an average of 0.15 year per pupil or $3.50 per an additional full year of school participation. By comparison, school uniforms cost $6 per child. Because provision of school uniforms increased school attendance at a somewhat lower rate than deworming, the school uniform program was less cost-effective at $99 per additional year of school participation. Even though deworming was associated with improved school attendance and was cost-effective in doing so, there was no discernible improvement in learning as measured by student test scores at the time of treatment.[24]

New findings from the original deworming RCT reveal its long-term benefits. A follow-up study conducted in 2007–09 located a large percentage of the original Kenyan school children, some 6 to 11 years after the experiment began. By then their median age was 22 years. Comparing those who attended schools where the treatment was provided with those where it was not reveals significant differences. Deworming is associated with more subsequent years enrolled in school, better self-reported adult health, more hours spent working, higher earnings among those employed in wage-paying jobs, and improvements in adult food consumption.[25] Consistent with a growing literature, these investments in the childhood years had significant pay-offs when these children reach adulthood because of improved childhood health and schooling.

More traditional school input issues involve reducing class size, increasing the availability of textbooks and other instructional materials, and improving teacher training. Appropriate class size is a matter of debate in school systems throughout the world, in rich and poor nations alike. The basic argument is straightforward. In smaller classes, teachers can provide more attention, even individualized work, for each student. Running a 25-pupil class of 7-year-olds (let alone 13-year-olds) is daunting enough, but what must it be like to teach a class two, three, or more times as large?

[24]Edward Miguel and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica* 72, no. 1 (2004).

[25]Michael Kremer, Sarah Baird, Joan Hamory Hicks, and Edward Miguel, "Worms at Work: Long-Run Impacts of Child Health Gains" (May 2011), www.economics.harvard.edu/faculty/kremer/files/KLPS-Labor_2011-05-13-Circulate-B-No-IRR.pdf.

Eric Hanushek, a senior fellow at the Hoover Institution and a leading expert on the economics of education in the United States, surveyed almost 100 studies on the determinants of student test scores in developing countries. Of the 30 studies that looked at class size, 8 found a statistically significant positive effect (the fewer students per teacher, the better students performed); 8 found a statistically significant negative effect (the fewer students per teacher, the *worse* students performed); and 14 found no significant effect at all.[26] We need to be careful interpreting these findings. They do not suggest that class size never matters; the results say only that there is no evidence of a robust and significant correlation (let alone causation) between class size and student performance.[27] In some instances, reducing class sizes is warranted; in others, the extra expenditure on teachers and additional classrooms yields no return. Without being able even to identify a simple correlation between reduced class size and student achievement, it is that much harder to move to the question of cost-effectiveness or economic efficiency of allocating scarce resources so that outcomes per dollar are equalized across alternative inputs.

The failure to easily identify strategies to improve student outcomes is not restricted to studies of student–teacher ratios. Studies of teacher training or textbook use yield a similar range of outcomes. According to the survey by economists Paul Glewwe and Michael Kremer, the introduction of textbooks in Jamaican primary schools improved reading scores, while in the Philippines the impact of textbooks was "unstable," including positive *and negative* effects on the mathematics and reading scores of first graders. An RCT in rural Kenya found no evidence that provision of textbooks improved scores for the average student. However, Kenyan students who were above average before textbooks were provided realized improvements. This may be because textbooks provided by the government were in English, often the second or even third language of many local children, and only the most able students had sufficient English language skills to benefit from the textbooks they received. What the textbook studies suggest is that the central problem may not be an absence of learning materials but systemic problems in the education system. Textbooks that help only the best students highlight problems with a centralized national

[26]Eric Hanushek, "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer* 10, no. 2 (1995). In an earlier study, Hanushek reached the same conclusion about the impact of class size on student performance in the United States ("The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24 [1986]).

[27]One of the difficulties of identifying the impact of class size on student learning is the problem of *endogeneity*. Class size may not be determined independently of other determinants of student performance. For example, if school administrators decide to keep class sizes small for students who are disruptive or have difficulty learning in the hope that teachers can manage such classes better if they have fewer students, then class size will be inversely correlated with student performance but not a cause of low performance. Similarly, if better-educated and higher-income communities effectively lobby for more teachers and smaller classes, class size will be positively correlated with better student performance even though better performance may mostly be due to having students with better-educated parents enrolled in smaller classes. Once again the causal effect of class size on performance would be hard to identify.

curriculum oriented to educating the children of the politically influential elite and failing to recognize the heterogeneity among students.

Most schools in developing nations can do better, and RCTs are providing useful insights, including the benefits of some technology-assisted approaches and linking teacher pay to student test scores.[28] One single intervention, however, will not work all the time. Local conditions make a difference, such as the accountability of teachers, principals, and other school officials.

IT IS ABOUT MORE THAN THE MONEY

If there has been a trend in thinking about education and development, it has gone from concern primarily over increasing the quantity of schooling, achieved by building more schools, hiring more teachers, and most recently, reducing the cost of attending to improving the quality of schools by improving the amounts and mix of inputs within schools themselves. Both these approaches have had some measure of success, with the huge increases in enrollments (discussed earlier) and success in learning outcomes, at least in some settings, due to different or better use of school inputs. But widespread and persistently low student achievement has resulted in a call not only for more money but also for reform of school systems. School reform and more resources need not be substitutes. Both may be required to realize better outcomes, but more money without school reforms will be insufficient to achieve these ends. Calls for school reforms echo those for institutional reforms in other areas of government. Improvements in governance are needed not only to improve the climate for private enterprises and encourage investment in physical capital but to improve investment in human capital.

One critical area of school reforms is to better motivate teachers, many of whom fail to show up for work and to engage in teaching when they do. With weak incentives, teacher motivation often is low. The need for improved accountability among teachers is starkly demonstrated by a nationwide study of teacher absence in India.[29] Researchers made three unannounced visits to 3,700 Indian primary schools. The presence or absence of teachers assigned to each school was determined by *direct physical verification*—that is, by a member of the research project looking for the teacher in the school building and not by checking logbooks or other records. On average, one in four primary school teachers was absent on the day of a visit. And only 45 percent (less than half) of primary school teachers were actively engaged in teaching their students at the time of the visit. Some percentage of these absences

[28]The use of RCTs has increased dramatically in recent years. J-PAL records many of the latest findings on its website at www.povertyactionlab.org/. Also, see Michael Kremer and Alaka Holla, "Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?" *Annual Review of Economics* 1 (September 2009).

[29]Michael Kremer et al., "Teacher Absence in India: A Snapshot," *Journal of the European Economic Association* 3, nos. 2–3 (April–May 2005).

and involvement in nonteaching matters was due to excused absences for health or other reasons or performing required administrative or other duties. But even after correcting for these factors the degree of absence remains high.

The authors of the India study looked at the correlates of teacher absence and did not find that relatively higher pay or more education among teachers reduces absences. Stronger ties to the local community do not seem to play much of a role either. Schools with better infrastructure (for example, with electricity or staff toilets), more frequent inspections, and closer proximity to roads have somewhat lowered absenteeism, but the overwhelming conclusion is that getting more teachers to do their jobs is a significant challenge. The research includes similar investigations in other countries. While India's teacher absence rate is on the high end (Uganda is higher at 27 percent), rates of 14 percent in Peru, 15 percent in Papua New Guinea, and 19 percent in Indonesia are not encouraging. Discussions of appropriate curriculums, better teaching methods, or more school resources almost seem secondary if so many teachers are not minimally engaged in the daily routines of educating their pupils. In explaining why teacher absence is so high, Michael Kremer and his co-authors note that teachers face little risk of being fired. They found that only one head teacher (school principal) in nearly 3,000 primary schools in India had ever dismissed a teacher for repeated absences. (An innovative solution to reducing teacher absence in some schools in India is discussed in Box 8–4.) The absence of repercussions is one explanation for high absence rates but the problem of accountability runs far deeper.

BOX 8-4 COMBATING *TEACHER* ABSENCE

When a student misses a class his or her learning suffers; when the teacher is absent the whole class suffers. Or does it? In an attempt to reduce teacher absence and assess its impact, Esther Duflo and colleagues collaborated with an Indian nongovernment organization (NGO), Seva Mandir, on an innovative policy experiment.

Seva Mandir runs single-teacher primary schools in tribal villages in a sparsely populated and remote rural area. Tens of millions of Indian children attend schools like these. Their goal is to teach basic skills to prepare students for the regular government-run school system. Teachers are employed by the NGO on flexible contracts and are not civil servants, as are most Indian schoolteachers. This made it easier to implement a program that provided monetary incentives to reduce teacher absences. Because of their remote location, it is difficult to monitor such schools. Teacher absence was at the alarming rate of 44 percent. This was despite Seva Mandir's threat to dismiss teachers with too many absences, although that threat rarely was carried out.

Seva Mandir selected 120 of its schools to participate in a randomized control trial (RCT). These one-room schools usually have about 20 students and run a six-hour school day. The teacher is drawn from the local community and has about 10 years of education. There is only one teacher per school, and when the teacher is absent the school is closed. Half of the schools, the treatment group, were given a camera with a tamper-proof date and time function. Each day, teachers were instructed to have a child in the school take a picture of the teacher and the other children at the beginning and end of the school day. A *valid school day* was defined as one in which at least five hours elapsed between the two pictures and when at least eight students were present in each picture. Teacher pay was a function of the number of valid days they worked based on the photographs submitted. The nontreatment group was paid their regular monthly salary. They continued to be told that they could be dismissed for poor performance, including too many unexcused absences. There was one unannounced visit to control schools each month. Average salaries ended up almost identical in the two groups.

Over an 18-month period, the absence rate in the treatment schools fell by half, to 22 percent. The use of cameras to monitor teacher presence completely eliminated extremely delinquent behavior—absences of over 50 percent. It also increased the number of teachers with perfect or very high attendance from under 1 percent in the control group to 36 percent in the treatment group.

Did students learn more? The incentives built into the program encouraged teachers only to be present. It did not reward spending more time teaching or improving student performance, the true goal of the intervention. Based on observations during the unannounced visits, teachers in treatment and control groups were equally likely to be engaged in teaching. Because the treatment group was present far more often, more teaching was taking place, equivalent to almost 3 days per month. After one year, this resulted in a 0.17 standard deviation increase in children's test scores in basic Hindi and math skills, and a significantly higher matriculation rate of children in the treatment group to government schools. These benefits were highly cost-effective. Outside of salaries, which were the same in the two groups, all other costs, including cameras, film, and administrative expenses, amounted to roughly $6 per child or to $3.58 per 0.10 increase in the standard deviation in test scores.

Seva Mandir continued to monitor attendance with cameras long after the RCT ended and improved attendance has been maintained. The effects appear to be long lasting. The success of this monitoring system, however, may not be applicable to government schools in India, in part, because of resistance from

teacher unions. But the potential of incentive schemes that reward teacher performance remains.

Source: Esther Duo, Rema Hanna, and Stephen P. Ryan, "Incentives Work: Getting Teachers to Come to School," *American Economic Review* (forthcoming).

Teacher absence or, for that matter, the absence of health workers or any number of other government workers is not the only example of the breakdown in accountability that compromises the delivery of public services in low- and middle-income-nations.[30] In another often-cited study, researchers followed the money allocated by Uganda's national government for local primary schools.[31] Despite the government authorizing 20 percent of public expenditures to education, most of it to primary education, a program to finance nonwage expenditures rarely resulted in money reaching the schools it was intended for. Over the period 1991–95, an average of only 13 percent of such grants actually reached the schools; almost three-quarters of the 250 primary schools surveyed in the study received less than 5 percent of what had been authorized. Local officials and politicians captured much of the money along the way.

The identification of the massive leakage of funds in Uganda led to a number of institutional improvements. The central government responded by launching an information campaign, where the amounts of centrally disbursed grants were reported in newspapers. Primary schools were also required to post notices of grants actually received. Local parent teacher associations (PTAs), which are very powerful in the Ugandan school system, thus were made aware of grants due their schools and could monitor outcomes. By reducing information asymmetries, it was possible to hold local officials more accountable. The information campaign, coupled with more attention from central government officials (including stiff penalties against offenders), increased the receipt of funds by schools from 20 percent in 1995 to 80 percent in 2001.[32]

The Ugandan example ultimately relied on local or community control as a means of increasing accountability, this time of local officials who received funds

[30]The problem of improving service delivery by the public sector is the focus of the *2004 World Development Report, Making Services Work for Poor People* (Washington, DC: World Bank and Oxford University Press, 2004). Specific chapters are devoted to basic education, health and nutrition, and drinking water, sanitation, and electricity.

[31]Ritva Reinikka and Jakob Svensson, "Local Capture: Evidence from a Central Government Transfer Program in Uganda," *Quarterly Journal of Economics* 119, no. 2 (May 2004).

[32]Ritva Reinikka and Jakob Svensson, "The Power of Information: Evidence from a Newspaper Campaign to Reduce Capture," Policy Research Paper 3239 (Washington, DC, World Bank, March 2004). See also Paul Hubbard, "Putting the Power of Transparency in Context: Information's Role in Reducing Corruption in Uganda's Education Sector," Working Paper Number 136 (Washington, DC, Center for Global Development, December 2007), which argues that the improved flow of resources to intended beneficiaries was the product of many reforms in government practice, including but not solely the result of more transparent information.

from the central government. Expanding local control, often through decentralization of authority over schools, is one of several reform strategies to hold all providers—officials who control funds as well teachers and principals—more accountable. The Educo program (Educación con Participación de la Comunidad) in El Salvador provides another example of this approach.

A prolonged civil war in the 1980s left the Salvadorian school system in a woeful state. One-third of the country's primary schools were closed, and in many regions, traditional government schools, and their teachers, were viewed with distrust by local people who may have been fighting against the government. During the war, some communities had recruited local teachers and established community schools in place of government-run ones. Seizing on this model, the government decided to incorporate community schools into the national system and encourage new ones to open. These *popular* schools enter into one-year renewable agreements with the ministry of education. The local schools are run by parent groups, which receive grants from the government, hire teachers on renewable contracts, and determine the salaries and other terms of these contracts. Turnover rates are high; teachers who are frequently absent or perform poorly are dismissed. Parents visit schools more often and are more involved than in traditional schools. The results of these arrangements are encouraging. Educo reached far into the countryside. Rural primary school enrollments grew rapidly with most new students enrolled in the *popular* schools. Student performance in Educo and traditional schools was comparable. This was an impressive outcome because Educo students tended to come from poorer and less-educated families than students who attended the traditional government schools.[33]

The relative success of Educo does not mean that all school systems should adopt a similar model or that greater parental involvement and control always improves school outcomes. Results from some RCTs in Kenya and India suggest otherwise.[34] But all these cases speak to the need for nations to experiment with alternative modes of service delivery. School reforms are part of a larger agenda of institutional reforms that many nations, whether low, middle, or high income, are pursuing to improve schooling, health, and other outcomes. Evidence from such reforms indicates that one size does *not* fit all and continued experimentation and evaluation is needed to identify what works in different settings. School systems need to enroll and retain more students and improve learning outcomes at all levels of education. Teachers must be adequately trained and motivated. Schools need clear objectives, adequate financing, and some autonomy on how to operate and achieve stated goals. They also must be held accountable to the taxpayers (at home and abroad) who finance school systems and, most important, to the students and their parents who count on schools to provide an education.

[33]World Bank, *2004 World Development Report,* pp. 131–32.

[34]Kremer and Holla, "Improving Education in the Developing World."

SUMMARY

- By historical standards, the last three decades have witnessed a revolutionary change in the number of men and *women* who have attended school and received a basic education. By 2010, 80 percent of the developing world's adults had attended school as children. These investments in human capital have profound implications for improving the well-being of much of humanity in the twenty-first century.
- Despite these achievements, much remains to be done. More than four out of every five children in the world live in low- and middle-income nations, and far too many of them have never attended school or failed to complete even four years of primary school. And those who attend school often learn far too little, with their education lagging well behind what comparably aged children master in high-income nations.
- Attending school remains a worthwhile private and social investment. In every nation, if you have more years of schooling, on average, you earn more than those with less schooling. Estimated rates of return, which take account of both the benefits and costs of schooling, indicate that schooling at all levels yields attractive returns when compared to other investments.
- But schooling is not a panacea. Many nations have expanded enrollments and sent their young people through years of schooling and still have not grown. In a bad economic environment, schooling, like other investments, can be wasted. Increasing the demand for all workers, educated and not educated, raises the private and social return on education and remains a key challenge to achieve economic development.
- Much also needs to be done to make schools more productive. Some developing nations underinvest in education overall; many misallocate resources by spending too much on the tertiary sector at the expense of primary and secondary levels; and most could improve the use of resources within schools themselves, including reducing the frequency of absent *teachers*.
- Decreasing user fees and other costs, offering households cash transfers conditional on school attendance, and improving child health and nutrition have all proven effective in getting the children of poor families to enroll in and more regularly attend school.
- Prescriptions for how to improve learning outcomes have proven harder to identify, although use of randomized controlled trials are offering many new insights. Changing the mix of school inputs has shown limited promise; making teachers, administrators, and government officials accountable for school outcomes may be the most important factor of all.

9 Health

Grab a pencil, here comes a pop quiz. Denmark, a high-income nation, had a population of 5.4 million people. With a death rate of 10 per 1,000, about 54,000 Danes died that year. Here is your first question: What was the median age of those who died? In other words, pick the age at which half of those who died were older and half were younger than that age. Now turn to Sierra Leone, one of the poorest nations in the world, which has a population about the same size as Denmark's. But Sierra Leone's death rate was much higher, estimated at 24 per 1,000, resulting in roughly 130,000 deaths. Here is your second question: What was the median age of those who died in Sierra Leone? Write down your predictions; we will provide the answers in a moment.

The difference in the age distribution of deaths between a very poor nation like Sierra Leone and a wealthy developed country like Denmark is that most deaths occur before age 5 in Sierra Leone, whereas most deaths occur among the elderly in Denmark. This is graphically illustrated by the age pyramids in Figure 9–1. We examined age pyramids in Chapter 7, contrasting those of rapidly versus slowly growing populations (Box 7–3). The age pyramids of Figure 9–1 look entirely different. They refer only to those who died in a given year. In Sierra Leone, more deaths occur in the first 4 years of life than in all other age brackets combined; in Denmark, child deaths rarely occur. Here are the answers to the quiz: The median age of death in Denmark is 77; in Sierra Leone, it is under 4. The striking difference between these numbers tells us a lot about life and death in the richest and poorest nations of the world.

One of the best indicators of a county's overall health status is the **under-five mortality rate**. This measure is the probability (expressed as per 1,000 live births)

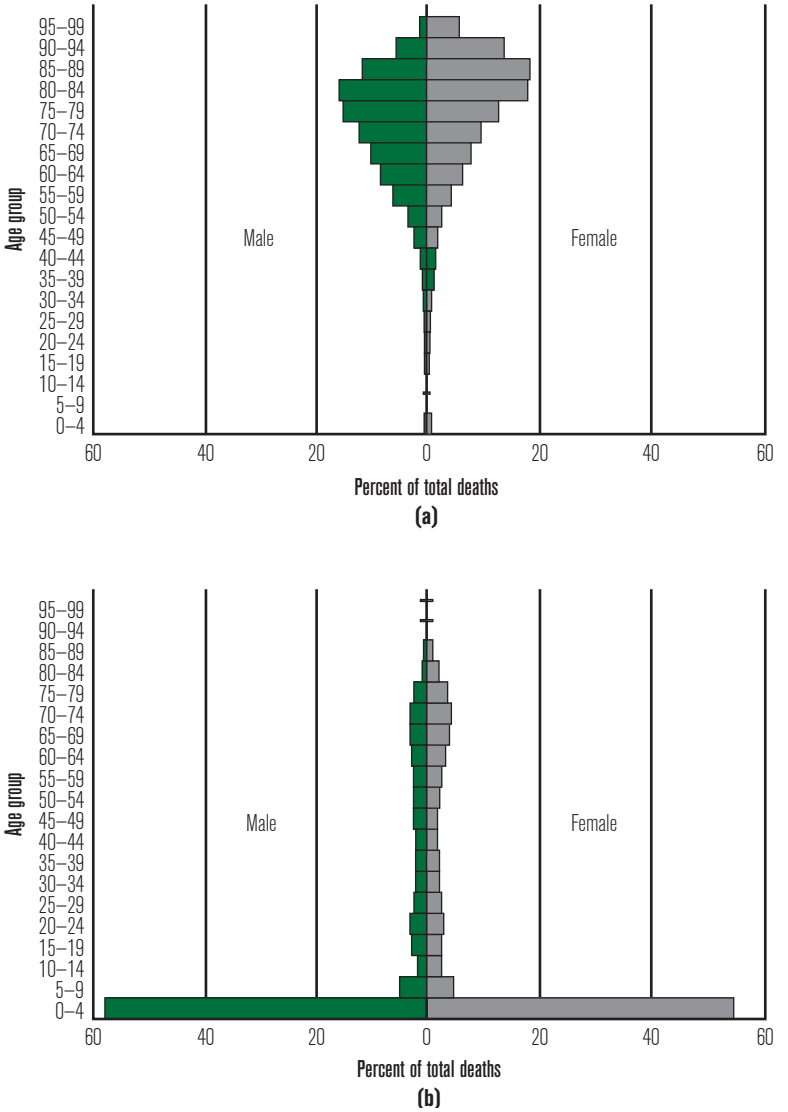


FIGURE 9–1 Distribution of Age at Death, 2005

(a) Denmark. (b) Sierra Leone. Values are projections.

Source: World Bank, *Health Financing Revisited*, processed (Washington, DC: World Bank, 2005).

that a child born in a specific year dies before reaching five years of age, if subjected to current age-specific mortality rates. In Sierra Leone, there were 194 deaths of children under-five per 1,000 live births compared to only 4 in Denmark (Table 9–1). About a quarter of all under-five mortality occurs in the first month of life in Sierra Leone. Early childhood is an especially risky period for Sierra Leone's children.

There are many explanations for the massive disparity in the chances that a child born in Sierra Leone versus one born in Denmark will survive and for the lower health status these numbers suggest for the rest of the population. At birth, less

TABLE 9-1 Selected Health-Related Measures for Sierra Leone and Denmark, Mid-2000s

	SIERRA LEONE	DENMARK
Mortality statistics		
Life expectancy at birth (years)		
Both sexes	40	79
Males	39	76
Females	42	81
Mortality rates (per 1,000)		
Neonatal (first 28 days)	56	3
Under 5 years	269	4
Adult (15–60 years) men	556	111
Adult (15–60 years) women	460	65
Healthy life expectancy at birth (HALE, years)		
Males	27	69
Females	30	71
Morbidity statistics		
Children under 5 years, stunted for age (%)	38	–
Children under 5 years, underweight for age (%)	25	–
Environmental risk factors		
Access to improved water sources, urban (%)	83	100
Access to improved water sources, rural (%)	32	100
Access to improved sanitation, urban (%)	20	100
Access to improved sanitation, rural (%)	5	100
Health services coverage		
Immunization coverage among 1-year-olds (%)		
Measles	67	89
Diphtheria, tetanus, pertussis, three doses	64	75
Births attended by skilled health personnel (%)	43	–
Health system statistics		
Number of physicians per 10,000	< 1	36
Number of nurses and midwives per 10,000	5	101
Hospital beds per 10,000	4	38
Per capita total expenditures on health (US$, PPP)	41	3,349

HALE, health-adjusted life expectancy; PPP, purchasing power parity.

Source: World Health Organization, Statistical Information System, *World Health Statistics 2009*, available at www.who.int/whosis/en, accessed February 2012.

than half of all mothers in Sierra Leone are tended to by a skilled health worker; in Denmark, such services are available to all mothers. In Sierra Leone, malnutrition is common, especially among children. One out of every three children under age five is stunted and one out of five is underweight. **Stunting** refers to the percentage of children under five years that have a height to age ratio more than two standard deviations below the **World Health Organization (WHO)** global reference median. Being **underweight** refers to the percentage of children under five years that have a weight to age ratio more than two standard deviations below the WHO global reference median. Both outcomes are rare in Denmark.

Sierra Leone and Denmark also differ significantly in terms of health services received and resources available to the health system. The majority of rural inhabitants in Sierra Leone have no access to clean water or improved sanitation, and many in the urban areas lack those services as well. Everyone in Denmark has access to safe drinking water and modern sanitation. Almost all 1-year-olds (89 percent) are immunized in Denmark for measles and receive three doses of the diphtheria, tetanus, and pertussis (DTP) or whooping cough vaccine, designated DTP3. About 60 percent of 1-year-olds in Sierra Leone are immunized for measles and for DTP3; thus 40 percent are not immunized and at risk of infection. A person who gets sick in Denmark has many more resources to turn to for medical care. Denmark has almost 50 times as many nurses and midwives per 10,000 persons as Sierra Leone (98 versus 2). There are 32 trained physicians per 10,000 people in Denmark; in Sierra Leone, there is less than 1. Denmark also spends almost 100 times as much per person per year on healthcare ($3,513 versus $32 in purchasing power parity [PPP]). Poor access to health services from childhood through adulthood increases adult mortality. In Sierra Leone, adult men (ages 15–60) are four times more likely to die than in Denmark; adult women are five times more likely.

The data for Denmark and Sierra Leone reflect the experience of their geographical regions, Europe and Africa. Africa has the highest rate of under-five mortality with 142 deaths per 1,000 compared to 14 per 1,000 for Europe, the region with the lowest rate[1] (Table 9–2). The eastern Mediterranean region, which includes 21 nations in North Africa and the Middle East, has the next highest regional under-five mortality rate at 78; and Southeast Asia, encompassing 11 nations (including Bangladesh, India, and Indonesia), is the third highest at 63. Poor health outcomes are correlated with low incomes, which in turn are associated with low levels of education and environmental risk factors, such as lack of access to safe drinking water and the absence of improved sanitation, especially in rural areas. Absence of medical personnel and low levels of health spending are other characteristics of high mortality among both children and adults. As we shall see, low incomes explain much, but certainly not all, of the disparity in health outcomes across countries and regions.

WHAT IS HEALTH?

The WHO defines *health* as a state of complete mental, physical, and social well-being and not merely the absence of disease.[2] However, such a complex construct would be difficult to measure and would likely vary between cultures and over time.

[1]WHO is the primary United Nations agency responsible for global health issues. WHO data rely on a regional breakdown that is different from the one used by the World Bank and frequently referred to throughout this textbook.

[2]Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, June 19–22, 1946. The definition has not been amended since 1948.

TABLE 9–2 Selected Health-Related Measures for World Health Organization Regions, 2009

	AFRICA	AMERICAS	SOUTHEAST ASIA	EUROPE	EASTERN MEDITERRANEAN	WESTERN PACIFIC
Mortality statistics						
Life expectancy at birth, males (years)	51	73	63	70	63	72
Life expectancy at birth, females (years)	54	78	66	78	66	77
Mortality rates (per 1,000)						
Neonatal (first 28 days)	40	11	35	10	38	17
Under 5 years	145	19	65	15	82	22
Adult (15–60 years) men	429	163	252	221	229	144
Adult (15–60 years) women	374	91	187	94	175	85
Healthy life expectancy at birth, males (HALE years)	45	65	56	64	55	65
Healthy life expectancy at birth, females (HALE years)	46	69	57	70	57	69
Morbidity statistics						
Prevalence of HIV among adults aged > 15 (per 100,000)	4,735	448	295	336	202	89
Environmental risk factors						
Access to improved water sources, urban (%)	82	98	94	100	93	98
Access to improved water sources, rural (%)	46	81	84	92	75	82
Access to improved sanitation, urban (%)	46	92	58	97	85	79
Access to improved sanitation, rural (%)	26	68	27	85	43	61
Health services coverage						
Immunization coverage among 1-year-olds (%)						
Measles	74	93	73	94	84	92
Diphtheria, tetanus, pertussis, three doses	74	93	69	96	87	92
Births attended by skilled health personnel (%)	46	92	48	96	59	92
Health system statistics						
Number of physicians per 10,000	2	19	5	32	10	14
Number of nurses and midwives per 10,000	11	49	12	79	15	20
Hospital beds per 10,000	10	24	9	63	14	33
Per capita total expenditures on health (US$, PPP)	111	2,788	85	1,719	259	461

HALE, health-adjusted life expectancy; PPP, purchasing power parity.

Source: *World Health Statistics 2009* (Geneva: World Health Organization, 2009).

The measures most often used to express health are those that describe the absence of health: mortality and morbidity statistics. **Mortality** measures deaths in a population; **morbidity** measures rates of disease and illness.

Because death is an unambiguous event indicating a complete failure of health, mortality statistics offer a summary of a population's health status and reveal much about a population's standard of living and healthcare. Most countries record and publish death rates with various levels of coverage and accuracy. In a very poor country, such as Sierra Leone, less than 25 percent of deaths were covered by the country's own vital registration system. In such cases, mortality rates are estimated from sample surveys or other existing data.

Perhaps the most common summary statistic used to provide a snapshot of a country's health status is life expectancy, which is derived from data on mortality. Because death rates are affected by factors such as age, sex, and race, life expectancy is often calculated for specific demographic subgroups. The life expectancy for a male born in 2008 in Sierra Leone was 48 years and 50 years for a female. A Danish baby boy born in 2008 had a life expectancy of 77 and a girl had a life expectancy of 81. To put these numbers in some historical perspective, life expectancy in Sierra Leone today is about what it was in the United States in 1900.

One might think that in a country where life expectancy is 49 (or 79) years, most people die by that age. That is clearly not the case, as Figure 9–1 demonstrates. Many people have shorter lives whereas others live well past the average life expectancy. Confusion arises because life expectancy is an average with variance around the mean; it is a "synthetic" statistic. It is based on the probability of surviving from one year to the next assuming that today's age-specific death rates remain unchanged into the future. But this is unlikely. Rising incomes and advances in medicine suggest that age-specific death rates are likely to fall in coming decades, expanding lifetimes beyond the predictions of current life expectancy estimates. However, new diseases, like HIV/AIDS, can cause the opposite, leading to shorter lifetimes and declining life expectancy.

Box 9–1 describes in detail how life expectancy is computed using data from Malaysia. Even though life expectancy in Malaysia was 69 years, people in the 70–75 age group had a life expectancy of 10 more years, a sort of "bonus" for surviving. In countries with high infant mortality, like Sierra Leone, life expectancy for children who survive the early years can be higher than at birth. For example, life expectancy at birth for Sierra Leone was 49 years in 2008. For the 5- to 9-year-old age group, life expectancy was 55 additional years for a total of 60 years for a 5-year-old—a full 11 years more than a newborn.[3] A 5-year-old child in Sierra Leone today will probably live even longer than 60 years. Now that years of civil war finally have ended, it is hoped that economic growth and development can begin. If this happens, age-specific death rates should start to fall, and Sierra Leone's children should tend to live beyond current estimates of life expectancy.

[3]WHO Global Health Observatory Repository, available at http://apps.who.int/ghodata; *World Health Statistics 2010,* World Health Organization, Geneva, 2010.

BOX 9-1 LIFE EXPECTANCY

Life expectancy is an estimate of the average number of additional years a person could expect to live if the age-specific death rates for a given year prevail for the rest of the person's life. Life expectancy is a hypothetical measure because it is based on current death rates, and actual death rates change (they usually fall) over the course of an individual's lifetime.

Life tables are used to calculate life expectancy; with life expectancy at birth the most commonly cited life expectancy measure. The table in this box contains selected portions of a life table for men in Malaysia in 1995. Age-specific death rates are applied to a hypothetical population of 100,000 people born in the same year. Column 1 shows the proportion of each age group dying in each age interval. These data are based on the observed mortality experience of the Malaysian population.

Column 2 shows the number of people alive at the beginning of each age interval, starting with 100,000 at birth. Each age group contains the population that survived from the immediately preceding group. Column 3 shows the number who would die within each age interval (Column 1 × Column 2 = Column 3). Column 4 shows the total number of person-years that would be lived within each age interval, including estimates of those who remain alive for only part of the interval. Column 5 shows the total number of years of life to be shared by the population in the age interval in all subsequent intervals. This measure takes into account the frequency of deaths that will occur in this and subsequent intervals. As age increases and the population shrinks, the total person-years that the survivors have to live necessarily diminish.

Abridged Life Table for Males in Malaysia, 1995

	1	2	3	4	5	6
AGE	PROPORTION DYING IN THE AGE INTERVAL	NUMBER LIVING AT THE BEGINNING OF THE AGE INTERVAL	NUMBER DYING DURING THE AGE INTERVAL	PERSONS-YEARS LIVED IN THE AGE INTERVAL	PERSONS-YEARS LIVED IN THIS AND SUBSEQUENT INTERVALS	YEARS OF LIFE REMAINING (LIFE EXPECTANCY)
< 1	0.01190	100,000	1,190	98,901	6,938,406	69.38
1–5	0.00341	98,810	337	394,437	6,839,505	69.22
5–10	0.00237	98,473	233	491,782	6,445,067	65.45
10–15	0.00270	98,240	265	490,536	5,953,285	60.60
–	–	–	–	–	–	–
65–70	0.16050	70,833	11,368	325,743	928,004	13.10
70–75	0.25762	59,464	15,319	259,024	602,260	10.13
75–80	0.34357	44,145	15,167	182,808	343,237	7.78
80+	1.00000	28,978	28,978	160,428	160,428	5.54

Source: Department of Statistics, Malaysia, 1997.

Life expectancy is shown in Column 6. The total person-years lived in a given interval plus subsequent intervals, when divided by the number of persons living at the beginning of that interval, equals life expectancy — the average number of years remaining for a person at a given age interval (Column 5 ÷ Column 2 = Column 6). For example, dividing the number of person-years associated with Malaysian men who survive to age 70 (602,260) by the number of these men (59,464) shows they have an additional life expectancy of 10.1 years. With age, life expectancy actually rises, a kind of "bonus" for surviving. The 59,464 Malaysian men who survive to age 70 can expect to live more than 10 additional years, well past their life expectancy at birth of 69 years.

Source: Adapted from, Arthur Haupt and Thomas T. Kane, *Population Handbook*, 5th ed. (Washington, DC: Population Reference Bureau, 2004), p. 29–30.

Cause-specific death rates provide information on why people die and can be a useful policy tool. Morbidity statistics provide information on how people live, whether they live in full health or experience disabilities that may limit their participation in work or family life. Researchers have developed several innovative approaches to measuring health status that address the limitations of morbidity and mortality statistics by combining the information from both into a single unit. The WHO uses a single measure of death and disability that recognizes that years lived with a disability are not the same as healthy years. **Health-adjusted life expectancy (HALE)** reduces life expectancy by years spent with disabilities, and disabilities are weighted according to their level of severity and duration. The balance remaining is the expected number of years of healthy life.[4]

Using the HALE measure, healthy life expectancy in Sierra Leone for males is cut by 14 years from 48 to 34 and for females it is reduced from 50 to 37. This is the lowest healthy life expectancy of all 192 WHO member countries. In Denmark, disabilities reduce healthy life expectancy from 77 to 70 for males; females live 73 healthy years compared to an average life expectancy of 81. The percentage of life lost to disability tends to be higher in poorer countries because some limitations strike children and young adults, such as injury, blindness, paralysis, and the debilitating effects of several tropical diseases such as malaria. People in the healthiest regions (Europe) lose about 11 percent of their lives to disability, versus 15 percent in the least healthy region (Africa; Table 9–2).

[4]*World Health Report 2002: Reducing Risks, Promoting Healthy Life* (Geneva: World Health Organization, 2002), statistical annex.

TRANSITIONS IN GLOBAL HEALTH

The past two centuries have seen a remarkable improvement in health and life expectancy. Until 1800, life expectancy at birth averaged around 30 years, but it could be lower. In France between 1740 and 1790, life expectancy of males fluctuated between 24 and 28 years. In England between the 1500s and the 1870s, life expectancy ranged from 28 to 42 with an average of only 35 years.[5] Many people died in infancy or early childhood while a few lived on to old age. Life expectancy has steadily increased since the late nineteenth century, making dramatic gains in the twentieth century. From 1960 to 2008, life expectancy around the world increased from 50 to 69 years (Table 9–3). Among the low- and middle-income countries, the increase is especially dramatic. Life expectancy was only 44 years in 1960; in 2008, it was 67 years, more than a 50 percent increase. By comparison, the increase in the high-income countries over this time period was 11 years, from 69 to 80. This smaller increase is because the high-income countries had achieved significant gains in life expectancy earlier in their history.

All regions of the world experienced gains in life expectancy since 1960, whether or not they also experienced economic growth. In East Asia, where economic growth has been rapid (and China dominates the regional average), life expectancy has risen from only 39 years in 1960 to 72 years in 2008. In 2008, per capita income in East Asia was only 15 percent that of high-income countries, but its life expectancy was already 90 percent the level achieved by high-income nations. Despite its slow growth between 1960 and 1990, South Asia added more than five years of increased life expectancy per decade. Latin America, even with its lost decades, by 2008, had achieved the world's second-highest regional life expectancy, at 73 years.

TABLE 9–3 Increased Life Expectancy, 1960–2007

	LIFE EXPECTANCY, YEARS			CHANGE IN YEARS PER DECADE	
REGION	1960	1990	2007	1960–1990	1990–2007
Low and middle income	44	63	67	6.3	2.4
East Asia and Pacific	39	67	72	9.3	2.9
Europe and Central Asia	n.a.	69	70	n.a.	0.6
Latin America and Caribbean	56	68	73	4.0	2.9
Middle East and North Africa	47	64	70	5.7	3.5
South Asia	43	59	64	5.3	2.9
Sub-Saharan Africa	41	50	51	3.0	0.6
High Income	69	76	79	2.3	1.8
World	50	65	69	5.0	2.4

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

[5]James C. Rile, *Rising Life Expectancy: A Global History* (Cambridge: Cambridge University Press, 2001).

But not all the news is this positive. Sub-Saharan Africa experienced an increase in life expectancy between 1960 and the early 1990s, but many African nations have recorded declines since then, as young adults and children died prematurely due to HIV/AIDS. Tragically, life expectancy in parts of southern Africa has been reduced by a decade or more. Life expectancy declined in some other countries as well. In Russia, life expectancy fell sharply in the 1990s. Between 1990 and 1994, life expectancy for Russian men fell from 64 years to 58 years and for Russian women from 74 to 71 years.[6] Increases in cardiovascular disease (heart disease and stroke) and injuries accounted for two-thirds of the decline in life expectancy. Causes for this decline include high rates of tobacco and alcohol consumption, poor nutrition, depression, and a deteriorating health system. The underlying causes of these determinants of increased morbidity and mortality have been traced to both the cumulative effects of often poor living standards under communist rule of the Soviet Union and to the upheaval and stresses associated with Russia's transition. It took two decades, but today life expectancy in Russia is back to the levels of 1990, the start of its economic transition.

The examples from Africa and Russia demonstrate that declines in health and life expectancy can take place rapidly. But these outcomes remain exceptions to the general trend of improving health and increasing life expectancy. The gaps in health outcomes between the richest and the poorest regions of the world have become markedly smaller since 1960. The difference in life expectancy between low- and middle-income compared to high-income nations was 25 years in 1960 but fell to only 13 years by 1990—the level it is also at today. Despite a lack of income convergence across nations (see Chapter 4), there is strong evidence of a convergence in life expectancy.[7]

THE EPIDEMIOLOGIC TRANSITION

Dramatic improvements in life expectancy and reductions in infant and child mortality during the past century resulted in equally dramatic demographic and socioeconomic changes. These changes led to the lower fertility rates and slower rates of population growth discussed in Chapter 7. Worldwide, youth dependency declined while elderly dependency rose. In 1970, 40 percent of the populations of all low- and middle-income nations were children 14 years and under. In 2009, the percentage had fallen to 29 percent. The elderly, those 65 and older, were only 4 percent of the population of developing nations in 1970 and today are still only 6 percent; they are 15 percent of the population in high-income countries. In coming decades, demographic change will

[6]Francis Notzon et al., "Causes of Declining Life Expectancy in Russia," *JAMA* 279, no. 10 (1998), 793–800.

[7]In the second half of the twentieth century, world life expectancy (computed as life expectancy by country weighted by each country's population) rose by about 18 years, while the weighted standard deviation *fell* from 13 to 7 years. See Charles Kenny, "Why Are We Worried About Income? Nearly Everything That Matters Is Converging," *World Development* 33, no. 1 (January 2005). See also Gary S. Becker, Tomas J. Philipson, and Rodrigo R. Soares, "The Quantity and Quality of Life and the Evolution of World Inequality," *American Economic Review* 95, no. 1 (2005), 277–91.

accelerate as the consequences of the developing world's transition from high to low population growth rates combine with falling child mortality and rising life expectancy.

As societies age and health improves, the pattern of disease and causes of death also shift in a generally predictable pattern. This shift in disease pattern is referred to as the **epidemiologic transition**. One early characterization of the epidemiologic transition identified three main stages: the age of pestilence and famine, the age of receding pandemics, and the age of degenerative and human-made diseases.[8] The age of pestilence and famine, which covered most of human history, was a time of frequent epidemics and famines. Chronic malnutrition existed, as did severe maternal and child health problems. Health outcomes were affected by environmental problems such as unsafe water, inadequate sanitation, and poor housing. During the age of receding pandemics, death rates fell as infectious disease and famines declined. As people lived longer, they began to experience heart disease and cancer in greater numbers. In the age of degenerative and human-made diseases, mortality rates are low and chronic degenerative diseases, such as heart disease, diabetes, and hypertension; lifestyle diseases, the result of smoking and excessive use of alcohol, for example; and environmental diseases, such as those caused by pollution, take the place of infectious diseases as the primary causes of death. Changes have been made to this initial theory, but it remains a useful framework.

Some researchers add other stages of epidemiologic transition, such as emerging and reemerging infectious disease.[9] Old infectious diseases may cause increased morbidity and mortality due to resistance to antimicrobial therapies, as is the case for drug-resistant tuberculosis, and new pathogens continue to emerge. HIV/AIDS is the best known of these new infections, although other new infectious diseases have emerged over the past 30 years, including the Ebola virus, severe acute respiratory syndrome (SARS), and recent strains of avian and H1N1 swine flu.

The twentieth century has seen a major shift in causes of death and disability from infectious diseases to noncommunicable disease. However, not all countries made this transition fully. Many of the poorest countries and poor subpopulations in middle-income countries still suffer from high rates of infectious diseases, maternal and perinatal conditions, and nutritional deficiencies that ceased to be problems in high-income nations. Table 9–4 demonstrates the extent of these disparities for 2004, the most recent update assembled by WHO. Globally, 60 percent of mortality is due to noncommunicable conditions such as cardiovascular disease and cancer, 30 percent of mortality is due to communicable conditions, and 10 percent is due to injuries. High-income countries have reduced mortality from communicable disease to only 7 percent, with 87 percent due to noncommunicable conditions. In stark

[8]A. R. Omran, "The Epidemiologic Transition: A Theory of the Epidemiology of Population Change," *Millbank Memorial Fund Quarterly* 49, no. 4 (1971), 509–37.

[9]Ronald Barrett, Christopher W. Kuzawa, Thomas McDade, and George J. Armelagos, "Emerging and Re-Emerging Infectious Diseases: The Third Epidemiologic Transition," *Annual Review of Anthropology* 27 (1998), 247–71.

TABLE 9–4 Mortality by Cause, 2002

CAUSE OF DEATH*	WORLD	AFRICA	AMERICAS	SOUTHEAST ASIA	EUROPE	EASTERN MEDITERRANEAN	WESTERN PACIFIC
I. Communicable diseases, maternal and perinatal conditions, and nutritional deficiencies (%)	32.1	71.9	14.7	39.3	5.9	41.7	14.2
II. Non-communicable conditions (%)	58.8	21.1	76.3	50.6	85.8	48.9	75.5
III. Injuries (%)	9.1	7.0	9.1	10.0	8.3	9.4	10.3

*Total mortality.

Source: *World Health Report 2004: Changing History* (Geneva, World Health Organization 2004), annex table 2.

contrast, 68 percent of mortality in sub-Saharan Africa remains due to communicable diseases and only 25 percent to noncommunicable conditions. The situation in India (characteristic of all of South Asia) and Latin America lies in between sub-Saharan Africa and the high-income countries. The differences in causes of death are also seen in Table 9–5. In high-income countries, such as Denmark, all but one of the 10 leading causes of death are from noncommunicable disease. In sub-Saharan Africa, the top four leading causes of mortality are infectious conditions.

China, a country that experienced rapid economic growth, also experienced a substantial epidemiologic transition. The distribution of the causes of death in China is rapidly approaching those of the high-income nations (Table 9–4). Heart diseases, cancers, and strokes are now the leading causes of death, accounting for approximately two-thirds of total deaths among adults 40 years and older. As recently as the early 1960s, famines in China accounted for tens of millions of deaths. China no longer faces the threat of famine but must instead contend with the more modern health risks of smoking; obesity; and the various causes of cancers, heart disease, and other noncommunicable conditions.

THE DETERMINANTS OF IMPROVED HEALTH

Improved health and increases in life expectancy are due to several factors. Major advances in agriculture and food distribution in Europe led to the disappearance of famine and starvation, a concern up to the nineteenth century in some parts of

TABLE 9-5 Leading Causes of Mortality, 2002

RANK	CAUSE*	PERCENT OF TOTAL
World		
1	(II) Ischemic heart disease	12.6
2	(II) Cerebrovascular disease	9.7
3	(I) Lower respiratory infections	6.8
4	(I) HIV/AIDS	4.9
5	(II) Chronic obstructive pulmonary disease	4.8
6	(I) Diarrheal diseases	3.2
7	(I) Tuberculosis	2.7
8	(I) Malaria	2.2
9	(II) Trachea/bronchus/lung cancers	2.2
10	(III) Road traffic accidents	2.1
Total for 10 causes		51.2
Africa		
1	(I) HIV/AIDS	19.6
2	(I) Malaria	10.7
3	(I) Lower respiratory infections	10.4
4	(I) Diarrheal diseases	6.6
5	(I) Childhood diseases	4.9
6	(II) Cerebrovascular disease	3.4
7	(I) Tuberculosis	3.3
8	(II) Ischemic heart disease	3.1
9	(III) Road traffic accidents	1.8
10	(III) Violence	1.3
Total for 10 causes		65.1
Europe		
1	(II) Ischemic heart disease	24.8
2	(II) Cerebrovascular disease	15.1
3	(II) Trachea/bronchus/lung cancers	3.8
4	(II) Lower respiratory infections	2.9
5	(II) Chronic obstructive pulmonary disease	2.7
6	(II) Colon/rectum cancer	2.4
7	(II) Hypertensive heart disease	1.9
8	(II) Cirrhosis of the liver	1.8
9	(III) Self-inflicted	1.7
10	(II) Stomach cancer	1.6
Total for 10 causes		58.7

*(I), communicable diseases; (II), noncommunicable diseases; (III) injuries.

Source: *World Health Report 2004: Changing History* (Geneva, World Health Organization 2004), annex table 2.

the region. Famines were common and devastating in East and South Asia throughout much of the twentieth century. With few exceptions, North Korea being one, famine no longer appears as a threat in most of Asia. The same cannot be said for sub-Saharan Africa, where famines still occur regularly. Even without famine, malnutrition continues to plague billions of people in Africa and elsewhere and contributes to lower life expectancy. Malnutrition makes individuals more susceptible to infections and less able to fight them off.

Rising incomes generally allow for better nutrition and housing and improved survival rates. Throughout the twentieth century, life expectancy has been strongly associated with per capita income. Life expectancy rises rapidly with income, especially at low levels of income. Increased income allows people, particularly the poor, to buy more food, better housing, and more health care. However, since 1900, on an almost decade by decade basis, life expectancy has shifted upward, so that more health is realized for a given income. This upward shift indicates that health depends on more than income.[10]

Public health measures such as clean water, sanitation, and food regulation contributed to the decline in child mortality in the late nineteenth century and continue to do so today. In the late nineteenth century, Robert Koch showed that the bacterium *Mycobacterium tuberculosis* causes tuberculosis and people began to understand germs. Many other pioneers of science and medicine, such as Louis Pasteur and Joseph Lister, made equally important discoveries. Simple precautions such as preparing food and disposing of waste hygienically, eliminating flies and rodents, and quarantining the sick had far-reaching benefits. The discovery that cholera and typhoid were transmitted through impure water dates to the 1850s, but access to safe drinking water is still far from universal. Medical technology became important to controlling infectious diseases in the 1930s, when antibacterial drugs and new vaccines were introduced. All these developments coupled with the rise of public health institutions that provide for improved sanitation, distribute vaccines, control disease vectors such as mosquitoes, and offer surveillance of disease outbreaks protect societies against the major infectious causes of death.

The critical role of science and institutions as determinants of health, independent of income, is reflected in the historical record. In 1900, life expectancy in the United States was 47 years and per capita income US$5,500 (2005). Compare this to the situation in low- and middle-income countries in 2009. Life expectancy was substantially higher at 67 years, even though per capita income, at about US$5,000 (2005, PPP) was about 10 percent lower. The explanation is straightforward. Penicillin and other antibiotics were unknown in 1900. Urban living conditions were overcrowded, and few vaccines were available. Infectious diseases took their toll on the U.S. population just as they do in many low-income nations today.

Finally, education plays an important role in improving health. This was not always the case. Child mortality differed little by education or income in the United States in the last decade of the nineteenth century, but these factors made a big difference by the early twentieth century. The implication is that affluence and education did not matter much until the underlying scientific knowledge was present. Better-educated individuals acquire and use new information more quickly. This helps explain the large differences in child mortality by mother's education observed in developing countries.[11]

[10]World Bank, "Health in Developing Countries: Successes and Challenges," in *World Development Report 1993: Investing in Health* (Oxford: Oxford University Press, 1993), pp. 17–36.

[11]World Bank, "Health in Developing Countries."

HEALTH, INCOME, AND GROWTH

As indicated before, in both rich and poor countries, there is a strong positive relationship between better health and economic growth. In Chapter 3, we saw that countries with higher life expectancy have faster economic growth. Figure 9–2 takes a different perspective, showing the relationship between under-five mortality and the *level* of income. Mortality rates are highest in the poorest countries, and they drop off sharply in conjunction with higher incomes. In countries with per capita income below US$2,000 (PPP), on average, 129 out of every 1,000 children do not live to their 5th birthday. Whereas in rich nations people reaching their 100th birthday is no longer that unusual, in poor nations 1 out of every 8 children still does not live to see his or her 5th birthday. For incomes between $4,000 and $6,000, the child mortality rate is much lower, falling to just 44 per 1,000. For countries with incomes above $10,000, only 12 out of 1,000 die early in life.

Because key measures of health are strongly related to both income levels and growth rates, it should be no surprise that they are also strongly related to poverty. Figure 9–3 makes the link with poverty, showing that countries with higher life expectancies have much lower rates of poverty. In countries with life expectancy below 45 years,

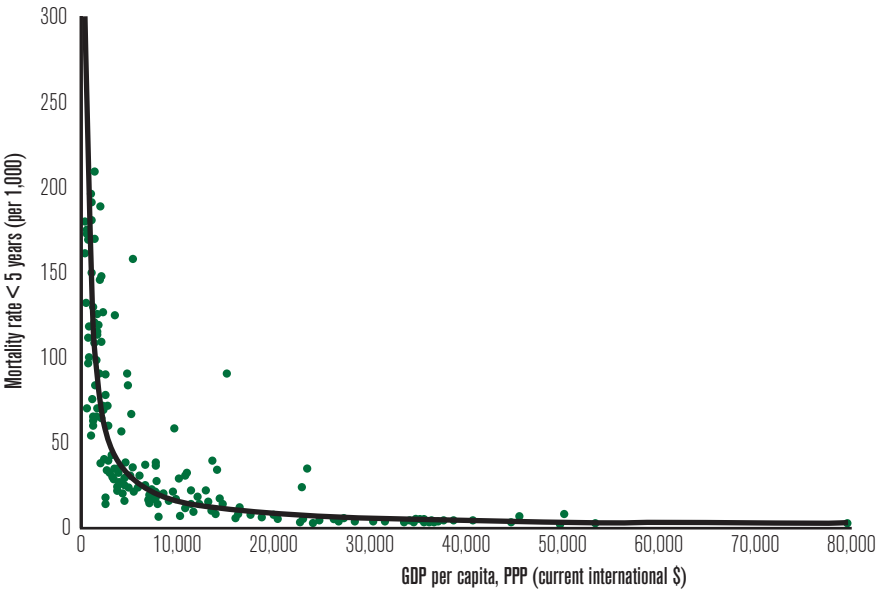


FIGURE 9–2 Child Mortality and Income, 2007

Based on data for 2007 from 168 countries of all income levels around the world.

GDP, gross domestic product; PPP, purchasing power parity.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

81 percent of the population lives below the $2-a-day poverty line; in countries with life expectancy greater than 70 years, only 17 percent live below the poverty line.

The strength of these relationships is not in doubt. The question is this, What is causing what? Higher income and reductions in poverty could lead to better health. But the causality could run in the opposite direction; improved health could lead to faster economic growth, higher incomes, and less poverty. In fact, both channels probably are at work. Better health and higher incomes work together in a mutually reinforcing and virtuous cycle, in which improvements in health support faster economic growth leading to higher incomes, and higher incomes facilitate even better health. Unfortunately, the cycle can also work in reverse, as declines in income can weaken health (as in Russia and other transition economies), and the onset of new diseases, such as HIV/AIDS, can lead to a fall in income.

INCOME AND HEALTH

As incomes increase, individuals, households, and societies at large are able to increase spending on a range of goods and services that directly or indirectly improve health. Individuals and households can buy more and better food, so there is a strong causal link from higher income to improved nutrition to better health.[12] Research

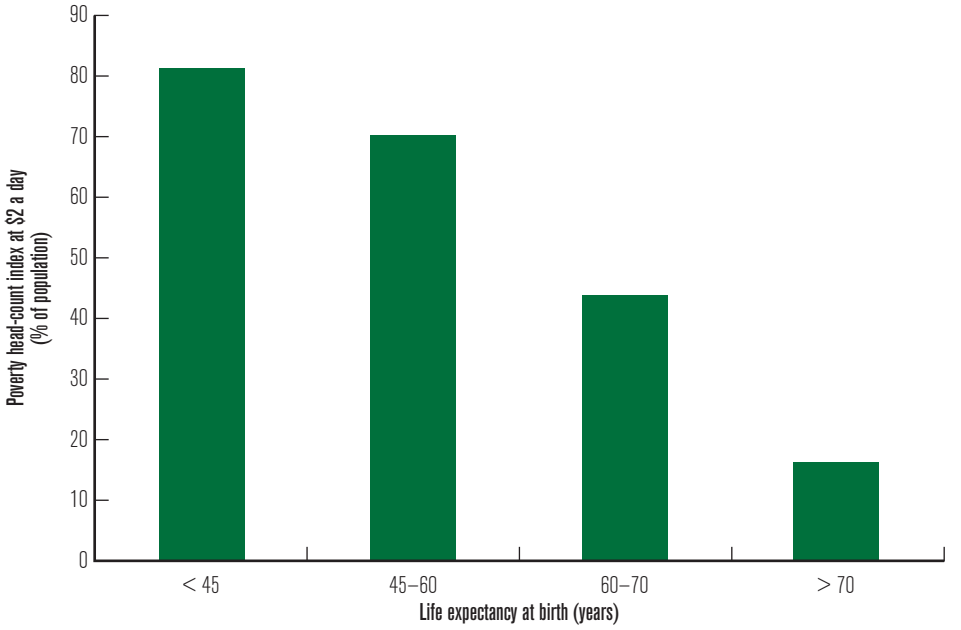


FIGURE 9-3 Poverty and Life Expectancy, 2000–2007 (Average)

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

[12]John Strauss and Duncan Thomas, "Health, Nutrition and Economic Development," *Journal of Economic Literature* 36 (June 1998).

shows that, as incomes rise, child height (an indicator of long-term nutrition) tends to increase in many low-income countries. With higher incomes, poor households can more easily afford clean water and basic sanitation facilities, either as an outhouse or through indoor plumbing, which can help control disease. Also, as incomes rise, poor households can afford better shelter, which in turn should keep them healthier. They are also more likely to be able and willing to seek medical care when it is needed, either because they can afford the transportation costs to the clinic or pay for medicines or other out-of-pocket expenses. For individuals living on subsistence incomes, paying for even the most basic medicines and healthcare can sometimes be a struggle. As incomes rise, such expenditures become more routine.

For society as a whole, as incomes rise, there is greater ability to build public health clinics and hospitals, train more doctors and nurses, and pay for public health services such as immunization campaigns or insect-spraying programs. Poor countries cannot afford to build the same kinds of public and private health systems that richer countries can, and they may not be able to afford to pay their doctors and nurses the kinds of salaries needed to keep them home rather than taking positions in richer countries. Interruptions in the supply of electricity can reduce the efficacy of medications—for example, many vaccines need to be kept refrigerated before inoculation, and when the electricity grid fails, the stock of vaccines is damaged. It is not uncommon in developing countries to find hospitals and clinics without medicines or clean water. Often, because operating budgets are so small and patients cannot afford to pay much, the patient's family must supply basic food and care to a family member in the hospital. As a society's income rises, it is much better able to pay for a reasonable healthcare system. Both sub-Saharan Africa and Latin America devote about 7 percent of GDP to health, including both private and public expenditures. In sub-Saharan Africa with 2009 GDP per capita of US$2,138 (PPP), spending 7 percent of national income on health translates to about $150 per person per year. In Latin America, where per capita incomes average a much higher US$10,555 (PPP), a similar 7 percent allocation to health expenditures provides $740 per person per year, which can buy far more hospital beds and medicine.

Economists Lant Pritchett and Lawrence Summers provide one estimate of the potential impact of rising incomes on health. They find that the long-run income elasticity of under-five mortality is between -0.2 and -0.4, meaning that each 10 percent increase in income is associated with a 2 to 4 percent decline in child mortality rates. According to this result, if income growth in low- and middle-income nations during the 1980s had been 1 percent higher, as many as half a million child deaths would have been averted in 1990 alone.[13] Pritchett and Summers' article is titled, "Wealthier Is Healthier," suggesting that improvements in income, holding other factors constant, lead to lower child mortality and increased life expectancy. The relationship between income levels and life expectancy is captured in Figure 9–4,

[13]Lant Pritchett and Lawrence H. Summers, "Wealthier Is Healthier," *Journal of Human Resources* 31, no. 4 (1996), 841–68.

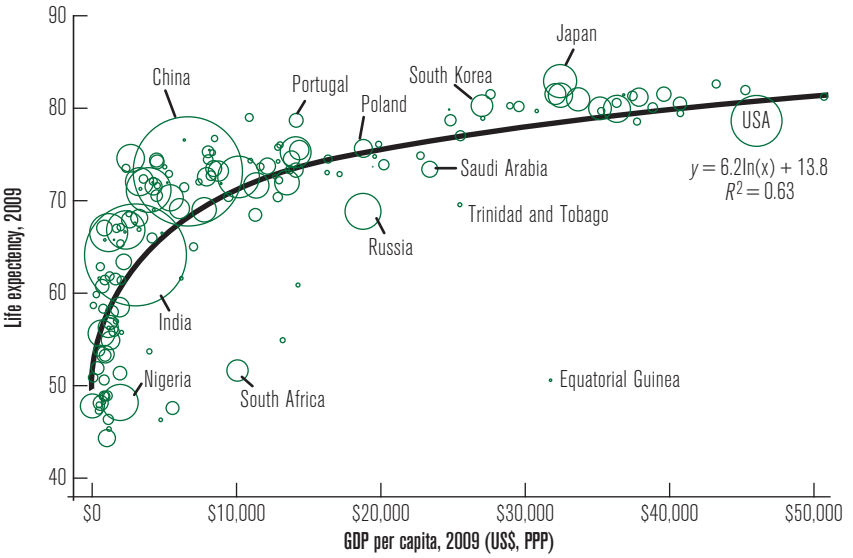


FIGURE 9–4 The Preston Curve: Life Expectancy Versus GDP Per Capita

GDP, gross domestic product; PPP, purchasing power parity.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

sometimes referred to as the **Preston curve**, after the work of demographer Samuel Preston. In this "bubble" graph, in which each bubble is proportional to the population of the country represented, life expectancy climbs rapidly at lower levels of per capita income and then levels off at higher incomes.

But as we have observed before, correlation does not imply causation. Income is not the only barrier to the use of even life-saving treatments and the significance of rising incomes in explaining health trends is, perhaps surprisingly, a matter of some debate. Preston makes this point in his 1975 article, arguing that most of the gains in life expectancy result from upward shifts in curves like the one depicted in Figure 9–4 rather than from movements along them.[14] A similar conclusion is reached by economists David Cutler, Angus Deaton, and Adrianna Lleras-Muney, who write,

> We tend to downplay the role of income. Over the broad sweep of history, improvements in health and income are both the consequence of new ideas and new technology, and one might or might not cause the other. Between rich and poor countries, health comes from institutional ability and political willingness to implement known technologies, neither of which is an automatic consequence of rising incomes.[15]

[14]Samuel Preston, "The Changing Relation between Mortality and Level of Economic Development," *Population Studies* 29, no. 2 (1975).

[15]David Cutler, Angus Deaton, and Adrianna Lleras-Muney, "The Determinants of Mortality," *Journal of Economic Perspectives* 20, no. 3 (2006), 116.

They cite historical improvements in life expectancy in countries across Europe in the nineteenth century, which happened at more or less the same time despite quite different rates of economic growth. They note that in the twentieth century both China and India made substantial gains in reducing child and infant mortality *before* the reforms that resulted in rapid economic growth. And once reforms were in place, further progress in infant mortality slowed down. Neither the historical record nor the cross-country evidence suggests "that economic growth will improve health without deliberate public action."

The creation and dissemination of medical knowledge and public health practices was central to the historical rise in life expectancy in the now-developed nations (Box 9–2). The diffusion of this knowledge to poor nations today remains a key determinant of health outcomes in developing countries. It is also a determinant somewhat independent of the per capita income levels of the poorer nations. This is another example of economic historian Alexander Gerschenkron's hypothesis about the advantages of backwardness. Poor countries can learn from the scientific knowledge, most of it generated in the developed nations, about disease transmission and interventions to promote health and extend people's lives.

While these cross-country associations between income levels and various measures of health indicate broad trends, the decisions of individual households are important as well. This is a theme of the widely acclaimed book *Poor Economics*, by Abhijit Banerjee and Esther Duflo. The MIT economists provide many examples of poor households making decisions that run counter to the health of their children and families that cannot be explained by low incomes alone. Take the case of oral rehydration therapy (ORT), a simple mixture of water, salts, and sugar that is an effective treatment for common, and often fatal, diarrheal diseases in children. It is extremely inexpensive, at under 50 cents per treatment, especially when viewed relative to its life-saving benefits. Poverty alone is an unlikely explanation for the failure to use ORT. Instead, in western India, mothers may not believe that ORT works and prefer that health workers give their children an antibiotic. Despite being prescribed oral rehydration salts for their children, these mothers may simply not use them. Banerjee and Duflo also discuss insecticide-treated bed nets that have proven effective at reducing exposure to mosquitoes that transmit malaria. They also are inexpensive—at less than $10 for a net that provides protection for up to five years. But even when bed nets (or childhood immunizations) are provided for free, the take-up rates are far from universal. More than low incomes must be involved.

Banerjee and Duflo conclude, "The poor seem to be trapped by the same kinds of problems that afflict the rest of us—lack of information, weak beliefs, and procrastination among them." People living in developed nations do not have to worry about chlorinating their water because it comes that way automatically. Poor households

BOX 9-2 "HOW BENEFICENT IS THE MARKET? A LOOK AT THE MODERN HISTORY OF MORTALITY"

Richard Easterlin, a distinguished economic historian and demographer, investigated the reasons behind the decline in mortality rates that began in Europe in the nineteenth century. While rising levels of income improved nutrition and generally contributed to increased resistance to disease, economic growth in this period also increased exposure to disease as industrialization gave rise to urbanization and the concentration of people in closer quarters.

In explaining the great improvement in life expectancy that occurred in parts of Europe in the nineteenth century, Easterlin identifies the centrality of scientific discoveries and the role of public health initiatives. Easterlin also reflects on the role of markets versus the state in achieving these outcomes:

> Only with the growth, first, of epidemiological and, then, bacterial knowledge did effective techniques emerge for controlling infectious disease. These techniques focused primarily on the prevention of the spread of disease—first via controlling the mode of transmission, and subsequently via immunization. It is these methods of prevention that have been chiefly responsible for the great improvement in life expectancy. In the last half century the advance of knowledge has added methods of curing disease to the arsenal available to fight infectious disease, particularly with the development of antibiotics, but the great bulk of the reduction in infectious disease has been accomplished largely by preventive methods.
>
> The control of infectious disease involves serious issues of market failure—information failures, externalities, public goods, principal-agent problems, and so forth. The market cannot be counted on for such things as the provision of pure water and milk, the proper disposal of sewage, control of pests such as mosquitoes and rats, the supply of uncontaminated food and other manufactured products, immunization of children and adults against major infectious diseases, and the dissemination of new knowledge regarding personal hygiene, infant and child care, food handling and preparation, care of the sick, and the like. There is also a serious market failure problem with regard to the distribution of anti-microbials because of externalities associated with the development of disease-resistant bacteria.
>
> The title of this article posed the question, how beneficent is the market? The ubiquity of market failure in the control of major infectious disease supplies the answer: if improvement of life expectancy is one's concern, the market cannot do the job . . . [T]he assumption that the market, in solving the problem of economic growth, will solve that of human development is belied by the lessons of experience. Rather than a story of the success of free market institutions, the history of mortality is testimony to the critical need for collective action.

Source: Excerpt from "How Beneficent is the Market? A Look at the Modern History of Morality," by Richard A. Easterlin. *European Review of Economic History 3*, December 1999. Copyright © 1999 Cambridge University Press. Reprinted by permission.

need to take active steps each day to ensure their water is safe, a far more difficult process even without the tough constraints the poor face. We trust the advice of our health professionals based on experience; given the low quality of medical care in many poor nations such trust is understandably absent. To overcome some of these constraints Banerjee and Duflo recommend *nudges*, whether providing small monetary incentives for childhood vaccines (2 pounds of dal per vaccination and a set of stainless-steel plates for completing the course) or just making it easier to do the right thing (a "one turn" chlorine dispenser next to a village well that makes chlorinating water as easy as possible).[16] Understanding the behavior of households and individuals is critical to improving health outcomes.

HEALTH AND PRODUCTIVITY

There is increasingly strong evidence of the causality also running in the opposite direction, in which improved health leads to faster economic growth, higher incomes, and reduced poverty. Several studies have shown that higher life expectancy in an early period (say, 1960) leads to faster subsequent economic growth (say, from 1960 to 2000). The relationship is usually found to have diminishing returns, meaning that an extra year of life expectancy at birth has a big effect on growth when life expectancy is 45 years but a smaller effect when life expectancy reaches 70 years. A typical estimate suggests that, after controlling for other factors that influence growth, each 10 percent improvement in life expectancy at birth translates into an increase of 0.3 to 0.4 percentage points per year in economic growth.

Better health affects both fundamental causes of economic growth: productivity gains and increased investment. Healthier people tend to be more economically productive because they are more energetic and mentally alert. A healthier worker is able to harvest more crops, build more furniture, assemble more computers, or make more service calls than a worker who is ill or lethargic. A study of sugarcane workers in Tanzania showed that workers with schistosomiasis, a disease caused by a parasitic worm that causes fever, aches, and fatigue, were able to cut much less cane and therefore earn much less in wages (because they were paid by the amount they cut) than workers in the same estate that did not have the disease.[17] Healthier workers not only are more productive while they work but also lose fewer workdays because of illness. The types of work available to many low-income, low-skilled workers put a premium on physical health because they typically rely more on strength and endurance than higher-skilled desk jobs. Moreover, aside from their own health, if a worker's

[16]Abhijit Banerjee and Esther Duflo, *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* (New York: Public Affairs, 2011), chap. 3.

[17]A. Fenwick and B. M. Figenschou, "The Effect of *Schistosoma mansoni* on the Productivity of Cane Workers in a Sugar Estate in Tanzania," *Bulletin of the World Health Organization* 47, no. 5 (September 1972), 567–72.

family is healthier, the worker loses fewer workdays in caring for family members that are ill. A dramatic example of the latter point involves HIV/AIDS. Firms in southern Africa, where HIV prevalence rates are the highest in the world, lose workdays due not only to the illnesses of employees but also in time lost to attending the high number of funerals of coworkers and family members.

A family member's health can also have an impact on children's education, which in turn has implications for future income. Families that cope with long-term illnesses, such as HIV/AIDS or tuberculosis of an adult, may rely on children to work the fields, find other ways to bring in income, or care for the sick family member, thus preventing the child from attending school and investing in his or her own human capital. The child's health also directly affects his or her schooling. Poor health can reduce cognitive ability in students and undermines schooling through absenteeism and shorter attention spans. As noted earlier, experimental evidence from Kenya finds that students in schools that received a simple treatment for hookworms had higher school enrollment rates and, in cases of wage employment, higher earnings as young adults. When child mortality declines, parents tend to have fewer children, allowing them to invest more in each child's health and education, as discussed in Chapter 7.

Childhood health can affect labor productivity years later when children grow older and join the workforce. The early effects of health on physical strength and cognitive ability can have lasting impacts. Nobel laureate Robert Fogel found a strong relationship among nutrition, body size, and labor productivity.[18] Research on Brazil inspired by Fogel's earlier work identifies a relationship between a worker's height and wages. Taller workers, who, after controlling for other factors, are likely to have had better nutrition during childhood, earn much higher wages, reflecting higher levels of productivity. The relationship in Brazil is strong: A 1 percent increase in height is associated with a nearly 8 percent increase in wages. When workers are healthier and more productive, firms are willing to pay them higher wages.

HEALTH AND INVESTMENT

Beyond its impact on labor productivity, improved health can influence the other key channel for achieving economic growth, increased saving and investment.

As people expect to live longer, they have greater incentives to make long-term investments in their farm or other business and their human capital. By contrast, people with poor health have to divert more of their income into medical expenses, reducing their capacity to save. Poor people facing catastrophic illness sometimes must sell some of their assets, such as livestock or farm tools, thus depleting their capital stock, to pay for medicines or simply to feed their family if they no longer can

[18]See, for example, Robert Fogel, "New Findings on Secular Nutrition and Mortality: Some Implications for Population Theory," in M. R. Rosensweig and O. Stark, eds., *Handbook of Population and Family Economics*, vol. 1a (Amsterdam: Elsevier Science, 1997) pp. 433–81.

work. Health can also affect public saving and investment. Governments fighting endemic diseases must allocate more of their spending to these purposes, reducing the amount available for building roads or other capital investment.

In South Africa, the high prevalence of HIV/AIDS is viewed as both a public health challenge and as a macroeconomic problem that reduces economic growth. The disease reduced the size of the workforce and its productivity at the same time as it has decreased the ability of households, firms, and the government to invest in the future. Households face lost labor income and increased expenditures on health; firms confront increased expenditures on health benefits and insurance, higher absenteeism and job turnover, and greater training costs due to premature deaths; and government finds that health expenditures crowd out other high-priority expenditures. Some studies suggest that HIV/AIDS may have cost South Africa a full percentage point of annual economic growth—a huge amount for an economy averaging GDP per capita growth of only 1.5 percent.[19]

Reduction or elimination of diseases can help substantially increase the amount of usable productive assets, including land. The classic example of disease impeding a critical investment is the construction of the Panama Canal in the late nineteenth century (Box 9–3). Contemporary examples include large tracts of land rendered uninhabitable by endemic disease. The dramatic reduction of river blindness in west Africa allowed increased farming in the fertile lands near riverbanks that previously were breeding grounds for the disease (see Box 14–4 for a discussion of the fight against river blindness in the context of foreign aid programs). Schistosomiasis makes it unsafe for people to enter lakes and streams in sections of Africa, and trypanosomiasis (sleeping sickness) restricts the range of the livestock industry. The reduction of malaria allowed farmers to work new lands in Malaysia, Sri Lanka, and the Terrai area of northern India. Singapore's economic prosperity was made possible partly because rampant malaria was brought under control on the island. Similarly, disease can undermine foreign investment, as investors shun environments where HIV/AIDS, malaria, tuberculosis, and other diseases are more prevalent.

THREE CRITICAL DISEASES

Dramatic changes are taking place in global health. The reduction of infectious disease allows the survival of many individuals to old age. Reductions in fertility rates contribute to a change in population structure so that the elderly represent a larger portion of

[19]Jeffrey D. Lewis, "Assessing the Demographic and Economic Impact of HIV/AIDS," in Kyle Kauffman and David Lindauer, eds., *AIDS and South Africa: The Social Expression of a Pandemic* (Hampshire, UK: Palgrave Macmillan, 2004).

BOX 9-3 MALARIA, YELLOW FEVER, AND THE PANAMA CANAL

On February 1, 1881, capitalized by over 100,000 mostly small investors, the French Compagnie Universelle du Canal Interocéanique began work on a canal that would cross the Isthmus of Panama and unite the Atlantic and Pacific Oceans. Ferdinand de Lesseps, builder of the Suez Canal, led the project. In the first months, the digging progressed slowly but steadily. Then, the rains began, and the crew soon discovered what they were up against: mile upon mile of impassable jungle, day upon day of torrential rain, insects, snakes, swamps, heat, and endemic disease—smallpox, malaria, and yellow fever.

In 1881, the company recorded about 60 deaths from disease. In 1882, the number doubled, and the following year, 420 died. The most common killers were malaria and yellow fever. Because the company often fired sick men to reduce medical costs, the numbers probably reflect low estimates. By the time the company halted the project and went out of business in December 1888, about $300 million had been spent, and 20,000 men had died.

In 1904 President Theodore Roosevelt instigated a treaty with Panama that gave the United States the right to build the canal and create the 10-mile-wide Canal Zone, which amounted to sovereign American territory surrounding the waterway. The U.S. Army dispatched surgeon Colonel William Gorgas to Panama to tackle malaria and yellow fever. Gorgas was fresh from Havana where he had helped eradicate yellow fever, following discoveries by his colleague Major Walter Reed and others that the disease was carried by a mosquito. The fight against yellow fever also substantially reduced malaria, building on discoveries by British bacteriologist Ronald Ross that the parasite that causes malaria was transmitted by the *Anopheles* mosquito.

The efforts to control the two diseases were successful. Yellow fever was totally eradicated. Deaths due to malaria in employees dropped from 11.6 per 1,000 in November 1906 to 1.2 per 1,000 in December 1909. But the impact spread well beyond the canal workers. Deaths from malaria in the Panamanian population fell from 16.2 per 1,000 in July 1906 to 2.6 per 1,000 in December 1909. The canal was completed in 1914, one of the greatest construction miracles of the early twentieth century. The project powerfully demonstrated that malaria and yellow fever could be controlled over a wide geographical area, paving the way for investment and increased economic activity.

Sources: Adapted from the Public Broadcasting Service's film series *The American Experience: The Story of Theodore Roosevelt,* "TR's Legacy—The Panama Canal," www.pbs.org/wgbh/amex/tr/panama.html; and Centers for Disease Control and Prevention, *Malaria: The Panama Canal,* www.cdc.gov/malaria/history/panama_canal.htm.

the population, imposing new demands on healthcare systems. The process of industrialization, urbanization, and modernization creates its own set of health problems. Pollution damages the environment and affects health. Increased tobacco use adds to the burden of diseases such as lung cancer and heart disease. Alcohol abuse, injuries, and stress create their own set of health problems. A shift in diet from vegetables and cereals to highly processed foodstuffs coupled with a decline in activity levels leads to rising obesity in both rich and poor nations and contributes to chronic illnesses.

While the burden of noncommunicable disease grows, the unfinished agenda of infectious diseases remains. There is a great disparity in what has been achieved in healthcare between the developed and developing nations and even between groups within nations. The rest of this chapter focuses on a handful of diseases that are preventable, treatable, or curable, yet pose significant problems in developing countries. This is followed by examples of how countries and the international community have addressed some of these and other diseases. A disease-based approach offers a snapshot of some of the health issues facing developing countries and some of the ways these challenges have been met.

Three of the developing world's most prominent and deadly infectious diseases are HIV/AIDS, malaria, and tuberculosis (Box 9–4). They were the first, fourth, and sixth leading causes of death in sub-Saharan Africa (Table 9–5). These three diseases have spread steadily in recent decades and together kill nearly 4.5 million people every year. Relative to high-income countries, the burden of these diseases (in terms of deaths per capita) was 23 times greater in developing countries, resulting in tremendous economic loss and social disintegration. One of the many tragedies of each of these diseases is that they are preventable and, with adequate resources and institutions, can be treated effectively.

HIV/AIDS

Although HIV/AIDS[20] has affected humans since at least the 1930s, it was virtually unknown until it was first recognized in 1981. Since then, the pandemic[21] has spread inexorably around the globe, exacting a terrible toll on the health and welfare of the world's poorest communities. More than 33 million people were living with HIV/AIDS in 2009, about 1 of every 200 people in the world, and almost 2 million people died that year of AIDS-related causes. HIV/AIDS is the leading cause of death among adults aged 15 to 59 worldwide and has killed an estimated 25 million people.

[20]This section draws on information from UNAIDS, 2010 *Report on the Global AIDS Epidemic* (Geneva: UNAIDS, 2010); Global Fund, *2004 Disease Report* (Geneva: Global Fund to Fight AIDS, Tuberculosis, and Malaria, 2004); and information from WHO (www.who.int) and the U.S. Centers for Disease Control and Prevention (CDC; www.cdc.gov).

[21]HIV/AIDS is usually referred to as a pandemic rather than an epidemic. Both terms often refer to the spread of an infectious disease. *Epidemics* refer to an illness that shows up in a larger number of cases than is normally expected. *Pandemics* refer to an even higher number of cases spread over a larger geographic area.

BOX 9-4 HIV/AIDS, MALARIA, AND TUBERCULOSIS: SOME BASICS

HIV stands for human immunodeficiency virus. Once infected with HIV, a person will always carry the virus. The disease primarily is a sexually transmitted disease, but it can also be transmitted through needles or contaminated blood. Over time, HIV infects and kills white blood cells called CD4 lymphocytes (or T cells), leaving the body unable to fight off infections and cancers. **AIDS** stands for acquired immune deficiency syndrome and is caused by HIV. A person with AIDS has an immune system so weakened by HIV that he or she usually becomes sick from one of several opportunistic infections or cancers, such as pneumonia, Kaposi sarcoma (KS), diarrhea, or tuberculosis. AIDS usually takes several years to develop from the time a person acquires HIV, usually between 2 and 10 years. AIDS can be treated but not cured. With successful antiretroviral (ARV) therapy, which requires medication for life, the body can remain healthy and fight off most viruses and bacteria, and people living with HIV can resume a relatively normal life.

Malaria is a serious and sometimes fatal disease caused by a plasmodium parasite transmitted by the female *Anopheles* mosquito. The mosquito transmits malaria by taking blood from an infected individual and passing it to another. Malaria can also be transmitted by blood transfusion and contaminated needles and syringes. Patients with malaria typically are very sick with high fevers, shaking chills, and flulike illness. Four species of plasmodium infect humans, but the most dangerous is *Plasmodium falciparum,* which causes the most severe disease and deaths and is most prevalent in sub-Saharan Africa and some parts of Southeast Asia. Although malaria can be a fatal disease, illness and death from malaria are largely preventable.

Tuberculosis (TB) is a bacterium that infects the lungs and is spread through the air when someone with active TB coughs, sneezes, talks, or otherwise releases microscopic droplets containing *Mycobacterium tuberculosis,* which then enter the respiratory system of another person. Because proximity and poor ventilation speed transmission, TB is a particular problem in densely populated or enclosed settings, such as urban slums or prisons. Each person with active TB infects an average of 10 to 15 people each year. Only about 10 percent of people infected with the bacterium develop active TB and the accompanying symptoms: fever, weight loss, chronic cough, chest pain, bloody sputum, loss of appetite, and night sweats. A third of those with active TB die within a few weeks or months if they are not treated. The remainder struggle with recurrent infections or the disease goes into remission, but it reemerges from time to time to cause pain, fever, and possibly death. The HIV epidemic fueled the resurgence of TB. HIV weakens the immune system, and people who are infected with HIV are

particularly vulnerable to TB. Antibiotics to treat and cure TB patients have been available for over 50 years. Successful treatment requires strict patient compliance. Antibiotics must be taken regularly for six months or more. Failure to do so has contributed to the spread of multidrug-resistant TB, which is much harder to cure.

Sources: "HIV InSite," University of California at San Francisco; WHO; Centers for Disease Control and Prevention; and the Global Fund to Fight AIDS, Tuberculosis, and Malaria.

Sub-Saharan Africa bears the largest burden, with two-thirds the world's infection, even though it is home to only one-tenth the world's population. Across sub-Saharan Africa nearly 1 in 20 adults carries the virus, but huge variations exist across the region. Some west African countries, such as Senegal, have maintained adult prevalence rates of less than 1 percent for more than a decade. In these low-prevalence countries, the disease is confined primarily within groups that practice high-risk behavior, such as sex workers and their clients. In southern Africa, HIV/AIDS has taken hold within the general population. In Swaziland, prevalence rates among pregnant women are a mind-boggling 42 percent. In three countries in southern Africa, more than one-fifth of all adults are HIV positive: Botswana, Lesotho, and Swaziland. Nearly 1.3 million people died of HIV/AIDS-related causes in 2009 in sub-Saharan Africa, more than two-thirds the global total.

The good news is that HIV epidemics in the region are either stable or declining. Among the five countries in sub-Saharan African with the largest HIV epidemics, four (Ethiopia, South Africa, Zambia, and Zimbabwe) reduced new HIV infections by more than a quarter between 2001 and 2009. Prevalence rates are falling in several African countries, including Kenya, Uganda, and Zimbabwe, representing an important achievement in turning the tide on the pandemic. However, even these trends convey a mixed message. Some claim that declines in prevalence rates in Uganda can be attributed at least partially to behavioral changes linked to the government's ABC program: promoting abstinence, being faithful to one partner, and using condoms. But studies demonstrated that, in some regions of the country, falling prevalence rates are less the result of behavioral change than of high mortality. Because of AIDS-related deaths, there were simply fewer HIV-positive adults in 2003 than there were a decade earlier.[22]

The pandemic is less severe in other regions but still a major threat that continues to spread. Adult prevalence rates stabilized around 1 percent in the Caribbean, where there is wide variation between and within countries. In 2009 adult HIV prevalence

[22]In the Rakai region in Uganda, researchers attribute approximately 5 percentage points of the 6.2 percent decline in HIV prevalence between 1994 and 2003 to increased mortality. UNAIDS, *AIDS Epidemic Update* (Geneva, UNAIDS, 2005), p. 26.

was 0.1 percent in Cuba but 3.1 percent in the Bahamas. In Asia, Thailand has implemented a successful campaign to limit the spread of the disease, and the rate of new infections fell more than 25 percent between 2001 and 2009 in India and Nepal; in other countries, however, the number of infected people is growing. With 2.4 million people living with HIV, India has the third largest number of infected people in the world behind South Africa and Nigeria. The epidemic is evolving in China, where unsafe blood transfusion and drug abuse traditionally accounted for the majority of HIV cases. Of the 740,000 HIV-positive people in China in 2009, 59 percent were infected through sexual transmission. The epidemic is worsening in eastern Europe and central Asia (the number of people living with HIV in this region has almost tripled since 2000), fueled by a surge in infections among people who inject drugs.

Women are increasingly vulnerable to HIV, and globally, girls and women account for more than half of those living with HIV. In Africa, the picture is especially grim for young women; those aged 15 to 24 are as much as eight times more likely than men to be HIV-positive. Socioeconomic factors contribute significantly to women's vulnerability to HIV infection. Most are infected by partners who practice high-risk behavior. Compounding this problem is that as many as 9 out of 10 HIV-positive people in sub-Saharan Africa do not know that they are infected. Social disempowerment and lack of access to HIV education or services further contribute to the soaring pandemic among women. In 24 sub-Saharan countries, two-thirds or more of young women aged 15 to 24 years lacked comprehensive knowledge of HIV transmission.[23] Children are also vulnerable. More than 2.5 million children are HIV-positive, and there were more than 0.25 million AIDS-related deaths among children in 2009. The vast majority of infected children acquire HIV through transmission from their mother during pregnancy, labor and delivery, or breastfeeding. Antiretroviral (ARVs) drugs, such as nevirapine, significantly reduce mother to child transmission but still are not always available in sub-Saharan Africa.

The HIV/AIDS pandemic has had a crippling effect on many countries that were making significant progress in health. Whereas most countries in sub-Saharan Africa had achieved steady increases in life expectancy in the 1960s, 1970s, and early 1980s, in several countries this progress stopped in the late 1980s and has since reversed. Figure 9–5 shows changes in life expectancy for six of the most severely infected countries. Tragically, life expectancy dropped from an average of 61 years in 1992 to just 47 years in 2007. But as Figure 9–5 suggests, some countries (Botswana, Namibia, Zimbabwe) seem to have reached a turning point, with life expectancy rising once again.

The effects on communities and families are devastating as parents, children, and community leaders become sick and die. Unlike most diseases, HIV/AIDS attacks primarily young adults and the economically active segment of the population. Other diseases tend to incapacitate the very young, the old, or those already weakened by other

[23]UNAIDS and World Health Organization, "HIV Infection Rates Decreasing in Several Countries but Global Number of People Living with HIV Continues to Rise," press release, November 21, 2005, available at www.who.int/hiv/epiupdate2005/en/index.html.

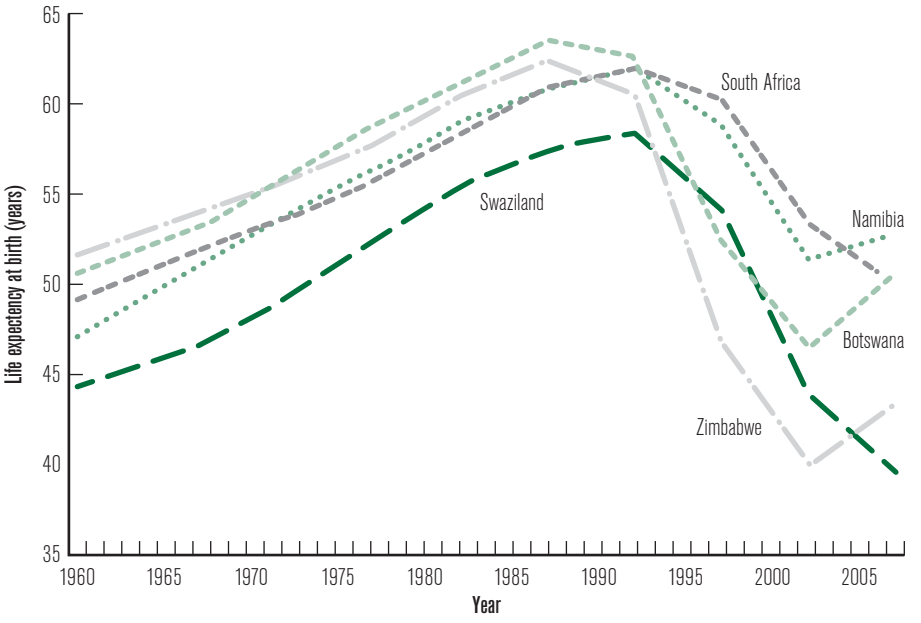


FIGURE 9–5 Life Expectancy at Birth in Selected Most HIV-Affected Countries

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

conditions. HIV/AIDS is different. In Africa, HIV disproportionately destroys the lives of society's most productive members. Households face catastrophe as income earners fall ill and eventually succumb to the disease. The long duration of AIDS illness makes it particularly costly. In addition to these losses, others in the family must stop work to care for those that are ill. Small farms and gardens go untended, and children may be taken out of school to tend to sick family members or to go to work to earn extra income. Families become further impoverished as they face increasing medical costs and have less income to pay for food, clothing, or other expenses. In some cases, funeral costs may be higher than healthcare costs and add to a family's burden. Studies in Tanzania and Thailand estimate that the cost of medical care and funerals can be greater than the household's annual income, triggering borrowing, asset sales, or other coping strategies that can have long-term financial welfare implications.[24] Many of the effects of HIV/AIDS are intergenerational: Worldwide there are now 16.6 million AIDS orphans who have had one or both parents die from the disease.

[24]Steven Russell, "The Economic Burden of Illness for Households: A Review of Cost of Illness and Coping Strategy Studies Focusing on Malaria, Tuberculosis and HIV/AIDS." Disease Control Priorities Project (DCPP) Working Paper 15, August 2003, available at http://www.dcp2.org/file/30/wp15.pdf. Funeral costs have become a significant expense to both households and firms. They have become so common that firms are beginning to limit the number of days that employees can take for funeral leave each month. This is an issue, in part, because family members often are buried in their traditional villages, requiring days of travel for mourners.

Combating the epidemic requires strong national and international commitment. Until scientists develop other interventions, such as a cure or preventive vaccine, the key to halting the spread of the disease is strong prevention programs coupled with adequate care and treatment. In richer countries, where people can afford treatment, the development of ARVs has made AIDS a chronic disease that can be managed, rather than a fatal disease, allowing those infected to resume a relatively normal lifestyle. Death rates for HIV/AIDS in Europe and North America fell by 80 percent in the four years after the introduction of antiretroviral therapy. Prophylactic treatment with ARVs in combination with other interventions has almost entirely eliminated HIV infection in infants in industrialized countries.[25] But for most people in the poorest countries, ARVs remain out of reach. At the end of 2009, about one in three Africans who needed treatment had access to these drugs. Beginning in 2003, access to ARVs in developing countries began to improve, and in a few countries, treatment coverage is now widespread. In Botswana, Cambodia, Croatia, Cuba, Guyana, Namibia, Romania, and Rwanda, more than 80 percent of those that need ARVs are now receiving them, but these are exceptions rather than the norm.

Delivering antiretroviral therapy on a global level requires not only good science and health policy but also political and economic initiatives. The most cost-effective way to provide the massive quantities of drugs required at hugely discounted prices is to encourage the manufacture of generic versions of these antiretroviral agents. Issues of trade-related intellectual property rights and patents must be considered here. One challenge is to encourage such manufacture while providing incentives for major pharmaceutical companies to develop newer medicines as patients develop resistance to the current stock.[26]

The international community was slow to respond to the growing crisis during the 1990s, but that has begun to change. In 1996, only about US$300 million was available to fight the disease globally. By 2009, an estimated US$15.9 billion was available, a figure that includes not just donor funds but the steadily increasing funding that comes from country governments and out-of-pocket spending by directly affected individuals and families. Yet this amount is still $10 billion short of what the Joint United Nations Programme on HIV/AIDS (UNAIDS) has estimated was required in 2010. Due to fiscal constraints in response to the global financial crisis, 2009 was the first time that international assistance did not increase from the previous year. Finding the needed funds will be a continuing challenge, and even when they are found, they will have significant opportunity costs because funding is diverted from other key challenges to fight HIV/AIDS.

[25]World Health Organization, *World Health Report 2004: Changing History* (Geneva: World Health Organization, 2004).

[26]Diane V. Havlir and Scott M. Hammer, "Patents Versus Patients? Antiretroviral Therapy in India," *New England Journal of Medicine* 353, no. 8 (August 25, 2005), 749–51.

MALARIA

Malaria[27] is estimated to claim the lives of 2,000 children every day and is a major contributor to illness in the developing world. WHO estimates that approximately half of the world's population is at risk of malaria in over 106 countries, primarily poor countries located in the tropics. Malaria contributes to almost 800,000 deaths and 225 million cases of severe illness each year. Africa, again, bears the heaviest burden, with about 78 percent of the cases of clinical malaria and 91 percent of the world's malaria deaths. Almost one-fifth of the deaths in children under five in Africa are caused by malaria. Malaria also continues to be a problem in Southeast Asia, India, Latin America, and some parts of Oceania.

Most diseases in developing countries have a disproportionate impact on the poor, but this is especially the case with malaria. Poor families are more likely to live in slum areas or in the countryside where malaria is common, less likely to be able to afford simple prevention steps (like insecticide-treated bed nets), and less likely to receive treatment once fever strikes. In the 1990s, more than half of malaria deaths occurred in the poorest 20 percent of the world's population, a higher percentage than for any other disease of major public health importance. This percentage is

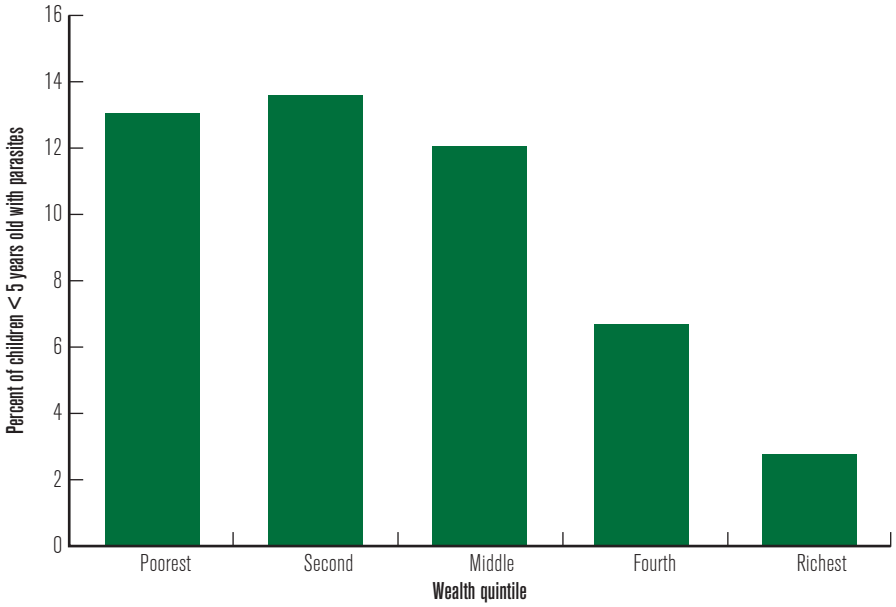


FIGURE 9–6 Parasite Prevalence Is Higher in Poor Children

Source: Government of the Republic of Zambia Ministry of Health, *Zambia National Malaria Indicator Survey 2008*, available at www.nmcc.org.zm/files/ZambiaMIS2008Final.pdf, accessed February 2012.

[27]This section draws on information from UNICEF and WHO, *2010 World Malaria Report*, (Geneva: World Health Organization, 2010); the Global Fund, 2004 *Disease Report*; and Tina Rosenberg, "What the World Needs Now Is DDT," *New York Times* (April 11, 2004).

probably higher today. Figure 9–6 shows that in Zambia 13 percent of children from the poorest quintiles had malaria, whereas less than 3 percent of children from the richest quintile were infected.

Many countries have begun to make progress in reducing malaria. Prevention strategies include the use of insecticide-treated bed nets for those at highest risk of malaria, intermittent preventive drug treatment of pregnant women, and spraying with insecticides and other forms of vector control. For those who become sick with malaria, the disease can be controlled with early treatment. As of yet, there is no vaccine against the disease, although research is under way to find one. Unfortunately, because this is a disease of the poor, commercial opportunities are limited for selling a vaccine and recouping expensive research and development costs, and research on this vaccine has been underfunded. Box 9–5 describes one idea aimed at overcoming this market failure.

BOX 9–5 MAKING MARKETS FOR VACCINES

Vaccines save 3 million lives a year, prevent long-term disability and illness in millions more, and are one of the most cost-effective ways of enhancing health and reducing poverty in developing countries. Yet, there are no vaccines for some of today's biggest killers, such as malaria, HIV, and tuberculosis; and new vaccines developed for hepatitis B, pneumococcus, rotavirus, and cervical cancer, which kill millions of people a year, are not affordable in developing countries. If vaccines are so cost-effective and save so many lives, why are vaccines not being developed and used for these killer diseases?

In some cases, the science is highly uncertain, such as for HIV. But in other cases, private firms that conduct much of the research and development (R&D) for vaccines have no financial incentive to do so for diseases that affect mainly poor countries. The costs of developing these vaccines are high, and incomes in many poor countries are too low to afford them. Private markets fail to develop vaccines because a price that is low enough for developing countries to afford is too low to create incentives for private companies to invest in developing new vaccines or production facilities for existing ones.

Many of the most widely used vaccines are aimed at diseases that affect children in both rich and poor countries, such as diphtheria, tetanus, pertussis, and polio. Because these vaccines are used around the world, the market is large enough to cover a pharmaceutical company's costs for both R&D and production. In some cases, vaccines could be sold at a higher price in richer countries, to recover R&D costs, and a lower price in poorer countries. But some of the

highest-priority diseases for developing countries lack the high priority for rich countries. AIDS, diarrheal diseases, malaria, and other diseases, which are some of the biggest killers in developing countries, are not nearly as high on the list for rich countries, so the total potential market for vaccines for these diseases is much more limited. The pharmaceutical industry tends to respond to where the financial pay-off is the greatest, serving the health needs of the richer countries while neglecting diseases concentrated in low-income countries.

The costs of development and the risks of failure for vaccines are high. Vaccine development can take 7 to 20 years for basic research, clinical testing, regulatory approval, production, and distribution, and even the most promising vaccine candidates can fail at any stage. Estimates of the total costs of developing a new vaccine vary but range from several hundred million dollars to $1.5 billion. It is not realistic to hope that private companies will spend the money necessary to develop urgently needed vaccines then just give them away or sell them at a low price to poor countries.

One way to overcome this market failure is for donors to provide an advance market commitment (AMC) to provide private firms the financial returns they need to develop new vaccines while keeping the cost low for consumers in low-income countries. Sponsors and donors, such as governments, international agencies, and foundations, make binding commitments to pay for a desired vaccine if and when it is developed. Developing countries could buy the vaccine at a low and affordable price, and sponsors would commit to providing the balance up to a guaranteed price, thus guaranteeing market returns for the vaccine developer that are comparable to other products. Once a specified number of vaccines was purchased at the guaranteed price (enough to cover R&D costs), the supplier would be committed to selling further treatments at a lower, affordable price without the sponsor's top-up. This step would ensure that developing countries could afford to continue purchasing the vaccine once payments under the commitment had been made.

A commitment creating a market comparable to the sales revenue of an average new medicine would cost about $3 billion. As an example, sponsors could make a legally binding commitment to purchase 200 million treatments of a malaria vaccine for $15 each. Poorer nations would decide to buy the vaccine at a low price (say, $1 per treatment) and the sponsors would provide the additional $14. After the first 200 million treatments, the company would be committed to sell additional treatments at $1, enough to cover ongoing production costs. These purchases, along with sales to the military, travelers, and some middle-income countries that can afford a higher price, would create the $3 billion market.

This approach complements rather than substitutes for other approaches for developing vaccines, such as support for public research facilities and universities.

R&D in public health often involves government-funded research into the "public good" of knowledge about diseases, which can then be appropriated by the private sector for commercial development. AMC provides clear incentives for private firms to invest in the development of critical vaccines. It also respects intellectual property rights for vaccines and removes pressures on firms to sell new products at a loss. For donors, there would be no cost unless a vaccine is actually developed, minimizing their risks. AMC would use public funds to make markets work more effectively to solve some of the developing countries' most pressing health problems.

In 2009 the first AMC was launched by the Global Alliance for Vaccines and Immunisations (now GAVI Alliance), a public–private partnership including major donors, pharmaceutical companies, and foundations. This pilot AMC is designed to accelerate access to existing vaccines against pneumococcal diseases. Such diseases include pneumonia, which is estimated to kill 1.6 million people per year, most of them children and about 90 percent occurring in developing nations. A pneumococcal vaccine has existed since 2000 and is part of regular immunization programs in developed countries. But the existing pneumococcal vaccine sells at over $70 per dose, far too expensive for widespread adoption in low-income countries. Funding from GAVI Alliance will permit the long-term price of the vaccine to fall to $3.50 and prevent an estimated 7 million childhood deaths by 2030.

Sources: The idea for the AMC was developed by a working group hosted by the Center for Global Development and is based on an earlier idea by Harvard University economist Michael Kremer. See Owen Barder, Michael Kremer, and Ruth Levine, *Making Markets for Vaccines: Ideas to Action* (Washington, DC: Center for Global Development, 2005). Much of the text in this box is drawn from this report. Also see Michael Kremer and Rachel Glennerster, *Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases* (Princeton, NJ: Princeton University Press, 2004).

Malaria can have substantial economic costs on top of its public health impact. Research by economists John Gallup and Jeffrey Sachs concluded that the presence of a high malaria burden reduces economic growth by 1.3 percent per year.[28] Although that figure probably incorporates the impact of other diseases that occur in tandem with malaria (such as HIV/AIDS), there is little doubt that malaria creates a heavy burden in countries with high prevalence rates. Malaria programs can be very cost-effective. One study, based on data from sub-Saharan Africa, estimated that the net benefits of a complete package of malaria interventions were about 18 times higher than the costs.[29]

[28]John Gallup and Jeffrey Sachs, "The Economic Burden of Malaria," *American Journal of Tropical Medicine and Hygiene Special Supplement* (June 2001).

[29]Global Fund, 2001 *Disease Report*, p. 37.

But even cost-effective programs can prove controversial. The insecticide of choice for combating malaria is dichlorodiphenyl-trichloroethane (DDT), but donors rarely pay for its use. The irony is that DDT was once used extensively and successfully in the United States, southern Europe, and many developing countries as a means of malaria control and eradication. But overuse of DDT in the United States, especially to control agricultural pests, was linked to environmental damage affecting wildlife. Author and social activist Rachel Carson brought attention to the problem in her seminal book *Silent Spring*, published in 1962. (The title refers to the impact of DDT on bird populations.) DDT use eventually was banned in the United States and other countries, but remains a highly effective and inexpensive (its patent expired long ago) means of eradicating the mosquito that transmits malaria. When sprayed on the interior walls of dwellings (the preferred mode for preventing the spread of malaria), DDT has minimal environmental consequences and has not been found harmful to humans. But given its pariah status in the West and the fear of maintaining a double-standard (selling an insecticide to poor countries that is banned in rich ones), DDT has not played a significant role in saving lives in Africa and other malaria-endemic regions for many years.

The level of international attention to and spending on malaria control in poor regions has been dismal. Funds by the major external financers of malaria control (the Global Fund, the World Bank, and the U.S. President's Malaria Initiative) stagnated at US$1.8 billion in 2010. Estimates by the WHO find that effective malaria control requires more than US$5 billion annually.

TUBERCULOSIS

Until 20 years ago, tuberculosis (TB)[30] was uncommon in the industrialized world and many assumed that the disease had been largely conquered. But, in recent years, TB has reemerged as a virulent killer. More than 2 billion people, one-third of the world's population, are infected with the TB bacterium. Most people who are infected carry the bacterium in their body without symptoms, but each year 9.4 million people develop active TB and exhibit fever and other symptoms. TB causes or contributes to 1.7 million deaths each year, with 90 percent of the deaths occurring in developing countries. About a quarter of those who die also are infected with HIV, which weakens the immune system and makes people more vulnerable to developing active TB.

As with other diseases, the poor are particularly vulnerable to TB. Studies in India have shown that the prevalence of TB is two to four times higher among groups with low income and no schooling. The poor are more likely to live in overcrowded conditions where the airborne bacterium can spread easily. Poor nutrition and

[30]This section draws on information from WHO, *Global Tuberculosis Control 2010* (Geneva: World Health Organization, 2010).

inadequate sanitation also add to the risk. And as with other diseases, once the poor are infected, they are less likely to be diagnosed and treated. The economic costs of TB can be significant, stemming from lost work for the patient and caregivers, extra nutritional needs, treatment costs, transportation to and from clinics, and withdrawal of children from schools. In the developing world, an adult with TB can lose an average of three to four months of work.[31] Every year in India alone, more than 300,000 children leave school because of their *parent's* TB. By some estimates, TB depletes the incomes of the world's poorest communities by up to $12 billion every year, and lost productivity can cost an economy on the order of 4 to 7 percent of GDP.[32]

The internationally recommended strategy to control TB is directly observed treatment, short course (DOTS), which combines a regular TB drug dosage with clinical observation visits. Full treatment takes many months, but unfortunately some patients stop taking their medicines during that period because they start to feel better or the drug supply is unreliable. This may lead to the emergence of drug-resistant TB. A particularly dangerous form of drug-resistant TB is multidrug-resistant TB (MDR-TB). Rates of MDR-TB are high in some countries, including Russia, and threaten TB-control efforts. From a public health perspective, poorly supervised or incomplete treatment of TB is worse than no treatment at all. When people fail to complete the treatment regime, they may remain infectious and develop a resistance to treatment. People they infect have the same drug-resistant strain. While drug-resistant TB generally is treatable, it requires extensive chemotherapy (up to two years of treatment), which is often prohibitively expensive (often more than 100 times more expensive than treatment of drug-susceptible TB) and more toxic to patients. Direct observation of treatment helps TB patients adhere to treatment. According to WHO, the DOTS strategy produces cure rates of up to 95 percent, even in the poorest countries. A six-month supply of drugs for treatment under the DOTS strategy costs as little as $10 per patient. The World Bank has identified the DOTS strategy as one of the most cost-effective of all health interventions.[33]

For DOTS to be effective, there must be access to TB sputum microscopy for TB detection; standardized short-course treatment of six to eight months, including direct observation of treatment; an uninterrupted supply of high-quality drugs; a reporting system to monitor patient outcome and program performance; and a political commitment to sustained TB control. Since its inception in 1995, more than 41 million TB patients have been treated under DOTS and 184 countries have adopted the DOTS strategy. However, 1.8 million people fail to get access to TB treatment each year.

[31]World Health Organization, *Treatment of Tuberculosis: Guidelines for National Programmes*, 3rd ed. (Geneva: World Health Organization, 2003).

[32]Global Fund, 2004 *Disease Report*, p. 26.

[33]World Health Organization, "Tuberculosis," Fact Sheet No. 104 (Geneva: World Health Organization, 2005).

WHAT WORKS? SOME SUCCESSES IN GLOBAL HEALTH

Despite many challenges to health in the developing world, there are also many successes. The examples that follow recount specific actions by the health sector that saved millions of lives and improved millions more. These cases are excerpted from *Millions Saved: Proven Successes in Global Health*, by Ruth Levine and the What Works Working Group.[34] (The What Works Working Group is a group of 15 development specialists convened by the Center for Global Development.) *Millions Saved* describes 17 successful public health interventions in the developing world. The programs highlighted were implemented on national, regional, or global scales, addressed problems of substantial public health significance, demonstrated a clear and measurable impact on a population's health, lasted at least five consecutive years, and were cost-effective.

The cases presented in *Millions Saved* demonstrate what the health sector can do, even in the poorest countries. Innovative interventions that involve the community can reach the most remote regions. These cases also demonstrate that governments in poor countries can get the job done, including being the chief sources of funds for the interventions. In almost all the cases, the public sector, so often maligned for its corruption and inefficiency, was responsible for delivering care to the affected populations. The interventions included technological developments as well as basic changes in behavior that had a great impact on health. In the control of guinea worm in Africa, for example, families learned to filter their water conscientiously; and in the fight against deaths from dehydration due to diarrheal disease in Bangladesh, mothers learned how to mix a simple salt and sugar solution and taught the technique to their daughters. Interventions can also benefit from international coalitions or partnerships. Such cooperative ventures can break through bureaucracies, provide funding, bring technical capabilities, and generate the political will to sustain an effort in the face of competing priorities. It is also possible to determine whether improvements in health outcomes are due to specific interventions; in most cases, because special efforts were made to collect data that look at outcomes. Finally, these cases illustrate that success comes in all shapes: disease-specific programs, initiatives that improve access and quality, traditional public health interventions, and legal and regulatory reforms all can work, individually or in combination.

[34]Ruth Levine and the What Works Working Group with Molly Kinder, *Millions Saved: Proven Successes in Global Health* (Washington, DC: Center for Global Development, 2004). We summarize 5 of the 17 cases presented. Interested students should consult *Millions Saved* for more complete coverage of these five cases and to learn about the other cases.

PREVENTING HIV/AIDS IN THAILAND

Well-documented stories of large-scale success in HIV prevention are few, although many small programs have been shown to be effective. But in Thailand, a large-scale program was successful in changing behaviors associated with increased risk of HIV among sex workers and those who use their services. Thai authorities initially recognized the severity of the situation in 1988, when the first wave of HIV infections spread among injecting-drug users. Between 1989 and 1990, they found that brothel-based sex workers infected with HIV tripled from 3.1 percent to 9.3 percent, and a year later it reached 15 percent. Over the same period, the proportion of male conscripts who were HIV positive when tested on entry to the army rose from 0.5 percent to 3 percent, a sixfold increase in only two years.

Although prostitution is illegal in Thailand, and there was fear that the government's intervention could imply that it tolerated or even condoned it, officials agreed that the higher priority was preventing HIV from spreading further. To do this the government launched the "100 percent condom program," in which all sex workers in brothels were required to use condoms with clients. There was one straightforward rule: no condom, no sex. By creating a monopoly environment, clients, most of whom preferred unprotected sex, could not go elsewhere to find it. Health officials provided free boxes of condoms and local police held meetings with brothel owners and sex workers. Men seeking treatment for sexually transmitted diseases were asked to name the brothel they had used, and health officials would then visit the establishment to provide more information. In principle, the police could shut down any brothel that failed to adopt the policy. Such sanctions were used a few times early on but authorities generally preferred to work with the brothels rather than alienate them.

The campaign was successful: Nationwide condom use in brothels increased from 14 percent in 1989 to more than 90 percent in 1992. The rate of new HIV infections fell fivefold between 1991 and 1993–95. The program may have prevented 200,000 HIV infections during the 1990s. While the program was successful among sex workers, the program did little to encourage men and women, especially among teens and young adults, to use condoms in casual but noncommercial sex. A substantial risk remains that HIV will spread through unprotected sex in Thailand. In addition, HIV continues to spread among injecting-drug users, where the rate is as high as 39 percent. The cost of treating AIDS with antiretroviral drugs, as well as the cost of treating opportunistic infections, is a major challenge facing the Thai government and its citizens. The expense of the government's ARV program, in part, was responsible for a two-thirds decline between 1997 and 2004 in the budget for HIV prevention.

CONTROLLING TUBERCULOSIS IN CHINA

In 1991, China revitalized its TB program and launched the 10-year Infectious and Endemic Disease Control Project to curb its TB epidemic in 13 of its 31 mainland provinces. The program adopted the DOTS strategy. Trained health workers watched

patients at local TB dispensaries every other day for six months as they swallowed their antibiotic treatment. Information on each treatment was sent to the county TB dispensary, and treatment outcomes were sent quarterly to the National Tuberculosis Project Office.

China achieved a 95 percent cure rate for new cases within two years of adopting DOTS, and a cure rate of 90 percent for those who had previously undergone unsuccessful treatment. The number of people with TB declined by over 37 percent between 1990 and 2000, and an estimated 30,000 TB deaths were prevented each year. The program cost $130 million in total. The World Bank and WHO estimate that successful treatment was achieved at less than $100 per person. One life year was saved for an estimated $15 to $20.

Despite China's success in curing TB, the program achieved lower-than-hoped-for rates of case detection. China is not alone in this shortcoming and had an experience similar to other high-burden countries. One of the main contributing factors to the low case detection rate was inadequate referral of suspected TB cases from hospitals to the TB dispensaries. Because hospitals can charge for TB diagnosis and treatment, they have little economic incentive to direct patients to the dispensaries. As a result, despite regulations requiring referrals to dispensaries, most TB patients are diagnosed in hospitals, where treatment is often abandoned prematurely.

TB remains a deadly threat in China. Hundreds of millions of people are infected and 10 percent of these are predicted to develop active TB. The government of China faces the challenge of maintaining high cure rates in the provinces covered by the project while scaling up the DOTS program to the remaining half of the population. In response to the 2004 SARS epidemic, the government established a web-based reporting system throughout the country, making it mandatory to report 37 infectious diseases, including TB, within 24 hours. From 2004 to 2007 the proportion of TB cases referred from hospitals to dispensaries increased from 59 percent to 78 percent, and in 2009 China had one of the highest case detection rates for high burden countries.

ERADICATING SMALLPOX

Smallpox, which had affected 10 to 15 million people globally in 1966 and resulted in 1.5 to 2 million deaths, has been completely eradicated. The last recorded case of smallpox occurred in Somalia in 1977. Its eradication has been heralded as one of the greatest achievements of public health in world history. In addition to the direct impact on smallpox, the campaign brought important benefits to other health issues, such as improvements in routine immunization. During the smallpox eradication campaign it was discovered that more than one vaccine could be given at a time, an idea now taken for granted. In 1970, the Expanded Program on Immunization was proposed, which sought to add several vaccines to routine smallpox inoculation. By 1990, 80 percent of the children throughout the developing world were receiving vaccines against six childhood killers, compared to only 5 percent when the program started.

The cost of smallpox eradication can be compared to its benefits, captured by the costs of the disease that have been averted. In developing countries, the costs averted include expenditures for caring for patients as well as lost economic productivity due to illness. In 1976, it was estimated that the cost of caring for someone with smallpox in India was about $3 per patient, which translated to an annual cost of $12 million for the country due to the widespread nature of the disease. Although $3 per patient may seem small, bear in mind that, in many developing countries, public health spending in total is typically only $7 to $8 per person. In addition to the costs of care, by one estimate, India lost about $700 million each year due to diminished economic performance. Assuming 1.5 million deaths due to smallpox worldwide in 1967, smallpox cost developing countries at least $1 billion each year at the start of the eradication campaign.

In the industrialized countries, eradication allowed governments to save the cost of vaccination programs that had been in place to prevent the reintroduction of the disease. In the United States, the bill for 5.6 million primary vaccinations and 8.6 million revaccinations in 1968 alone was $92.8 million, or about $6.50 per vaccination. With other indirect costs of the vaccination program, expenditures in all developed countries were around $350 million per year. Therefore, combining developing and industrialized countries, the estimated global costs, both direct and indirect, of smallpox in the late 1960s was more than $1.35 billion per year. This figure represents the economic benefits of eradication.

The ultimate expenditures of the intensified eradication program were around $23 million per year from 1967 to 1979, financed with $98 million from international contributions and $200 million from the endemic countries. Thus the benefits of eradication far outweighed the costs. It has since been calculated that the largest donor, the United States, saved, in terms of costs averted, the total of all its contributions every 26 days, making smallpox prevention through vaccination one of the most cost-beneficial health interventions of our time.

ELIMINATING POLIO IN LATIN AMERICA

The world's largest public health campaign, the Global Polio Eradication Initiative, follows in the footsteps of smallpox eradication. In 1988, 125 countries were endemic for polio with an estimated 350,000 cases. By 2006, the number of cases had dropped to below 1,400 and only 6 countries and territories remained polio endemic: Afghanistan, Egypt, India, Niger, Nigeria, and the Palestinian Territories. However, poliovirus continues to spread to previously polio-free regions. In late 2004 and 2005, 10 previously polio-free countries were reinfected. Worldwide success is due to the coordinated efforts of regional and international polio eradication campaigns that have immunized hundreds of millions of children. The regional polio campaign in the Americas eliminated polio in just six years.

Polio, short for poliomyelitis, is caused by the intestinal poliovirus. Its most feared effect, paralysis, develops in less than 1 percent of all victims, when the virus affects the central nervous system. The most serious form of the disease causes paralysis that leaves a person unable to swallow or breath. Respiratory support is needed to keep patients alive and mortality runs as high as 40 percent.

In the 1930s and 1940, after a series of polio outbreaks, the American public called for a vaccine. The mobilization effort was led by the disease's most famous victim, U.S. President Franklin D. Roosevelt. In 1938, President Roosevelt created the National Foundation for Infantile Paralysis, later renamed the March of Dimes, to raise funds "a dime at a time" to support the quest for a vaccine. In 1952, the campaign paid off with Dr. Jonas Salk's discovery of an inactivated polio vaccine (IPV). Between 1955 and 1961, more than 300 million doses were administered in the United States, resulting in a 90 percent drop in the incidence of polio. In 1961, a second scientific breakthrough resulted in a new form of the vaccine. Dr. Albert Sabin's oral polio vaccine (OPV) had several advantages over the previous vaccine. The vaccine prevented paralysis, as did the IPV, but OPV went further by helping to halt person-to-person transmission. At approximately $0.05 per dose, OPV was cheaper than its predecessor, and because it is an oral vaccine that requires no needles, it is easier to administer on a wide scale by volunteers.

Successful vaccination programs in Latin America eventually led the Pan American Health Organization in 1985 to launch a program to eradicate polio from the Americas. The immunization strategy of the campaign centered around three primary components: achieving and maintaining high immunization coverage, prompt identification of new cases, and aggressive control of outbreaks. In countries where polio was endemic, national vaccine days (NVDs) were held twice a year, one to two months apart, reaching nearly all children younger than five. The NVDs were designed to vaccinate as many children as possible. While aggressive vaccination strategies helped slow polio's transmission in most of the region, the disease still lingered.

Operation Mop-Up was launched in 1989 to aggressively tackle the virus in its final bastions. The initiative targeted the communities where polio cases had been reported, where coverage was low, or where overcrowding, poor sanitation, weak healthcare infrastructure or heavy migration pervaded. In these communities, house-to-house vaccination campaigns were held to finally wipe out the disease. The last reported case in Latin America was in Peru in 1991. Polio was declared eradicated from the Americas a few years later when no further cases emerged.

Administration of an oral polio vaccine proved both inexpensive and cost-effective. The cost of immunizing a child with three doses of the polio vaccine (along with diphtheria, tetanus, and pertussis vaccine) is just $21. Even without taking into consideration such benefits as increased productivity, strengthened capacity to fight other diseases, and reduced pain and suffering, the polio eradication campaign was

economically justified based on the savings of medical costs for treatment and rehabilitation alone. The first five years of the polio campaign cost $120 million: $74 from national sources and $46 million from international donors. Taking into consideration the savings from treatment avoided, donor contributions paid for themselves in 15 years. Today, war and politics, not money, are the primary barriers to the total eradication of polio.

PREVENTING DEATHS FROM DIARRHEAL DISEASE

Diarrheal disease is one of the leading causes of death among children, causing nearly 20 percent of all child deaths. Worldwide, dehydration from diarrhea kills between 1.4 and 2.5 million babies each year, mostly in developing countries where children suffer an average of three episodes a year. Nearly 20 out of 1,000 children die of diarrheal disease before the age of two. This is a great public health failure because such deaths easily can be prevented at very low cost. Nevertheless, there has been some progress. Many countries, including Egypt, as documented later, have succeeded in disseminating knowledge about a life-saving treatment that resulted in dramatic declines in mortality associated with diarrhea and a large improvement in infant and child survival.

Diarrhea is an intestinal disorder characterized by abnormally frequent and watery stools. It is caused by a number of agents, including bacteria, protozoa, and viruses. Unclean water, dirty hands during eating, and spoiled food are the primary sources of transmission. The most effective modes of prevention include improved water supply and sanitation, improved hygiene, and immunization against measles (which can cause diarrhea). Dehydration is the most serious effect of diarrhea. When fluid loss reaches 10 percent of body weight, dehydration becomes fatal. In cases that are not fatal, dehydration can leave a child susceptible to future infections.

Avoiding death from dehydration requires the swift restoration of lost fluids and electrolytes. Until the development of oral rehydration therapy in the 1960s, the only effective treatment was through intravenous (IV) infusions in a hospital or clinic. IV therapy is not a treatment of choice in the developing world because of its high cost, the hardship of traveling to clinics, and shortage of both trained personnel and supplies. Many people turn to drugs, including antibiotics, which can stop diarrhea. However, the majority of these drugs have no proven health benefit and some can cause dangerous side effects.

Oral rehydration therapy has been called the most important medical discovery of the twentieth century. Developed in the 1960s in Bangladesh and India, it consists of a simple solution of water, salt, and sugar that is as effective in halting dehydration as intravenous therapy. It is immensely cheaper at just a few cents per dose, safer and easier to administer, and more practical for mass treatment. The effectiveness of ORT was proven during cholera outbreaks in refugee camps on the border of

India and Bangladesh in the 1970s. In 1971, the war for independence in what is now Bangladesh created 10 million refugees. The unsanitary conditions in overcrowded refugee camps resulted in outbreaks of cholera with fatality rates approaching 30 percent. Resources were not available for mass treatment with IV fluids. Oral treatment was proposed as an alternative. The results were extraordinary. Cholera fatalities in the camp using ORT dropped to less than 4 percent, compared to 20 to 30 percent in camps treated with intravenous therapy.

Egypt was one of the pioneers of national-level administration of oral rehydration therapy. In 1977, 1 in 20 children died of diarrhea before his or her first birthday. That year packets of oral rehydration salts were introduced at public clinics and for sale at pharmacies, but few mothers were aware of the treatment and even fewer used it. In 1982, only 10 to 20 percent of diarrhea cases used the packets of oral rehydration salts. Physicians also did not recommend them.

In 1980, the government launched the Strengthening of Rural Health Delivery Project to test how mothers and physicians could be persuaded to use ORT. Initially, in 29 rural villages, nurses taught mothers in their homes how to use ORT and physicians were educated about the therapy. ORT use rose dramatically, and as a result, child mortality was 38 percent lower than in control villages, and diarrhea-associated mortality during the peak season was 45 percent lower. Based on the success of these community trials, in 1981 Egypt began a massive program to promote ORT use among the country's 41 million residents. Financial and technical support came from the U.S. Agency for International Development (USAID) and the public health organization John Snow, Inc. The program involved the entire Ministry of Health, other branches of government, WHO, and UNICEF. The program worked through the existing health infrastructure in order to strengthen the capacity of the health services to deliver care.[35]

The program used several innovative approaches to increase the use of ORT. Packets were redesigned in smaller quantities and the project logo became the most recognized product label in Egypt. Production and distribution channels were developed. Health workers, nurses, and physicians were trained. A mass media campaign was launched in 1984 that took advantage of the 90 percent of households that had televisions. The campaign was very successful. By 1986, nearly 99 percent of mothers were aware of ORT. Infant mortality dropped by 36 percent and child (under-five) mortality by 43 percent between 1982 and 1987. Mortality due to diarrhea dropped 82 percent among infants and 62 percent among children during this same period. The program achieved success with an extremely cost-effective intervention. The average cost per child treated with ORT was less than $6, and the cost per death averted was between $100 and $200.

[35]M. el-Rafie et al., "The Effect of Diarrheal Disease Control on Infant and Childhood Mortality in Egypt: Report from the National Control of Diarrheal Diseases Project," *The Lancet* 335, no. 8690 (March 17, 1990).

ORT continues to be the most cost-effective means of treating dehydration, and its use has been adapted to the unique challenges found in different countries. In Bangladesh, where 90 percent of the over 100 million residents lives in rural areas with poor transportation in a country 10 times the size of Egypt, distribution of ORT packets was not feasible. In 1980, a program to promote ORT in rural Bangladesh began by training workers to go door to door and teach mothers about dehydration and ORT. Mothers were also taught how to make a homemade solution by mixing a three-finger pinch of salt, a fistful of sugar, and a liter of water. (Today, packaged oral rehydration salts are available in most of the country.) Between 1980 and 1990, 13 million mothers were taught to make oral rehydration mixtures. An evaluation of more than 7,000 households found that between 25 and 52 percent of cases of severe diarrhea used the mixture. Today, the usage rate of ORT in Bangladesh is 80 percent, and ORT is part of Bangladeshi culture. An increase in the use of ORT across the globe has slashed diarrhea mortality rates in children by at least half. ORT saves the lives of an estimated 1 million children each year.

LESSONS LEARNED

There is no simple prescription for success that comes out of comparing the cases presented in *Millions Saved*. Nor can one easily place blame for the dismal failures that let HIV/AIDS ravage southern Africa or prolong malaria's devastating toll on so many. Poverty is one of the explanations for the failures but it is not sufficient. Even poor nations have had success in combating diseases. The health sector is one with pervasive market failures: There are all the common problems of negative externalities, principal-agent problems, information failures, and public goods. An inability to resolve these issues reflects equally pervasive government failures at the local, national, and international levels.

Successful health initiatives are characterized by strong leadership. Former World Bank president Robert McNamara was personally committed to controlling river blindness in West Africa, the Thai government had charismatic leaders with the vision to launch a program to prevent the spread of HIV/AIDS, and Egyptian officials stood firmly behind plans to expand ORT from community trials to a national program. In contrast, South African presidents Nelson Mandela and Thabo Mbeki generally ignored HIV/AIDS as an issue until the disease had already turned into a pandemic and national tragedy. Malaria, after having been resolved in the developed nations, has been neglected elsewhere.

In addition to strong leadership and program champions, one needs a combination of a technological solution and an affordable delivery system. Affordability can require concessions from patent holders, as has happened with generic ARVs mass produced by pharmaceutical companies in Brazil and India. Public–private partnerships also have worked, in which drugs are provided at cost and distribution is handled by government authorities. Donors, whether public or private, can

play important roles. The Global Fund to Fight AIDS, Tuberculosis and Malaria; USAID; the World Bank; and others have provided critical finance as has the private Bill & Melinda Gates Foundation. Local commitment is equally essential. Delivering improved health requires the actions of millions of individuals, whether those trained as doctors, nurses, or engineers or relatively unskilled workers and volunteers who watch patients take their medications as part of DOTS programs or help administer oral vaccines in remote villages.

HEALTH CHALLENGES

The challenge for the twenty-first century is to continue the battle against communicable disease while developing strategies to combat emerging epidemics of noncommunicable conditions. The developing world continues to face the illnesses of poverty. Over 8 million children and 300,000 mothers die each year, even though most of these deaths can be avoided through cost-effective vaccine programs, rehydration therapy for diarrheal diseases, improved nutrition status, and better birthing practices. Many people believe that infectious diseases in low- and middle-income countries should be a high priority because the technical means exist to control them and they disproportionately affect the young. The Millennium Development Goals (MDGs) placed maternal and child health as a high priority and an integral part of poverty reduction. Over the past decade, the MDGs have moved maternal and child health from a primarily technical concern to one that is increasingly seen as a "moral and political imperative."[36]

The battle against infectious disease will be ongoing this century. New diseases will emerge and, in a more globalized world, move quickly across borders. Drug resistance will make disease eradication much harder, as the discovery of new drugs races against the ability of microbes to adapt and mutate into even more virulent strains. But infectious diseases are not the only challenge the low- and middle-income nations face. The *2002 World Health Report: Reducing Risks, Promoting Healthy Life* highlighted other factors that have adverse effects on health. The report identified 10 major global risk factors in terms of the burden of disease. Listed in order of their expected health risks, they are underweight children; unsafe sex; high blood pressure; tobacco consumption; alcohol consumption; unsafe water, sanitation, and hygiene; iron deficiency; indoor smoke from solid fuels; high cholesterol; and obesity and physical inactivity. These 10 risk factors already account for more than one-third of deaths worldwide.

[36] *World Health Report 2005: Make Every Mother and Child Count* (Geneva: World Health Organization, 2005), p. 3.

Underweight children are strongly related to poverty as are unsafe water, inadequate sanitation, and indoor air pollution. Unsafe sex is the main factor in the spread of HIV/AIDS, with a major impact in the poor countries of Africa and Asia. Some of these risk factors, such as high blood pressure and high cholesterol, tobacco and excessive alcohol consumption, obesity and physical inactivity, are more commonly associated with wealthy societies. However, with epidemiologic transitions, poor and middle-income countries increasingly face these risks as well. For improvements in health and life expectancy to continue, it will be necessary to address this *double burden* of disease in the decades ahead.

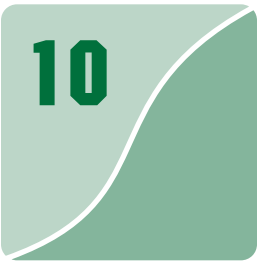
SUMMARY

- Life expectancy is among the most common measures for assessing health outcomes. In 2008, the high-income nations had life expectancy at birth of 80 years; in Latin America, 73 years; in South Asia, 64 years; and in sub-Saharan Africa, only 52 years. Most of the differences can be explained by the much higher rates of mortality during the first 5 years of life in poorer regions.
- Since the 1960s, the gap in life expectancy between nations, unlike the gap in per capita incomes, has fallen rapidly. Many low- and middle-income nations have made substantial and historically unprecedented progress in increasing life expectancy. This stands as one of the great successes of human development in the past half century although there are exceptions, including several nations in sub-Saharan Africa.
- In many developing nations, where life expectancy has risen and fertility fallen, the age structure of the population has changed and so has the pattern of disease. This has resulted in an epidemiologic transition. Heart disease, a noncommunicable disease, is now the leading cause of death worldwide. Diseases common in high-income nations, including heart disease and cancers, are becoming more prevalent in low- and middle-income nations. But infectious diseases, which take a relatively small toll in high-income nations, remain a major threat to children and adults in developing countries.
- Improving health and rising levels of per capita income are well correlated. But significant improvements in health can occur even at low incomes. Preventative measures long practiced in developed countries are vital. They include access to clean water and proper sanitation, control of insects and other disease vectors, and widespread vaccination programs. In addition to these public health measures, education, especially of females, is correlated with better health outcomes.

- Better health helps children remain at school and makes workers more productive in their fields and at their jobs. Better health increases the opportunity nations and individuals face to save and invest in their futures. Higher incomes, in turn, permit governments and families to devote more resources, whether for water, sanitation, vaccines, drugs, or health workers, to improving health.
- Infectious disease continues to plague poor nations. HIV/AIDS, malaria, and tuberculosis are three of the best known. Together they kill 4.5 million people a year, accounting for about 8 percent of all deaths worldwide. Other infectious diseases kill millions more, especially children, and debilitate those who are sick but do not die. Many of these infectious diseases, including HIV/AIDS, malaria, and tuberculosis, are preventable and with adequate resources and institutions can be treated effectively. That this is not happening is both an economic failure and a human tragedy.
- The spread of HIV/AIDS, the failure to attack malaria, and the reemergence of TB, including new drug-resistant strains, provide evidence of what has gone wrong in addressing world health. But there are also abundant examples of health successes, ranging from the eradication of smallpox to the near eradication of polio to the diffusion of oral rehydration therapies as a means of saving children from diarrheal diseases.
- The challenge of the twenty-first century will be for low- and middle-income countries to win the battle against both old and new infectious diseases, while addressing the increasing prevalence of chronic noncommunicable diseases long associated with higher incomes.

PART THREE

Macroeconomic Policies for Development



Investment and Savings

In Chapter 4 we learned that economic growth occurs because of increases in a country's capital stock and in its employed labor force together with increases in the productivity of those two factors of production. In this section of the book we go in depth into the mechanisms that lead to increases in the capital stock and to the increases in productivity that are embodied in capital. Capital increases because individuals and companies decide to invest their funds in the purchase of machinery, buildings, and other capital equipment. If they invest wisely, their production and profits increase. If they invest poorly the capital stock and profits may not increase at all. A farmer in Kenya will plant coffee trees and expect to reap a return when the coffee beans are ripe several years in the future. A large corporation in China will invest in furnaces and rolling mills to produce finished steel products.

To invest, however, the farmer or corporation must first obtain the funds and other resources that make their investments possible. The Kenyan farmer must have money to buy coffee seedlings and to hire labor to help in the preparation of the land. The Chinese steel corporation must find the funds to pay workers to build the furnaces and rolling mills and to purchase the best steel-rolling technology from abroad or from another corporation in China. The funds required may come from savings from the farmer's own income, may be borrowed from members of his or her extended family, or may be a loan from a microcredit institution. The large corporation may sell ownership shares in the company, use its own retained earnings, or borrow from a major bank. All of these sources of funds require that some person or business first save at least part of their income. If all income is consumed for food by the farmer or spent on an expensive limousine for the steel plant manager, there will

be nothing left over to invest. Both the farmer and the steel corporation, however, can borrow from someone else who has saved and use the money to invest, paying it back at some future time.

In this chapter, we start first with the fundamentals of what constitutes good investments—what leads to investments that promote growth either of the individual farmer, the large corporation, or the country as a whole? The principles of good investments from a country's point of view apply equally to investments made by individuals, by private companies, or by governments. However, differences in ownership, in who is making the investment decisions, have a powerful influence on whether the investor applies the principles of good investment correctly. Government officials do not have the same objectives as private companies, and large corporations can pursue goals quite different from a small farmer or proprietor of a five person factory. Foreign investors may have different incentives and goals from domestic investors. As already noted, no one can invest either well or poorly unless someone or some company or government agency first saves. We will briefly explore why households, corporations, and governments save. Of particular interest is that some of the saving done for many developing countries is done by foreigners, not just by citizens of the country itself.

In some cases investors save and invest their own money, but more commonly those who invest are different from those who save. The gap between the savers and investors is filled by a wide variety of financial institutions ranging from banks and the stock market to the government itself. This separation of investors and savers makes possible much higher rates of investment than would otherwise be the case, but it also depends on institutions and government policies that can behave in ways that lead to economic crises, as occurred throughout much of the world in 2008–09. There is, of course, much more to the role of the government in managing the economy than just what it can accomplish through a central bank. Government is both an investor itself and a spender of funds in a wide variety of areas, ranging from income transfers to the poor to paying the salaries of civil servants and the military, that have little to do with investment. In Chapter 11 we explore how the government raises the revenue it needs to make these expenditures and how it uses these resources to help cope with economic crises as they occur.

Chapter 12 deals with the role of the financial sector first in its role as the intermediary that mobilizes savings from individuals and then allocates savings to investors. Efficient management of the financial sector, however, can be undermined by inflation and thus control of inflation is a critical component of any effective growth strategy. The chapter reviews how developing countries have tried to control inflation and avoid financial crises and how their success or failure in these efforts has fostered or undermined the development of the intermediation role of the sector. Chapter 13 explores in greater depth the central role of developing country debt in triggering many financial crises. Developing country debt that leads to crises also involves foreign savings, both private and government, that is then lent to developing countries by international

banks and other financial institutions including multilateral agencies, such as the International Monetary Fund (IMF) and World Bank. Chapter 14 focuses on the role of foreign aid to developing countries (a major form of foreign savings made by aid-giving governments that is transferred to developing countries).

USING INVESTMENT PRODUCTIVELY: COST–BENEFIT ANALYSIS

In many low-income countries, there appear to be abundant opportunities for increased investment that could have significant economic and social benefits, but these investments do not happen. Roads are absent or of poor quality, electricity and telecommunications facilities are in short supply and unreliable, and there seem to be many workers with few factories to employ them. Sometimes, this is because investment funds are not always available, but it is also because the investments that do take place are not always particularly productive. But this in itself is a puzzle: The Solow growth model discussed in Chapter 4 suggests that, in low-income countries where capital is scarce, new investment has the *potential* to be highly productive but that potential is not always realized. Questions about the quantity and productivity of investment become more complicated when foreigners make the investments. Although foreign capital appears to be a ready source of additional investment funds, foreign investment does not always have a good reputation, and in some cases, even where it is productive, it is not always welcomed.

How does one judge whether an investment is likely to be productive or not? One of the most powerful tools for answering this question is **cost–benefit analysis**, which is also referred to as **project appraisal**. This tool is applicable to both private and public investment and to investment made by foreign firms as well. Public or government investors, however, should use the tool somewhat differently from private investors because their objectives are different.

PRESENT VALUE

The investor's calculation normally involves four steps. First, the firm will forecast the **net cash flow** of an investment, measuring the difference between the cash revenues from the sale of the product and the cash outlays on investment, material inputs, salaries and wages, purchased services, and other items.

Second, investors have to adjust their valuation of costs and benefits, depending on whether they happen soon or far into the future. Investors recognize that cash received in the future is less valuable than cash received immediately because in the interim they could earn interest (or profits) on these funds by investing them in

bonds or savings accounts (or in additional, revenue-earning production facilities). Any firm or individual asked to choose between $1,000 today or $1,000 next year would take the money now and place it in a savings account earning, say, 2 percent a year. Then, after one year, the interest payment would boost the savings account balance to $1,020. Equivalently, the prospect of $1,000 a year from now should be evaluated as equivalent today to only $1,000/1.02 = $980. This process of reducing the value of future flows is called **discounting**. Because the interest rate provides a first approximation of the rate at which future cash flows must be discounted, it is usually called the **discount rate**. Discounting takes place over several years, so the discount rate must be compounded over time. In the second year another 2 percent would be earned on the balance of $1,020, and increase it to $1,040.40. The payment of $1,000 two years from now then would be discounted to yield a present value of only $1,000/1.040 = $961. A general expression for the present value, *P*, is

$$P = F/(1 + i)^t \qquad [10\text{–}1]$$

where *F* is the value in the future, *i* is the discount rate, and *t* is time, usually represented by the number of years. As the interest rate or the number of years increases, the present value of the future cash flow decreases. Note that because project appraisals are most conveniently done at **constant prices**, netting out inflation, the discount rate is a **real rate of interest**, net of inflation.

Investment projects usually have negative cash flows in the early years, when there are large expenditures, then generate positive cash flows in later years, as the new facilities begin to earn revenues. Tree crops such as palm oil and rubber take several years before they become productive and must be cut down and replanted when production falls off because of aging. Such a **time profile** of net cash flow is depicted in Figure 10–1; it is the most common of several possible profiles. To summarize the value of this net cash flow in a single number, each year's net cash flow is divided by the respective discount factor, and the resulting present values are added to give the **net present value (NPV)**:

$$NPV = \sum (B_t - C_t)/(1 + i)^t \qquad [10\text{–}2]$$

where B_t and C_t are the benefits (revenue) and costs in each year *t*, and *i* is the discount rate. For a firm, the correct discount rate is the average cost at which additional funds may be obtained from all sources (that is, the firm's cost of capital).

If the NPV of a project happens to equal 0, the project will yield a net cash flow just large enough to cover all the projects costs. In that case, when NPV = 0, the discount rate has a special name, the internal rate of return (IRR), a term we introduced in the discussion of education as an investment in Chapter 8. If the net present value is positive, then the project can cover all its financial costs with some profit left over for the firm. If negative, the project cannot cover its financial costs and should not be undertaken. The higher the net present value, the better the project's financial

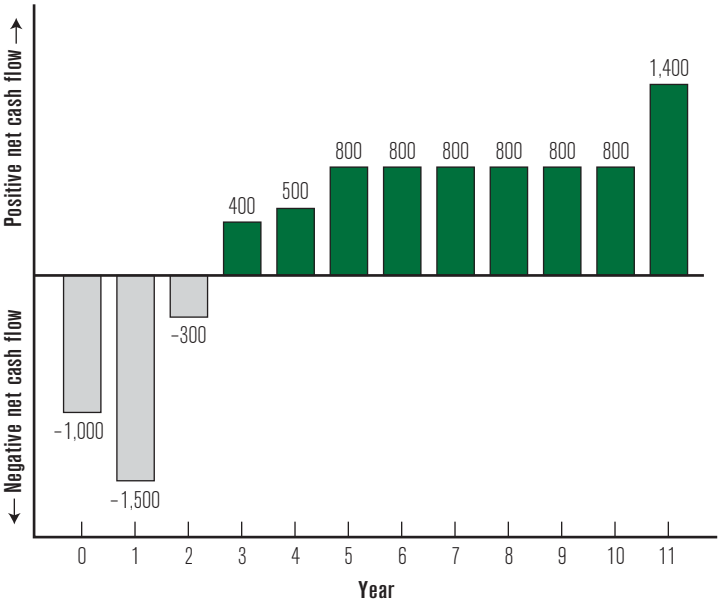


FIGURE 10–1 Time Profile for Investment: Net Cash Flow

The cash flow of a project can be represented by a bar diagram. Cash outflows are shown by bars below the horizontal axis; inflows, by bars above the axis. Year 0 and 1 show investment in construction and equipment, hence the negative cash flows. Year 2 is the startup period, years 3 and 4 show gradually increasing output and cash inflows, and years 5 through 10 show steady output and cash inflows. The project is assumed to end in year 11, when the salvage value of equipment swells the cash inflow.

profile. If the cash flow of Figure 10–1 is discounted at a rate of 8 percent,[1] the net present value is a positive $1,358 (Table 10–1). This calculation indicates that the investment will earn enough to repay the total costs ($2,500 over years 0 and 1) with a (discounted) surplus of $1,358.

Third, investors must consider the risks to both the costs and the benefits of their investments. Simple NPV analysis such as that in Table 10–1 can be misleading because it generates a single summary number for the answer, giving a false sense of precision. In reality, investors cannot be sure in advance of all the costs or the cash flows their investment might generate. Fuel prices might skyrocket or electricity supplies might be cut off, forcing the investor either to buy a new generator or to curtail production. The prices for their products might be either much higher or much lower than the original projections. Political instability might increase costs and disrupt

[1]The return on investment in developing countries can be much higher and much lower than 8 percent, depending on circumstances. We have used 8 percent in our example because it is a common rate of return for such investments.

TABLE 10-1 Net Present Value

YEAR	CASH FLOW FROM FIGURE 10-1 (DOLLARS)	DISCOUNT FACTOR AT 8%*	PRESENT VALUE† (DOLLARS)
0	−1,000	1	−1,000
1	−1,500	0.926	−1,389
2	−300	0.857	−257
3	400	0.794	318
4	500	0.735	368
5	800	0.681	544
6	800	0.63	504
7	800	0.583	467
8	800	0.54	432
9	800	0.5	400
10	800	0.463	371
11	1,400	0.429	600
Net Present Value‡			1,358

*Calculated as $1/(1 + i)^t$ from equation 10–2.
†Calculated as the cash flow times the discount factor.
‡Sum of the present value of each year.

the firm's ability to purchase supplies or sell its products. Therefore, prudent investors calculate a range of NPVs for their investment based on best case and worse case scenarios, testing the sensitivity of the outcome to changes in specific costs or benefits. An investor may decide not to make an investment even if it generates a positive NPV under baseline calculations if the risks around actually achieving that outcome are too great.

Fourth, investors must compare across projects. Just because a project generates a positive NPV does not mean it is the best possible use of investment funds, especially if the firm has a set of alternative projects to consider and a budget that can accommodate several but not all of them. Under these circumstances, investors should select the set of projects that will yield the highest total net present value for the entire investment budget.

Up to this point the methodology is the same for both private and public investors. The next steps calculate the costs and benefits from an economywide perspective. Private investors are typically only interested in whether a project provides an adequate rate of return or profit to themselves. Governments, in contrast, should be concerned with whether a project contributes to the growth of the economy as a whole or to social welfare more broadly. This concern of governments applies both to public investment undertaken by the government itself and to the impact of private investments. Governments are often called on to support private investments in various ways, and their decisions in this regard need to be based on the economy-wide impact of the project, not just on whether the private investors will make a profit or not.

OPPORTUNITY COSTS

Every investment uses goods and services that could otherwise be used for other purposes. In evaluating the investment from an economywide perspective, these goods and services should be valued in terms of the net benefits they would have provided if used in some alternative project—that is, they have **opportunity costs**. An investment in a dam requires saving that otherwise could be invested in a rural road or in elementary schools. Cotton used in a textile factory otherwise could have been exported and sold abroad. The labor used to build a road might have otherwise been used to sink water wells or grow millet. For a private firm, the cost is the market value that it pays for wages, goods, and services. For society at large, the cost is the value of the resources in terms of their next best use. Public sector investments, in particular, should be evaluated using opportunity costs rather than market costs where the two differ.

The opportunity cost to society might differ from the market wage rate if the government pays wages higher than the market clearing value. If 10 farm workers take jobs building a new highway, their opportunity cost is the value of the lost agricultural output due to their leaving the farm, net of the nonlabor recurrent costs of producing that output. This reduction in net output is the value of the **marginal revenue product** and, in this situation, is equal to the opportunity cost of labor.[2] Similarly, if the highway project means that savings will be drawn away from other projects that on average would have earned a return of 5 percent, then the opportunity cost of capital is 5 percent, and this should be used as the discount rate in evaluating the project, even if the government borrows money at 2 percent.

In some countries, the market price of foreign exchange might differ from the opportunity cost, especially in countries where governments manipulate or control the exchange rate. Cotton used in a textile mill otherwise might have been exported; if so, its opportunity cost would be the foreign currency it would have earned as an export. If the cloth produced by the mill would have been imported in the absence of the project, its opportunity cost (a benefit in this case) would be the foreign exchange that otherwise would have been spent on cloth imports.

SHADOW PRICES

The opportunity costs of goods and services for the economy as a whole are called **shadow prices** or **economic opportunity costs**. The first approximation of a shadow price (for land, labor, capital, and foreign exchange) is the price paid by private participants in the market. Interference in the market distorts market prices from their economic opportunity cost: taxes and subsidies of all kinds, monopoly power, minimum wages, interest rate controls, tariffs and import quotas, price controls, and

[2]The value of the marginal product of a factor of production can be calculated as the price of a commodity multiplied by the additional physical output that results when one unit of the factor is added to the production process with all other factors held constant (e.g., that factor's marginal product).

so forth. When a private firm undertakes investment analysis, it conducts a **commercial project appraisal** based on market prices and focusing on profitability. When governments consider public investments, they want to measure the full economic impact of the investment on the country. To do so, they must use shadow prices when market prices do not fully reflect true scarcity values. This kind of evaluation is called **economic project appraisal**. To estimate shadow prices, market prices must be adjusted to reflect full opportunity costs.

Estimating shadow prices is not easy and requires intimate knowledge of the workings of an economy, both its macroeconomic relationships and the microeconomic behavior of its factor markets. But such adjustment can be important. In many countries, the shadow foreign exchange rate is higher than the official rate in terms of local currency per dollar. In most cases, this reflects the impact of import duties and quotas or intervention by central banks in establishing or modifying the exchange rate. As a consequence, any export project that earns more foreign exchange than it uses or any import-substituting project that saves more than it uses gets a boost from the shadow exchange rate. Although the salaries and wages of skilled employees usually require no adjustment from market to shadow prices, frequently the opportunity cost of unskilled workers is lower than the wage in formal urban labor markets. Therefore, any project using unskilled labor, especially if it is located in a rural area, gets a boost because the shadow wage reduces costs without changing benefits.

Discount rates might also be adjusted from market interest rates to shadow discount rates if a government treats capital as a very scarce factor of production or if it has a strong preference for projects that realize their benefits relatively quickly. This approach would tend to discourage any public investment project with high initial investment costs, long gestation periods, and low net cash flows. A higher shadow discount rate would favor projects that generate their net benefits early because these can be reinvested in other productive projects for continued growth and projects that use relatively abundant factors of production, especially labor, instead of scarce ones, like capital.

An illustration of the power of shadow pricing is contained in Table 10–2, which depicts two projects with identical cash flows. One project (the textile mill) earns more foreign exchange but uses more labor. Because the shadow wage rate is below the market rate, the economic NPV of both projects is raised, but the more labor-intensive textile project benefits more (shown in Part 2 of the table). When the shadow exchange rate is applied, the NPV of the textile mill is raised considerably, and that of the telecommunications system becomes negative (Part 3).

WELFARE WEIGHTS

Under some circumstances, governments want to incorporate broader social goals into the evaluation of public sector projects, such as the impact on income distribution, poverty alleviation, or environmental degradation. To do so, governments can use welfare weights to further adjust shadow prices in a project analysis. This process

TABLE 10–2 Effects of Shadow Pricing on Cost–Benefit Analysis

PROJECT	INVESTMENT (FIRST YEAR)	NET ANNUAL CASH FLOW (NEXT 5 YEARS)	NET PRESENT VALUE (10%)
1. Take two projects with identical cash flows. Project A earns more net foreign exchange and uses more labor than project B:			
A. Textile mill, of which	−1,000	+300	+137
Net foreign exchanged earned	−500	+400	
Wages paid	−350	−100	
B. Telecommunications system, of which	−1,000	+300	+137
Net foreign exchange earned	−800	0	
Wages paid	−100	−50	
2. The shadow wage is 75% of market wage, so all wage costs are reduced by 25%. This results in the following net cash flows:			
A. Textile mill	−913	+325	+319
B. Telecommunications system	−975	+313	+212
3. The shadow exchange rate is 20% above official rate, so the net foreign exchange flow is raised by 20%. This results in the following net cash flows:			
A. Textile mill	−1,100	+380	+340
B. Telecommunication system	−1,160	+300	−23

might place a higher value on net additional income to certain target groups, such as the very poor. Then projects generating more income for these groups have higher NPVs than otherwise and tend to be selected more frequently. Similarly, it might adjust benefits to capture environmental goals or further reduce the shadow wage rate if it was particularly interested in creating new jobs. When welfare weights are introduced or shadow prices are further adjusted to reflect social goals, the process is called **social project appraisal**.

The method is potentially powerful but has its dangers. Welfare weights are arbitrary, subject to policy makers' or politicians' judgment. This, in itself, is not bad, but these weights can so overwhelm economically based shadow prices that project selection comes down to a choice based almost entirely on arbitrary weights. This gives a false sense of precision and sometimes can be misused. In most cases, it is best to do the calculations without these weights and then look separately at the welfare implications of the project.

BARRIERS TO PRODUCTIVE PUBLIC AND PRIVATE INVESTMENT

In Chapter 3 we explored at a macro level many of the variables that influence the rate of growth of gross domestic product (GDP) in a country. Many of these same variables have an influence on the productivity of investment, both public and private.

Political instability can lower the rate of return and increase the risk, particularly to private investors but to government investors as well. Macroeconomic instability can distort investment decisions away from investments that promote long-term growth and toward investments that protect the investor from inflation. Excessive investment in high-income housing and office buildings often characterized the high inflation period that characterized some South American countries in the 1950s through the 1980s. Investment by governments and the private sector in education and health increase the quality of the labor force and that in turn raises the rate of return on investment in projects other than the investments in education and health themselves. More open economies provide opportunities for high return investments not only for exporters but also for any investment that needs to use imported inputs. In closed economies imported inputs are often very expensive if they are available at all.

Our focus here, however, is on what in the macroeconomic analysis in Chapter 3 is referred to as effective governance and institutions. There are a wide range of elements that are included under the rubric of good governance and effective institutions. We focus here at the micro level on two: the extent and nature of government intervention in the economy and on corruption. The two are closely related. As we have indicated in many chapters in this book, government regulation is often necessary to correct for market failures that could do damage to the environment, to health, or to growth. In all countries, however, government intervention can create opportunities for private investors and public officials alike to seek what economists call rents. **Rent-seeking** behavior refers to activities by private and public officials to extract value from a project without providing any compensation in return or contributing any increased productivity to the project. An example would be for an investor in a cement plant to negotiate with the government to provide high tariffs on imported cement so that the investor could raise the domestic price on cement, thus earning a higher level of profit. The cement producer in this case has done nothing to raise the quality of the cement or to increase the efficiency with which the cement is produced. The increase in profits is due solely to this investor's ability to persuade the government to raise the cement tariff and hence the domestic cement price. Purchasers of the cement then must pay more money to the producer without getting any increased value for themselves. Not only can rent-seeking behavior lead to poor policy choices, but the rent-seeking behavior itself entails the diversion of potentially substantial resources from productive uses.

The relevance of these concepts to this chapter is that excessive or inappropriate kinds of regulation together with high levels of corruption are a major reason public investment projects in low- and middle-income countries often have a low rate of economywide return. They can also explain why many private investors also face a low rate of return on their investments despite trying to provide a badly needed product or service that under more normal circumstances would provide the investor with a substantial profit. The World Bank has compiled quantitative indicators of how government interventions affect the ease of doing business in individual countries. Transparency International, a nongovernmental organization (NGO), also

has carried out and collected surveys of perceptions of the level of corruption in most of these same countries. We take up each of these two efforts in turn.

BARRIERS TO DOING BUSINESS

Starting and running a business is often described as an individual coming up with a good idea (a new or better product, an old product in a new location, or a new way to produce a good), building a factory, hiring workers to produce the good, and creating a marketing office to sell it to customers. In most countries that individual must first get a license to open that business and in many countries it takes a large number of such licenses. In Malaysia, for example, in the early years of the twenty-first century, it took 70 licenses or permits to open a new hotel. The business must then hire labor, and typically there are many regulations governing the hiring of labor. Many of these regulations make good sense in the right context. Employers in many countries, including the United States, are not allowed to discriminate in hiring based on gender, race, or religion. More problematic are the rules in countries that make it extremely difficult to dismiss employees even if their performance is subpar or an economic downturn limits the work that is available. This new business, early on and throughout its existence, must enter into numerous contracts with suppliers, with clients, and with its own employees. The question then becomes who if anyone will enforce these contracts. In a high-income country it is typically the legal system and the judiciary that is responsible for deciding whether a contract should be enforced or not, and these decisions are made according to rules typically written into law by the legislature. In many low- and middle-income countries, however, the judiciary is poorly trained, has little power, and is often corrupt. Enforcement of contracts tends to depend more on the power and influence of the company seeking (or trying to avoid) enforcement.

The World Bank has measured the degree to which these many government interventions affect the ability to open a new business or carry on an existing one. They report on a number of government interventions that are widely believed to affect the ease of doing business in a country. For each of these indicators, they measure the time it took to get the necessary permits, the number of permits or procedures required, the cost of getting these permits, and a variety of other indicators in areas such as the employment of labor or obtaining credit. The full list of categories is as follows: Starting a Business, Dealing with Construction Permits, Employing Workers, Registering Property, Getting Credit, Protecting Investors, Paying Taxes, Trading across Borders, Enforcing Contracts, and Closing a Business.[3] Table 10–3 presents data for a number of economies on their overall ranking as a place where it is easy or difficult to do business.

[3]World Bank, *Doing Business 2010: Reforming through Difficult Times* (Washington, DC: World Bank, 2010).

TABLE 10-3 The Business Environment

COUNTRY	DOING BUSINESS RANK, 2011	CORRUPTION PERCEPTION RANK, 2011
Singapore	**1**	**5**
United States	**4**	**24**
Georgia	16	64
Thailand	17	80
Japan	**20**	**14**
Mauritius	23	46
Latvia	21	61
South Africa	35	64
Taiwan	**25**	**32**
Chile	39	22
Spain	**44**	**31**
Turkey	71	61
Jamaica	88	86
Pakistan	*105*	*134*
China	91	75
Vietnam	98	112
Kenya	109	154
Egypt	110	112
Greece	**100**	**80**
Bangladesh	*122*	*120*
Russia	120	143
Nigeria	133	143
Brazil	126	73
India	132	95
Iran	144	120
Philippines	136	129
Haiti	*174*	*175*
Afghanistan	*160*	*180*
Bolivia	153	118
Venezuela	177	172
Chad	*183*	*168*
Central African Republic	*182*	*154*

Bold, high-income economies; italics, low-income economies.

Sources: The World Bank, *Doing Business 2011*,4 (Washington, DC: The World Bank) available at www.doingbusiness.org/Rankings. Transparency International, *Corruption Perception Index 2011* (Berlin: Transparency International) available at www.transparency.org/policy_research/surveys_indices/cpi/2011.

Virtually all the high-income countries are ranked among the top 30 (out of 183) in the overall ease of doing business, with Singapore occupying the top spot. Six middle-income nations also made it into the top 30 in 2010: Georgia, Thailand, Mauritius, Malaysia, Lithuania, and Latvia. We find some poor performers among the high-income nations: Taiwan, Spain, Italy, Brunei, and Greece. Most middle-income countries rank in the top half of this index but there are notable exceptions, such as Russia, Brazil, India, and Iran. The bottom 20 countries are 17 sub-Saharan African countries, some of them plagued by war (Chad, Congo), plus Timor-Leste, Laos, and Venezuela.

Most of these countries are among the poorest in the world, although Venezuela is a major oil exporter. The Central African Republic ranked as the most difficult nation in the world with which to do business in 2010, according to this index.

The reasons for the differences in these rankings between the highly and lowly ranked can be best understood by looking at the specific categories measured. In Singapore, only 3 procedures were required to start a business, and it took only 3 days to complete those procedures, there were no restrictions on firing unwanted employees (rated 0 out of 100, with 0 indicating no difficulty), and the fired employee was entitled to only 4 weeks of severance pay. Registering property involved 3 procedures and took a mere 5 days. In Kenya, a country ranked close to the middle on ease of doing business, 12 procedures were required to start a business and it took 34 days to complete them; the difficulty of firing workers was rated 30 out of 100 and 47 *weeks* of severance pay was required. In addition, there were 8 procedures to register property, which took 64 days to complete. In the bottom quarter was Bolivia, where the 15 procedures involved in starting a business took 50 days; the difficulty of firing an employee was rated 100 out of 100, meaning you could not fire an employee so severance pay was not relevant. Registering property in Bolivia involved 8 procedures that took 2 months to complete. For these and other countries, the money cost of completing all the necessary procedures was a relatively low share of per capita income in high-ranked countries and a much larger share of per capita income in low ranked countries.

Time is also money for an investor, thus the more procedures that exist and the longer they take, the less likely the investor will go ahead with the investment. The result is that in countries with an unfriendly business climate, investments are lower, the costs of doing business are higher, and the rate of return on investments is lower. All of this translates at the macro level into a lower rate of growth.

Excessive regulation that creates any number of obstacles to doing business is not just a matter of lost time or the direct legal costs of the procedures. Large numbers of procedures involving licenses and permits also create broad opportunities for rent seeking and outright **corruption**. The difference between corruption and rent seeking is that corruption involves rent seeking that is illegal, whereas many forms of rent seeking are perfectly legal even though they damage the productivity of the economy. Large numbers of regulatory procedures are not the only cause of widespread corruption. Construction, especially public construction, often involves large elements of rent seeking and corruption. Road construction and the building of government offices involve costs that are difficult to measure for someone not involved directly in the business, making it easier for a construction contract to include hidden costs needed to bribe politicians to provide the necessary project funding and to pay off those in charge of granting construction permits.

Data on perceptions of corruption in countries around the world estimated by Transparency International, a non-profit organization based in Berlin but with a global network, are presented in Table 10–3. Even a casual comparison of the two

columns in the table reveals the rough correlation between the ease of doing business and the perception of corruption. Countries with friendly business procedures typically have low levels of corruption, and the reverse is equally true.

Complete elimination of corruption from public construction and from government business relations is virtually impossible. Even the wealthiest countries regularly discover that an official has taken a bribe to issue a permit and a politician has received illegal payments so that a vote to fund a public investment project would go forward. It is often assumed that the solution to the problem of corruption is to increase the number of police investigating corruption and raising the penalties for being caught. Certainly law enforcement has a role in curbing corruption, but law enforcement alone cannot accomplish much if the opportunities for corruption are pervasive (or if law enforcement is itself corrupt). In China, the death penalty has frequently been applied to cases of corruption, but there is little evidence that corruption has declined as a result.

Key elements in the fight against corruption are similar to what it takes to improve the climate for doing business. The number of regulations should be kept to the minimum necessary to control fundamental violations of public health, the environment, or other important social goals. The officials administering those regulations should have as little discretion as possible on whether to grant a permit. The more discretion they have, the more opportunity they have to delay and obfuscate the procedures until they receive illegal payments to do what they should be doing without payment. The process should be fully transparent with clear deadlines for the officials making the decisions. Putting the whole process online where all applicants can go to find out the status of their application has proved to be a powerful tool in countries where computer literacy is the norm. For public construction contracts the key is to have competitive bidding with controls to ensure that the companies bidding do not collude with each other or with the official contract awarding agency. If all of these steps are taken, the amount of corruption that remains will be less, and the anticorruption law enforcement agencies will not be overwhelmed and will be able to root out much of the corruption that remains. That, in essence, is the Singapore story, and Singapore is ranked 3 on Transparency International's 2009 Corruption Perception Index. Central African Republic is ranked 158.

Overcoming corruption becomes almost impossible if the corruption starts not at lower levels of government, such as in the customs office where those issuing import permits work, but instead at the highest levels of government, including those closest to the president or prime minister of the country. The motive for high-level corruption is partly personal greed, as the family of Ferdinand Marcos in the Philippines demonstrated, but it is also a result of the difficulties most countries have in finding legal and transparent ways of funding the political organization needed to win elections or hold on to power in more authoritarian situations. Raising money for a political organization is easy if the politician can offer favored access to government decisions in exchange. Two presidents of the Republic of Korea and the son of

another were convicted of receiving hundreds of millions of dollars in this way. More than a few low-income countries have a serious problem in this regard, and even the American political process has yet to find a way to fund elections in a way in which buying political influence is minimal.

FOREIGN DIRECT INVESTMENT

The vast majority of investment in developing countries is undertaken by local citizens, either through small individual projects, larger investments by local firms, or by the government. But in some countries, an increasingly larger share of investment is made by foreign individuals or firms. Since 1990, **foreign direct investment (FDI)** has been the largest of all international capital flows to middle-income developing countries, larger than either cross-border bank lending or foreign aid (in low-income countries, FDI is smaller). FDI is defined as a long-term investment in which a nonresident entity exerts significant management control (usually at least 10 percent of voting stock) over an enterprise in the host country. A second type of cross-border investment is **portfolio equity**, in which an investor takes a smaller stake in an enterprise, either through a direct purchase or through a stock exchange. Portfolio equity flows to developing countries, starting from a low base, grew rapidly in the early 1990s, but then fell sharply following the financial crises of 1997 and 1998. A similar pattern occurred in the last decade. Portfolio equity increased 10-fold between 2000 and 2007, only to turn *negative* in 2008 as investors withdrew their investments. With the end of the most recent financial crisis such flows have resumed, and by the end of 2009 were approaching 2007 levels. Even at their peak, however, portfolio flows were still only one-quarter the size of FDI and about half the size of long-term international commercial bank loans.

In the 1960s and 1970s, many developing countries were suspicious of FDI and often took steps to actively discourage it. At the time, because of the colonial history in many countries and the sometimes offensive behavior of certain foreign investors in taking advantage of weak political and legal systems to gain monopoly rights, make huge profits, and influence domestic politics, this suspicion often was well founded. Starting in the mid-1980s, however, these attitudes began to change, and developing countries increasingly sought to attract FDI as a means of financing investment, creating jobs, and importing technology and ideas. FDI to developing countries grew rapidly from $22 billion in 1990 to $148 billion in 2000 to almost $600 billion in 2008 (more than half of all foreign capital flows). The sharp increase was due partly to worldwide advances in technology in communications and transportation (and the accompanying reduction in costs) and was closely related to the rapid expansion in world trade during the period. It was also the result of much greater enthusiasm on

the part of developing countries in trying to attract FDI. The financial crisis reduced these flows by 2009, but they still were more than double the level of 2000.[4]

FDI continues to generate extensive controversy on a range of issues, including repatriation of profits, loss of control over domestic natural resources to foreign-owned entities, the extent of the transfer of technology, the impact on tax revenues, and the content of incentive packages. For some, foreign ownership of domestic companies should always be discouraged, even if it brings economic benefits. For others, FDI is a critical avenue to integrate the domestic economy with global markets that brings with it new technology and skills. The evidence on the relationship between FDI and development suggests that it is hard to make broad generalizations, and the impact depends critically on the purpose and type of the investment as well as policies and institutions in the recipient country. FDI in firms producing manufactured products sold on competitive global markets is more likely to be positively related to economic growth and other development outcomes, whereas FDI in firms producing primary products or manufactures or services for protected domestic markets tends to have more mixed outcomes.

FDI PATTERNS AND PRODUCTS

The majority of direct foreign investment in developing countries is undertaken by **multinational corporations (MNCs)**. A multinational corporation is a firm that controls assets of enterprises located in countries other than its home country in which foreign operations are central to its profitability. Multinational enterprises are quite diverse, coming in all sizes, from all regions of the world, and are engaged in a wide variety of activities. The majority of MNCs are from industrialized countries, but a growing number are based in developing countries. Six developing countries had companies in *Fortune* magazine's list of the world's largest 500 firms in 2010, including China (46 entries), India (8), and Brazil (7). Not all of these firms operate outside of their own nations, but most of them do. Many MNCs are private corporations, but not all: *Fortune*'s top 500 includes a number of giant, state-owned companies producing petroleum, such as China National Petroleum (ranked 10 in the world) and Mexico's PEMEX (64).[5] But multinationals are not always large; small companies, especially in East and Southeast Asia, have been investing overseas for many years, particularly producing labor-intensive manufactured products such as textiles, shoes, toys, and electronics.

It is no surprise that the vast majority of FDI comes from rich-country investors. Globally, most multinational investment is also directed toward other wealthy countries, but the share aimed at developing countries has been rising. In the late 1980s, less than one-fourth of global FDI flows were directed toward developing countries.

[4]World Bank, *Global Development Finance, 2011,* 36 (Washington, DC: World Bank, 2010).
[5]Walmart Stores, Inc. is the largest corporation in the world.

The developing country share of global FDI flows increased rapidly during the early 1990s, peaking at 40 percent in 1994 and averaging 31 percent over that decade. By 2000, the developing country share had fallen below 20 percent; yet, that share grew throughout the first decade of the 21st century, reaching 46 percent by 2010.[6] Then, in the wake of the financial crisis of 2008–09, total FDI flows fell steeply, but the amounts going to low- and middle-income economies stayed relatively stable. However, FDI (as well as portfolio equity) is highly concentrated in a few developing countries. By 2010, half of all FDI in developing countries went to just 10 countries. China (including Hong Kong), received one-quarter of the total at the same time as China was becoming a major investor in Africa and elsewhere. Russia is another large recipient of FDI, followed by Brazil and India. By contrast, the UN's category of least-developed nations, making up much of sub-Saharan Africa, received just 5 percent of the total.[7]

FDI is aimed at a very wide range of activities that are sometimes difficult to classify, but three broad categories stand out:

- Natural resource-based activities, such as petroleum, minerals, and agricultural production. Firms engaged in these activities tend to be quite large and the investments capital intensive. These investments are usually negotiated directly with the host government, and the government is often a partner in the investment.
- Manufacturing and services aimed at the domestic market in the host country, including consumer goods (such as processed food or apparel); capital-intensive products such as steel or chemicals; and a range of services such as transportation, communication, finance, electricity, business services, and retail trade. In many cases, these activities are protected at least partially against competition from imports through tariffs or other restrictions.
- Labor-intensive manufacturing aimed for export on world markets, including apparel, electronics, food processing, footwear, textiles, and toys. Firms in these activities tend to be efficient and competitive, but they also can move quickly from one country to another in response to changes in production costs or macroeconomic or political instability.

BENEFITS AND DRAWBACKS OF FDI

Viewed most narrowly, FDI is a **source of capital** that adds to total investment. FDI is relatively small compared to domestic investment, but its share has grown rapidly in the last three decades. It accounted for between 2.2 and 3.6 percent of GDP in all

[6]United National Conference on Trade and Development (UNCTAD) on-line data, UNCTADSTAT http://unctadstat.unctad.org/ReportFolders/reportFolders.aspx, accessed March 23, 2012.

[7]United Nations Conference on Trade and Development, *World Investment Report 2010*, annex table A-1.

developing countries throughout the first decade of the twenty-first century compared to close to 1 percent in the early 1990s and less than 1 percent in the 1980s. In terms of investment, FDI amounted to about 12 percent of gross capital formation during the past decade,[8] but its importance varies widely. In some countries, FDI can be equivalent to well over 10 percent of GDP in some years, especially in countries with oil or other natural resources (for example, Kazakhstan). In other countries, FDI is less than 1 percent. By the end of 2010, Vietnam was averaging foreign direct investments equal to 9 percent of its GDP; Bangladesh attracted barely 1 percent.

Foreign capital in general tends to be more volatile than domestic capital, but FDI is usually more stable than other forms of private foreign capital. One key reason is that FDI is often attracted by either resource endowments or long-term fundamental economic strengths, factors that generally do not change as quickly as interest rates and exchange rates, which are stronger determinants of short-term bank loans and portfolio capital. Moreover, because FDI is characterized by significant management control in large, fixed-production facilities such as factories and mines, investors are less likely (and less able) to flee during economic downturns. During the 1997–98 Asian financial crises, FDI fell somewhat but much less than bank loans and portfolio capital, which dropped sharply. FDI in emerging market countries fell between 2008 and 2009.

Along with providing new capital, recipient countries hope that FDI will add to the demand for labor and **generate employment**. It is not surprising that because FDI is a relatively small source of investment in most developing countries, its contribution to employment also tends to be small, accounting for less than 5 percent of total employment in most developing countries. However, the impact on employment varies greatly depending on the activity. FDI focused on resource-based capital-intensive industries, such as mining or petroleum, creates relatively few jobs, whereas investments in labor-intensive manufacturing create more jobs. The impact of FDI on wages and working conditions is controversial, with some people charging that MNCs pay low wages and impose harsh working conditions. While conditions vary, and this may be true in some instances, on the whole, evidence suggests that MNCs pay higher average wages and have better average working conditions than do domestic firms.[9]

Another benefit of at least some kinds of FDI, especially FDI in manufactured exports, is that it can help increase **specialization** in production. When firms in developing countries are part of a global production process, they can focus on the particular activity in which they have a comparative advantage and produce most efficiently. Although domestically financed investment can also engage in very specialized production, FDI has the advantage of stronger links to other parts of a global supply chain. Firms in the automobile business need not make the entire car (and

[8]Calculated from World Bank, "World Development Indicators," http://databank.worldbank.org.

[9]Edward Graham, *Fighting the Wrong Enemy: Antiglobal Activists and Multinational Enterprises* (Washington, DC: Institute for International Economics, 2000), chap. 4.

if they did, they might not be able to compete on world markets). Instead, they can specialize in assembly, production of basic components (such as door parts or wiring), manufacturing of more sophisticated parts (for example, engines or transmissions), design, or other aspects of the production process. In today's production environment, it is difficult to classify an automobile as American, German, Japanese, or Korean because the parts are made all over the world. Moreover, the particular activity that makes sense in an economy can change over time as resource endowments, skills, and economic policies change. When Intel made its first investments in the electronics industry in Malaysia in the early 1970s, it focused on low-skill activities such as simple assembly. Over the years, as the Malaysian workforce gained new skills and more Malaysians became experienced managers, activity shifted toward production of basic parts (such as keyboards), then to more-advanced components (such as microprocessors), and later to testing and research. More recently, electronics production in the Philippines has followed a similar progression.

FDI can also bring with it **access to world markets**. Developing countries capable of producing at competitive costs often find it difficult to penetrate foreign markets. Many multinationals, particularly in natural resources, chemicals, and other heavy industries, are vertically integrated, oligopolistic firms, for which many transactions take place within the firm. Multinationals develop preferential access to customers by fashioning and adhering to long-term contracts in standardized products, such as petroleum, or by acquiring a reputation for delivering a specialized product of satisfactory quality on a reliable schedule, as in electronics and engineering. On their own, firms from a developing country often require years to overcome such marketing advantages, and an affiliation with an MNC can help accelerate the process.

Among the most important potential benefits from foreign investment is the **transfer of technology, skills, and ideas**. Because much of the world's research and development activity takes place within North America, Europe, and East Asia, firms from these areas are a potentially rich source of innovative products, machinery, manufacturing processes, marketing methods, quality control, and managerial approaches. MNCs potentially can bring these new ideas and technologies with them, helping reduce costs and increase productivity in the host country. FDI is most valuable when these kinds of benefits spread beyond the local affiliate of the MNC itself and help other local firms and enterprises, generating positive externalities called **spillovers**. For example, local firms competing against the MNC may observe and adopt the technology brought from abroad to try to improve their own productivity. A domestic furniture maker might learn about new machinery, a special technique to finish wood products, or a different way to organize production that could lower its costs and increase its competitiveness. Competition from the MNC can push rival firms to reduce their costs through more-efficient use of all inputs and by introducing new technologies and techniques, generating an indirect benefit to all firms that buy from the MNC and its competitors.

While this kind of "horizontal" spillover to competitors can be important, MNCs clearly have the incentive to limit technology transfer to their rivals. They have stronger incentives to encourage "vertical" spillovers to firms operating up and down the supply chain that either sell to or buy from the MNC. MNCs want their local suppliers to produce at the lowest possible cost, so it is in their interest to work with suppliers to increase the suppliers' productivity by introducing new methods to improve quality and reliability and reduce costs. These changes provide indirect benefits to other firms that buy from these suppliers.[10]

One important way these kinds of spillovers can occur is through the **training of workers and managers**. Most developing countries do not have many well-trained managers with experience in organizing and operating large industrial projects, such as those undertaken by multinational firms. MNCs have the incentive to train local managers and workers (both their own employees and their supplier's) and, in some circumstances, will send employees abroad to the parent company for training in production methods or management techniques. It is not uncommon for local managers to spend six months or more working in the parent company and then return to senior positions in the local affiliate. The immediate impact is to help increase productivity and profitability of the MNC. However, over time, these managers and workers may move to other firms or even start their own companies, bringing this new knowledge with them. When textile companies began to move from Korea and Taiwan to Indonesia in the 1980s, initially most senior managers were expatriates. Over time, more Indonesians were trained to fill those positions, and eventually some started their own companies, competing against the MNCs. When Mauritius first began to produce textiles and footwear in the early 1980s, virtually 100 percent of the export firms were foreign owned, but within 15 years, foreign firms controlled just 50 percent of total equity capital.[11]

However, just as FDI can create positive spillovers, it can create negative ones. MNCs might create air or water pollution or cause other environmental damage, generating a negative externality for the host country. (Of course, local investors also cause environmental damage, in some cases even worse damage.) While some MNCs have been guilty of causing extreme environmental damage, others bring with them cleaner and more efficient production techniques that help reduce environmental damage.

FDI in protected or inefficient activities can lead to net economic losses rather than gains by misallocating capital, labor, and other resources. If a government encourages an MNC to invest in a petrochemicals company that can be profitable

[10]See Howard Pack and Kamal Saggi, "Vertical Technology Transfer via International Outsourcing," *Journal of Development Economics* 65, no. 2 (2001), 389–415; Garrick Blalock and Paul Gertler, "Foreign Direct Investment and Externalities: The Case for Public Intervention," in Theodore Moran, Edward Graham, and Magnus Blomstrom, eds., *Does Foreign Direct Investment Promote Development?* (Washington, DC: Institute for International Economics and Center for Global Development, 2004).

[11]See Blalock and Gertler, "Foreign Direct Investment and Externalities."

only with government subsidies or regulations that limit competition (like high import tariffs on competing products), the costs to the country are likely to be higher than the benefits. Firms operating under protection tend to use more outdated technologies and are less likely to be able to compete eventually on world markets. The firm might boast large profits and appear to be thriving, but the costs to taxpayers and to downstream customers forced to pay higher prices might be even larger.

Perhaps the biggest concern about FDI is the loss of local control over business. MNCs can drive out local businesses, and even where this might be economically efficient it can be politically very unpopular. A large efficient MNC can drive out small local business that operate at higher costs, particularly in countries with high trade barriers and other regulations where local firms may be relatively weak. Also, MNCs are more likely to repatriate their profits abroad, although some reinvest locally (and many local businesses also send their profits abroad). Concerns are most acute for the largest MNCs, whose size and control over resources often match and sometimes outstrip that of the recipient country governments. Investment by a multinational corporation raises the specter of interference by, and dependence on, foreign economic powers beyond the control of the host country. In some cases valid social preferences over the control and distribution of income might outweigh economic efficiency arguments.

FDI AND GROWTH

Ultimately, broad statements about whether FDI is beneficial or harmful are difficult to make and can be misleading. Because FDI is so varied, much depends on the specific activity, the actions taken by the MNC and the government, and the reaction by local suppliers, competitors, and customers. Debate continues as to the existence and magnitude of spillovers, both positive and negative. Research examining the relationships between FDI and growth has reached mixed conclusions. Some research has found a positive relationship, particularly when the workforce in the recipient country has achieved a minimum level of education, whereas other studies have found no relationship or even a negative one.[12]

Clearer patterns begin to emerge when researchers take into account the purpose and context of the FDI. Foreign investment that produces for the domestic

[12]The research on FDI, spillovers, and growth is extensive. For an example of a study finding a positive relationship in the presence of a minimum level of education in the recipient country, see E. Borensztein, J. DeGregorio, and Jong-Wha Lee, "How Does Foreign Investment Affect Growth?" *Journal of International Economics* 45, no. 1 (1998), 115–72. For an influential study finding negative spillovers in Venezuela, see Brian Aitken and Ann E. Harrison, "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela," *American Economic Review* 89, no. 3 (June 1999), 605–18. For a recent survey, see Robert Lipsey and Fredrick Sjöholm, "The Impact of Inward FDI on Host Countries: Why Such Different Answers?" in Theodore Moran, Edward Graham, and Magnus Blomstrom, eds., *Does Foreign Direct Investment Promote Development?* (Washington, DC: Institute for International Economics and Center for Global Development, 2004).

market and relies heavily on subsidies or protection from competition is much less likely to be beneficial and may even generate economic losses for the host country. FDI in natural resource-based industries depends heavily on the impact of the industries themselves. As described in Chapter 18, while primary product exports have at times helped spur development in some countries (such as diamonds in Botswana and petroleum in Indonesia and Malaysia), in other countries, natural resource abundance has been as much as curse as a blessing, such as oil in Nigeria and Venezuela, copper in Zambia, and diamonds in the Democratic Republic of the Congo and Sierra Leone.

FDI aimed at firms producing manufactured exports and operating in competitive global markets most often is found to have a positive relationship with growth and development. This type of activity tends to be economically efficient and conducive to importation of new technologies, training new workers and managers, and positive spillovers to suppliers and even competitors. Georgetown University economist Theodore Moran finds significant differences in operating characteristics between subsidiaries that are integrated into the international sourcing networks of their parent multinationals and those that serve protected domestic markets. Where parent firms use local affiliates as a part of a strategy to remain competitive in international markets, they maintain those affiliates at the cutting edge of technology, management, and quality control. But Moran finds that FDI in protected industries can hinder an economy, as these firms typically use older technologies and operate under restrictions that raise costs, hinder exports, and reduce productivity gains.[13]

POLICIES TOWARD FOREIGN DIRECT INVESTMENT

Developing countries interested in attracting FDI and ensuring maximum benefit generally use some combination of three broad strategies:

- Improve the general environment for all kinds of investment (foreign or domestic) by strengthening infrastructure, reducing red tape, and improving the quality of labor.
- Introduce specific policies and incentives to attract FDI, such as export processing zones, worker training, import protection, or tax holidays.
- Impose requirements on MNCs (such as limits on equity holdings or repatriation of profits) in an attempt to capture more benefits locally.

IMPROVE THE GENERAL INVESTMENT ENVIRONMENT The least-controversial approach, and one that can bring benefits beyond those associated with FDI, is to take steps to improve the general environment for investment and business operations. As

[13]Theodore Moran, "How Does FDI Affect Host Country Development? Using Industry Case Studies to Make Reliable Generalizations," in Theodore Moran, Edward Graham, and Magnus Blomstrom, eds., *Does Foreign Direct Investment Promote Development?* (Washington, DC: Institute for International Economics and Center for Global Development, 2004).

discussed earlier, businesses in poor countries usually face much larger regulatory burdens and higher costs than those in rich countries. To try to improve the general investment environment, governments can improve road and ports infrastructure, invest in utilities to make them more reliable and less expensive, reduce trade tariffs, strengthen the judicial process and the rule of law, and reduce unnecessary regulations and red tape that add to the cost of doing business. These steps should help attract new investment, both domestic and foreign, and help make existing investments more productive and profitable.

INTRODUCE POLICIES AIMED SPECIFICALLY AT ATTRACTING FDI The question of whether governments should take specific measures to encourage FDI even more than domestic investment turns on views on the existence and size of spillovers: If FDI brings with it positive spillovers that generate benefits to firms other than the MNC, then policies and incentives aimed specifically at attracting FDI may be appropriate. Where there are few if any positive spillovers and possibly negative ones, the argument for government action and special treatment is much weaker.

Three kinds of specific policies are common. First, host governments can provide effective and timely **information** to potential investors that markets might not provide on their own. Investors might not come to a particular country if they simply do not know much about it. Many governments in developing countries have established investment promotion agencies to market their countries and provide information on business costs, port facilities, natural resources, levels of education, climate, and other factors that might make their country attractive to investors. In some countries, this kind of promotion can help, but in others it seems to make little difference.

Second, governments can undertake specific expenditures aimed at making their country more attractive for FDI. Many governments have established industrial parks or **export-processing zones (EPZs)** to increase investment and manufactured exports. The basic idea is that building infrastructure and improving the investment climate for an entire economy are likely to take a long time, so as an interim measure, the government could establish an enclave located near port facilities in which infrastructure is of high quality, utilities are reliable and of relatively low cost, and there are fewer regulations and less red tape. Typically, countries welcome both domestic and foreign investment in EPZs, but FDI tends to account for a greater share of the investment. Most of the countries that have been highly successful in producing manufactured exports have established EPZs or similar facilities, including EPZs in Malaysia, Korea, Mauritius, the Dominican Republic, and Indonesia; special economic zones in China; and maquiladoras in Mexico and similar stand-alone facilities in Tunisia and other countries. But EPZs and related facilities are not always successful and have failed when they have been poorly located, badly managed, or in other ways added to rather than reduced the producer's costs (see Chapter 19).

Third and more controversial, governments can provide specific incentives to MNCs to increase their profitability, including protection from import competition, subsidies, or tax breaks. Governments provide **protection** by introducing tariffs and quotas to reduce imports of competing goods or provide outright monopoly control over local markets. These steps are most relevant for FDI aimed at producing for the domestic market, rather than for exports. Because import protection and monopoly control create higher domestic prices and profits, in effect local consumers pay higher prices as a transfer to the multinationals' foreign stockholders. Because evidence suggests that FDI aimed for protected domestic markets brings with it the least benefits (and sometimes net economic losses), it is often difficult to justify this kind of intervention.

Perhaps the most controversial of all steps taken to encourage FDI is income tax incentives. While these incentives can take a wide variety of forms, the most common is income **tax holidays**, which exempt firms from paying taxes on corporate income, usually for three to six years. Most countries otherwise would tax profits at rates anywhere from 20 percent to as high as 50 percent. For tax holidays to help the multinationals, they must be creditable against income taxes due to their home country governments. Otherwise, if home countries tax firms on worldwide income, as all industrial countries except France do, then the MNC would pay less tax to the developing country but more to its home country. In effect, taxes foregone by the developing country would simply be transferred to tax revenues of the multinational's home country. Most industrial countries now permit their firms to take credit for tax holidays granted abroad through tax treaties negotiated between host- and home-country governments.

There is substantial debate about the circumstances, if any, under which tax holidays are justified. Most studies conclude that, for many kinds of FDI, income tax holidays have only marginal effects on multinational investment decisions and reward the multinationals for doing what they would have done in any case. This is especially true for firms attracted by natural resources or those intending to produce in protected domestic markets of the host countries. Under these circumstances, MNCs are attracted by specific characteristics of the host country, such as the presence of gold or petroleum deposits or a large domestic consumer market. These MNCs may have few other serious location options, so tax treatment may not be a critical factor in their investment decision. Moreover, tax holidays cannot turn an unprofitable investment into a profitable one: Taxes become relevant only once a firm is earning profits.

Evidence suggests that export-oriented, labor-intensive, "footloose" industries may be more sensitive to tax holidays and other incentives because they have a much wider set of location options. Firms producing electronics products, shoes, textiles, clothing, games, and toys can choose among dozens of countries that offer the basics of political and economic stability, relatively low transport costs, and a pool of unskilled and semiskilled workers. Under these circumstances, tax holidays may be an important factor for the MNC in choosing its location. Because research suggests that these kinds of investments are more likely to provide positive spillovers to the

host country, limited tax holidays may be justified. These steps always are controversial, however, because local firms and other MNCs quickly demand similar tax treatment, and it may be difficult for governments to resist this pressure.

REQUIREMENTS AND RESTRICTIONS ON FDI Many governments impose restrictions and requirements on MNCs in an attempt to capture as much as possible of the expected benefits from foreign direct investment, including performance requirements, local ownership requirements, labor requirements, and restrictions on profit repatriation. It is possible that these policies can increase the benefits to the host country; however, under most circumstances, they discourage FDI.

Many host countries have made it mandatory for foreign investors to sell a specified share of equity, usually at least 51 percent, to local partners to form **joint ventures**. Through local ownership requirements, host governments hope to ensure the transfer of technology and managerial skills, limit the repatriation of profits, and maintain local control. However, parent multinationals often are more reluctant to allow diffusion of technology to joint ventures than to wholly owned subsidiaries. Moreover, many local joint-venture partnerships are pro forma arrangements involving local elites close to the centers of political power with little knowledge or expertise in business that simply receive occasional payment from the MNC as a figurehead partner. This is not always the case, however. In some cases, skilled local partners eventually can take greater control and possibly buy out the MNC.

Many countries impose **domestic content requirements** on MNCs, requiring them to purchase a certain share of their inputs locally (as opposed to through imports), often with the share increasing over time. The idea is to encourage local suppliers and ensure stronger links between the MNC and local firms. Ideally, MNCs would choose to purchase locally in any case, but when they are forced to do so it usually indicates that the local suppliers are not producing at low enough prices and high enough quality to compete with imports. These requirements add directly to the costs of the MNC and can discourage investment. Domestic content requirements today, however, typically violate the rules of the World Trade Organization (WTO) and thus cannot be used by developing countries subject to those rules. Similar production requirements can be aimed at labor or technology transfer. Policies that make foreign firms use local personnel are aimed not only at job creation but also at increasing absorptive capacity for the transfer of technology from multinationals. Developing countries have tried to promote technology transfer by imposing standards requiring multinational firms to import only the most-advanced capital equipment rather than used machinery. Other common restrictions include ceilings on repatriation of profits to the parent corporation and stiff taxes on profit remittances. As with domestic content requirements, if not carefully designed, these restrictions can deter investment and may do little in the long run to help transfer benefits to local entities.

Overall, the accumulated evidence indicates that some types of FDI bring large benefits and others bring small or even negative benefits. This suggests that the

best general policy is to get the basics right to make the investment climate more attractive for both domestic and foreign investment: reliable transportation, power, and communications facilities; freedom from unnecessary or onerous government regulations; capable institutions; adherence to the rule of law; macroeconomic and political stability; and labor markets in which wages are matched by productivity.

SAVINGS

As pointed out at the beginning of this chapter investment is not possible without first some person, company, or government saving part of their income. The saver can then invest the funds saved directly or deposit them in an intermediary (a bank or other financial institution) that specializes in identifying promising investments and lending it the money needed to make those investments a reality.

In the early years after the end of World War II it was commonly assumed that a lack of savings was at the heart of why so many of the countries in the world were poor. Because people in those countries were so poor, it was argued, they had to use all of their income to maintain a subsistence level of income and had nothing left over to save and invest. This problem was seen as the most fundamental of the various vicious circles of poverty that plagued low-income countries. To break the circle and get these countries on the path out of poverty, an outside source of savings (and hence investment) had to be found. The answer for many was to increase foreign aid provided by high-income countries to those mired in this rut. Arguments for foreign aid of this sort are still heard today but the rationale for most analysts has moved away from this view. (Foreign aid and the rationale for it are the subject of Chapter 14.)

Foreign aid is a form of **foreign savings** that is transferred to low-income countries and either invested or consumed. Other forms of foreign savings include FDI (discussed earlier in this chapter); portfolio equity by foreigners in developing country stock markets; and loans to developing countries by private (usually high-income-country) financial institutions, high-income governments, and international financial institutions (such as the World Bank, the IMF, and the regional international development banks).

Domestic sources of savings, however, are far larger in most developing countries than all of the sources of foreign savings combined. **Domestic savings** is made up of household, domestic corporation, and government savings. Domestic savings tend on average to be higher in high- and middle-income countries than in low-income countries but there is enormous variation within each group, and the savings rate among low-and middle-income countries has risen from two or three decades ago. China had a savings rate in 2005–09 of roughly 50 percent of GDP, an extraordinarily high rate. India, a country with lower income than China's, had a savings rate of over 30 percent

in the same period, also a high amount. But some nations save very little; Nepal saved about 10 percent of its GDP, whereas Ghana and Guatemala had domestic savings rates of 5 percent or less. High-income countries such as the United States and the United Kingdom had savings rates below 15 percent of GDP. Savings and investment data for a selection of countries are presented in Table 10–4.

How does a low-income country with a low savings rate increase that rate to contribute to a higher rate of growth in income? Can government policy have a major influence on a country's savings rate? Since the nineteenth century it has been assumed that higher incomes would raise the savings rate. In all countries the highest income earners in the population typically save far more than the lowest income earners. It would seem to follow that if a country's income increased, the average rate of savings of the country as a whole would rise along with income. But when economists set out to measure the change in savings rate over time, they quickly discovered

TABLE 10–4 Domestic Capital Formation (Investment) and Savings

	GROSS CAPITAL FORMATION (% OF GDP)		GROSS DOMESTIC SAVINGS (% OF GDP)	
	1975–1979	2005–2009	1975–1979	2005–2009
Economy				
Low income	14	23	7	10
Middle income	26	28	25	30
High income	24	20	24	20
Region				
East Asia and Pacific	30	38	30	44
Europe and Central Asia	n.a.	23	n.a.	24
Latin America and Caribbean	25	21	23	23
Middle East and North Africa	30	26*	24	32*
Sub-Saharan Africa	25	21	23	16
South Asia	18	33	16	28
Country				
Brazil	24	18	21	19
China	32	44	32	51
Egypt	31	20	16	16
Ghana	9	22	9	5
Guatemala	20	18	16	3
India	20	36	19	32
Indonesia	24	27	29	30
Mexico	24	25	23	23
Nepal	16	29	12	10
Russian Federation	n.a.	22	n.a.	32
South Africa	27	20	31	18
United Kingdom	20	17	20	14
United States	20	18	20	13

*Data from 2005–2007 only.
GDP, gross domestic product.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

that as average per capita income rose, the national average domestic savings rate did not change much if at all.[14] It turns out that we do not really even know whether it is a rise in savings that leads to a rise in investment that in turn leads to higher growth or whether the causation goes in reverse. Is it instead the rise in the rate of growth brought about by various reforms in government policy that leads to a higher rate of domestic savings? In the economies of South Korea and Taiwan, for example, reforms in economic policy in the early 1960s jump-started growth at a time when the savings rate, particularly the savings rate in South Korea, was very low. Rapid growth during the 1960s and into the 1970s then led to a rising rate of domestic savings and domestic investment that sustained the high growth sparked by the earlier reforms.

Although there is much not known about what determines a country's overall rate of domestic savings, we do know something about what drives household savings, typically the most important part of the overall level of domestic savings. We also can say something about government savings but much less about savings by domestic corporations. We will look at each of these sources of domestic savings in turn. The chapter ends with a return to the issue of foreign savings and what drives whether a country is a net lender (that transfers savings abroad) or a net borrower (which receives other countries' savings, as in the case of FDI).

HOUSEHOLD SAVING AND CONSUMPTION

Individuals and households save for two main reasons. First, they want to generate higher future income by saving and investing some of their current income. Second, they want to save some current income to protect themselves from the risk of unexpected falls in income in the future, often from a serious illness on the part of the major income earner; the loss of employment for some other reason; a drought that undermines next year's food crop; the loss of productive assets from fire, flood, theft, or disease (that kills crops or livestock); or other loss of income. This precautionary motive can be particularly important for poor people living on subsistence levels, for whom a sudden loss of income can be catastrophic, or where insurance markets are not well developed and individuals, families, and businesses must self-insure against the risks of a sudden loss of income or assets. In most low-income countries insurance markets are weak, especially in poor rural areas.

Given these motivations, a key decision is how much to save and how much to consume. The more a person saves, the more he or she can invest to increase future income and the more he or she can safeguard against future shocks. But saving comes at a cost: The only way to save is to reduce current consumption. Households

[14]This discovery led to the works of Milton Friedman on the permanent income hypothesis and James Duesenberry on the relative income hypothesis, which, in different ways, attempted to explain this paradox.

will sacrifice current consumption to save only if they believe either that the ensuing investment will increase their income and consumption in the future or they will need to draw on that saving to finance basic consumption in the event of a severe income shock. A wide variety of factors might influence saving: the current level of income, anticipated future levels of income, the number of dependents that must be supported, the opportunities for productive investment, and the quality of the financial institutions through which some saving might be channeled.

The **life-cycle model** of household savings captures an important element of why some household savings in developing countries is high whereas with other kinds of households it is low. The life-cycle model, associated most closely with Nobel Prize–winning economist Franco Modigliani, is specific about how saving and consumption would be expected to vary systematically during a person's lifetime.[15] In this model, young adults tend to have lower saving rates because they have lower incomes (and expect higher incomes in the future) and are raising children. Indeed, many people would be expected actually to dissave (go into debt) during this stage of life. Saving rates tend to rise and peak toward the middle and end of a person's working years, when incomes are higher and there are fewer consumption-related expenses for children. During this stage of life, people accumulate the bulk of their saving to be used during retirement. Once workers retire, their income falls and they again dissave by drawing down on their previous saving. Thus saving rates tend to be low or even negative during the younger years, high during middle age, and negative again after retirement.

The life-cycle hypothesis suggests that the demographic structure of a society may have a strong effect on overall saving rates. As discussed in Chapter 7, all societies tend to pass through a demographic transition with three basic stages: (1) high birth and death rates, with low population growth, (2) falling death rates with continued high birth rates and consequently high population growth rates, and (3) falling birth rates with continued low death rates and lower population growth. As countries pass through these stages, the share of the population too young to work, of working age, or retired tends to change dramatically, and saving rates may change with them. A society in stage 2 of the demographic transition is likely to have a relatively large number of surviving children because of falling death rates and continued high birth rates. Children, of course, earn very little or no income, so their consumption effectively is negative saving. At this stage of the demographic transition, each worker in the society has more young dependents to care for, so saving rates are expected to be relatively low. Later on, as a country enters stage 3 of the transition, the number of

[15]See Franco Modigliani and Richard Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. Kurihara, ed., *Post Keynesian Economics* (New Brunswick, NJ: Rutgers University Press, 1954); Alberto Aldo and Franco Modigliani, "The Life-Cycle Hypothesis of Saving: Aggregate Implications and Tests," *American Economic Review* 53, no. 1, (March 1963), 55–84. For a classic early work on the subject, see Ansley J. Coale and Edgar M. Hoover, *Population Growth and Economic Development in Low-Income Countries* (Princeton, NJ: Princeton University Press, 1958).

children tends to fall and the large number of children from the previous generation enters the workforce. As a result, society as a whole tends to have a far larger share of workers and fewer dependents. Because each worker has fewer children to care for, saving rates are expected to rise. Later still, these workers retire and saving rates tend to decline. This hypothesis is consistent with an income-saving relationship in which saving rates are high for middle-income countries and lower for high-income countries. Higher-income countries tend to have longer life expectancies and more people living in retirement, which would suggest lower average saving rates for these countries.

Figure 10–2 shows the relationship in 100 low- and middle-income economies between saving rates and the *dependency ratio*, which compares the size of the dependent population (children and retirees, the population aged 0 to 15 years and those over 60) to the working-age population. The figure shows a tendency for saving rates to be lower in countries with a larger number of dependents relative to the workforce. Note that there is significant variation around the line. Gabon and Tajikistan, for example, both have similar levels of dependency, but Gabon had a gross domestic savings rate of almost 50 percent while Tajikistan had one of *negative* 21 percent. Clearly, more than the dependency burden determines the rate of national savings. Some studies that controlled for more variables found little connection between dependency ratios and saving rates, while others detected a relationship.[16] The relationship seems to be more robust with respect to the young-age dependency ratio (the ratio of the population between 0 and 15 years of age to the working-age population). The evidence is weaker with respect to the old-age dependency ratio, which compares the size of the population aged over 60 to the working-age population, suggesting that this cohort may not reduce its saving as much as the life-cycle theory suggests.

The life-cycle hypothesis may capture a reason for the high savings rates in East Asia. When the East Asian economies such as South Korea and Taiwan began to grow rapidly, their societies also went through a rapid demographic transition as birth rates came down soon after the decline in death rates, leading to a population largely concentrated in the working ages of 15 to 65 and a declining young population. As time passed, those of working age gradually retired and with increasing life expec-

[16]Studies that find no relationship between demographic structure and saving include Angus Deaton, *Understanding Consumption* (Oxford: Clarendon Press, 1992); M. Gersovitz, "Savings and Development," in Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 1 (Amsterdam: North-Holland, 1988). Studies that reach the opposite conclusion include Sebastian Edwards, "Why Are Saving Rates So Different across Countries? An International Comparative Analysis," NBER Working Paper 5097 (Cambridge, MA: NBER, 1996); Paul R. Masson et al., "International Evidence on the Determinants of Private Saving" *The World Bank Economic Review* 12, no. 3 (September 1998), 483–501; Norman Loayza, Klaus Schmidt-Hebbel, and Luis Serven, "What Drives Private Saving across the World?" *Review of Economics and Statistics* 82, no. 2 (May 2000), 165–81; Steven Radelet, Jeffrey Sachs, and Jong-Wha Lee, "Economic Growth in Asia" (Cambridge, MA: Harvard Institute for International Development, May 1997).

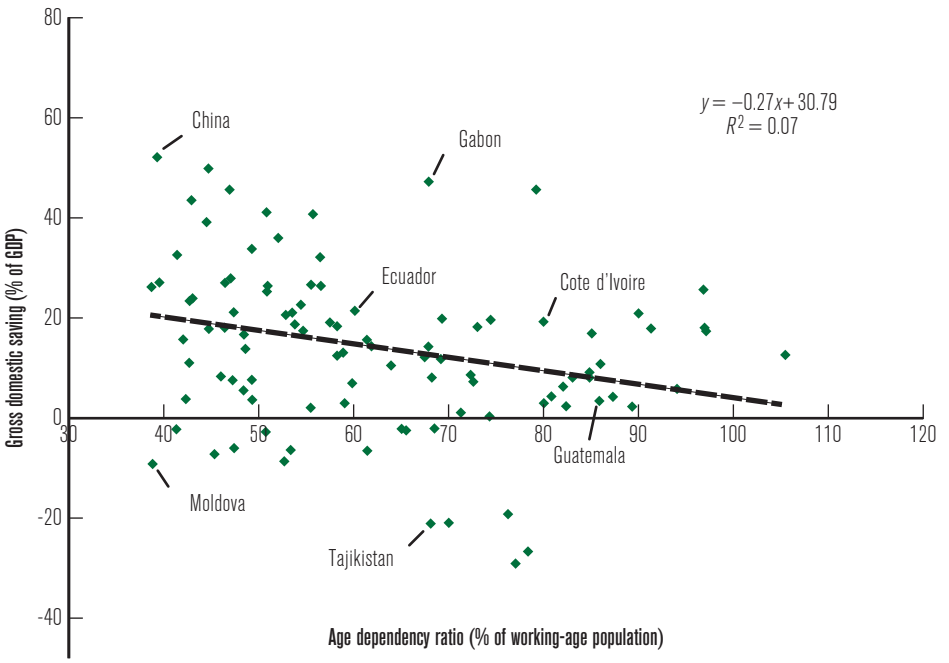


FIGURE 10–2 Gross Domestic Saving and the Age Dependency Ratio, 2009

GDP, gross domestic product.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

tancy, elderly dependents made up a rapidly rising share of the population, but this did not occur until after these economies had experienced three or more decades of high savings, high investment, and rapid economic growth. China's one child policy has produced a similar phenomenon. The one child policy was instituted in the first part of the 1970s and had taken full effect by the time the economic reform began in 1978 that jump-started economic growth. Rapid growth thus benefited from and was sustained to an important degree by the dwindling share of dependent young people, which in turn made high rates of household savings possible. A rising share of elderly dependents in China is only now beginning to occur and will not have a major impact on the overall dependency ratio for another decade.

CORPORATE SAVING

Corporate saving is relatively small in most developing countries primarily because the corporate sector generally is small. For a variety of reasons, there are fewer pressures and incentives in developing countries for doing business in the corporate form (as opposed to operating as an unincorporated business). The principal reasons for

organizing as a corporation in the private sector are to limit the liability of enterprise owners to amounts invested in a business and to facilitate enterprise finance through the issue of equity shares (stocks). Although these advantages are substantial in higher-income countries with well-developed commercial codes, civil court systems, and capital markets, they are smaller in most developing countries, where the collection of commercial claims (for example, company debts) through the courts is relatively difficult and where capital markets are poorly developed, when they exist at all.

As usual, however, there are important exceptions to these generalizations. In some developing countries, corporations are both numerous and quite large. As noted earlier, *Fortune* magazine's annual list of the 500 largest corporations in the world (in 2010) includes MNCs from developing nations such as China's Sinopec Corporation, Shanghai Automotive Industry Corporation, and the Industrial and Commercial Bank of China. The giant firms in India included Tata Steel and in Brazil the financial conglomerate Banco Bradesco. But less than 15 percent of the *Fortune* list was from low- and middle-income nations and two-thirds of them were located in China. The vast majority of developing countries had none. There are, of course, many good-size corporations that do not make it into the *Fortune* 500 largest list. Thus in most developing countries, even when corporations of say a billion dollars in sales are accounted for, such corporations do not account for a large share of private-sector business activity and do not provide a high proportion of domestic saving.

In all but a few of the highest-income developing countries, the great bulk of private sector farming and commercial and manufacturing activity is conducted by unincorporated, typically family-owned enterprises. Some of these businesses fall into the category of medium-scale establishments (from 20 to 99 workers). The great majority are small-scale operations with fewer than 20 employees. The noncorporate sector, including most small family farms and urban informal sector enterprises, manages to generate more than half of all domestic saving in most developing countries, and this sector is the only consistent source of surplus in the sense that the sector's saving exceeds its investment. For those closely held, largely family-owned and -managed firms, enterprise profits become an important part not only of corporate saving but also of gross household income. The available evidence indicates, and economic theory suggests, that household saving accounts for the overwhelming share of private saving in developing countries and the chief source of household saving is probably household income from unincorporated enterprises.

GOVERNMENT SAVING

Unlike private saving, there are no well-developed theories for government saving and consumption behavior. Government saving arises when tax revenues exceed public consumption expenditures (that is, government spending excluding

public investment expenditures).[17] Government saving in most countries tends to be smaller than private saving (although not always), typically averaging between 3 and 10 percent of GDP in most developing countries. In countries where governments run large budget deficits, government saving is negative. Government saving also tends to be much smaller in low-income countries than in middle- and upper-income countries. One reason is that poorer countries generally do not generate as much tax revenue as richer countries.

Even though government saving usually is smaller than private saving, historically it has been given strong emphasis in development policy. This is partly because private saving is often seen as inherently constrained by such factors as low per capita income and high private consumption propensities among wealthy families with the greatest capacity for saving. But it is also because government saving can be controlled more directly by policy makers, and public sector investment financed by that saving (roads, bridges, ports) plays an important role in supporting economic growth.

Government saving can be changed by either increasing revenues or reducing consumption expenditures. Historically, most effort has focused on increasing tax revenues, either by strengthening tax structures and collection systems or by altering tax rates. In this line of thinking, tax ratios in many developing countries are too low, and policy changes aimed at increasing tax revenues help increase domestic saving (and ultimately the rate of economic growth). But there are several steps in between, and the extent to which this occurs depends on three key questions:

1. To what extent can governments increase the tax ratio, if they want to do so?
2. Will an increase in the tax ratio lead to an increase government saving?
3. Will an increase in government saving increase total domestic saving?

On the first question, it is not easy to judge the appropriate level of taxation in developing countries. If taxes are too low, governments cannot provide essential services, such as basic education, public health, a well-functioning judicial system, basic security, and public infrastructure. Tax rates that are too high (or fall too heavily on some groups) can undermine private entrepreneurs and reduce the incentives for new investment and economic growth. As a general matter, developing countries generally do not (and cannot be expected to) have tax ratios as high as is common in wealthier countries (except for developing countries with significant natural-resource endowments, which are easier to tax). Moreover, it is not easy to increase tax collection in developing countries, if for no other reason than their much lower per capita income allows a much smaller margin for taxation after subsistence needs are met. Figure 10–3 shows that in the 1990s, whereas typical tax ratios for low-income countries averaged around 15 percent, the ratio of government revenue to GDP in the industrialized countries averaged over 30 percent.

[17]In some countries saving by government-owned enterprises contributes to government saving, but it generally plays a minor role, so our discussion of government saving is confined to budgetary saving.

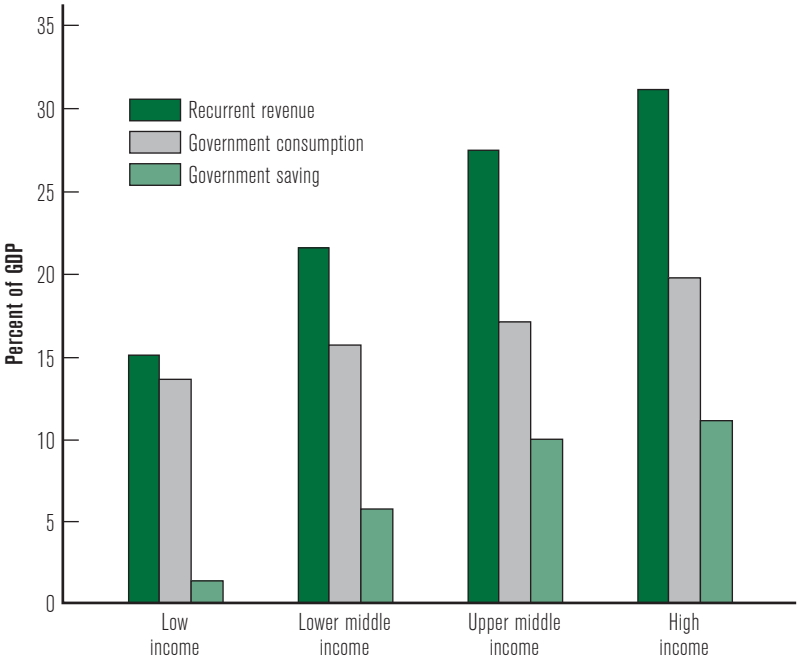


FIGURE 10–3 Government Revenue, Consumption, and Saving, 1990–2002

Tax revenues as a share of income tend to be much smaller in low-income countries than in high-income countries. Government consumption also grows with income but not nearly as much. Therefore, government saving generally is much smaller in low-income countries than in high-income countries.

GDP, gross domestic product.

Turning to the second question, a higher tax ratio does not necessarily lead to a higher government saving rate. It depends on what the government does with the revenue. To the extent that it spends the revenue on consumption (that is, increased government employment or higher salaries), government saving does not increase; to the extent it uses the revenue to reduce a budget deficit or for investment, government saving increases. More specifically, the impact of increased tax revenue on the government saving rate depends on the government's propensity to save (as opposed to consume) an additional peso or rupiah of tax revenue. In many countries, the government's marginal propensity to save out of taxes has been sufficiently small that higher taxes often have had relatively little effect on government saving and in some cases none at all.

It is worth emphasizing that increasing government consumption (rather than saving) may or may not be bad policy. Some consumption expenditures may be a waste, such as fancy cars, luxurious government offices, or a rapid buildup of military purchases. But other consumption expenditures may be vitally important, such as purchasing vaccines; making salary adjustments to keep and attract qualified civil servants; maintaining roads, schools, health facilities, and communication networks; and strengthening the judicial system or financial institutions.

Government consumption tends to rise less sharply with income than does revenue, as shown in Figure 10–3. There is much less difference in government consumption ratios between low-income countries (14 percent) and high-income countries (19 percent) than there is in tax ratios. This difference leads to a substantial divergence in average government saving rates. The poorest countries tend to have very low (and sometimes negative) government saving rates; government saving rates tend to rise with income as tax ratios increase.

On the third question, perhaps counter intuitively, an increase in government saving does not automatically increase domestic saving. How can this be when government saving plus private saving equals domestic saving? It depends on how private saving responds to the increase in government saving. If the government raises taxes, government saving might increase, but private individuals might respond by lowering their saving to pay the higher taxes, thereby dampening the impact on aggregate domestic saving. Thus the effect on saving of raising taxes depends on both the government's and the private sector's propensity to save out of the marginal dollar of income. As a general statement, increased tax revenue increases the domestic saving rate only if the government's propensity to save is greater than the private sector's propensity to save.

Debates about the private sector's reaction to increased government saving date back at least as far as the early part of the nineteenth century to the writings of the great British economist David Ricardo. Ricardo postulated—and largely rejected—the idea that, under certain circumstances, a change in taxes will have absolutely no effect on total domestic saving. Ricardo's theory has come to be known as **Ricardian equivalence**: Any increase in public saving is offset by an equivalent decline in private saving, with total domestic saving remaining unchanged. The basic idea is that when private actors receive a tax cut and see that government spending does not change, they anticipate a commensurate tax increase in the future and prepare for it by increasing their current saving. Thus, if the government reduces taxes, government saving may fall, but households and firms increase private saving in anticipation of a future tax increase. The fall in government saving is offset exactly by a rise in private saving, and the tax cut has no impact on total saving. Similarly, a tax increase may lead to higher government saving, but private saving is expected to decline by the same amount with no change in total saving. Ricardo's original idea was rejuvenated and formalized by economist Robert Barro in 1979.[18]

Ultimately, of course, this is an empirical issue that comes down to measuring the extent to which private agents alter their saving behavior in response to changes in government tax and expenditure policies. The abundance of evidence suggests partial, but only partial, truth to Ricardo's hypothesis. Many studies have shown that an increase in government saving tends to be associated with a decline in private saving, but by less

[18]Robert Barro, "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82, no. 6 (November–December 1974), 1095–117.

than the strict one-to-one relationship postulated by Ricardo. Most studies have found an offset of between 0.40 and 0.65, meaning that a 1 percentage point of GDP increase in the government saving rate is associated with a 0.40- to 0.65-percentage-point decline in the private saving rate. This would imply a net increase in the total domestic saving rate of between 0.35 and 0.60 percentage points.[19]

The salutory impact of higher government saving may go beyond the effect on domestic saving. Because higher government saving is derived largely from smaller budget deficits, it reduces the government's borrowing requirements. Government deficits in developing countries typically are financed through domestic money creation (that is, borrowing from the central bank, which prints money to fund the deficit) or by borrowing in either local or overseas markets. Larger budget deficits can lead to higher rates of inflation (from money creation), larger government debts, or both. Lower budget deficits (or higher public saving) tend to have the opposite outcomes: lower inflation and indebtedness. Thus higher public saving is likely to enhance general macroeconomic stability.

FOREIGN SAVING

Historically, many countries have augmented domestic saving with foreign saving to help finance investment and growth. The United States relied heavily on foreign saving, particularly during the antebellum period from 1835 to 1860 and again in the late nineteenth century to finance, among other things, the expansion of the railway system. Likewise, Russia used foreign saving to help propel its development in the three decades before World War I and the Communist revolution. However, not all countries have followed this pattern: Japan became a wealthy nation even though it actively discouraged inflows of foreign saving and investment throughout its history. Therefore, foreign saving can help development but is not necessarily essential for it.

In decades past, some developing countries actively tried to shut out private capital flows; today almost all developing countries actively try to encourage them. But controversy surrounds foreign aid, foreign investment, short-term private capital flows, and the debt that has accrued from foreign borrowing. The remainder of this chapter provides a brief overview of some basic concepts and data on foreign saving. Foreign saving in the form of FDI was discussed earlier in this chapter; foreign savings transferred through loans to developing countries and the resulting accumulation of foreign debt is the subject of Chapter 13, and more in-depth analyses of foreign aid as a source of savings is the subject of Chapter 14. The appendix to Chapter 15 provides related background on balance-of-payments accounting.

[19]See, for example, Edwards, "Why Are Saving Rates So Different across Countries?;" Masson et al., "International Evidence on the Determinants of Private Saving"; Loayza et al., "What Drives Private Saving across the World?"; Radelet et al., "Economic Growth in Asia."

The relationships between investment, domestic saving, foreign saving, and the trade balance can best be seen from standard national accounts identities, which tell us that total investment (I) must equal total saving (S), which in turn consists of domestic saving (S_d) and foreign saving (S_f). To see this, recall from the income side of the national accounts that

$$Y = C + I + G + X - M \quad [10\text{–}3]$$

where G is government consumption, X is exports of goods and services, and M is imports of goods and services. On the expenditure side, all income must be consumed, saved, or given to the government as taxes:

$$Y = C + S_p + T \quad [10\text{–}4]$$

Note that the saving term (S_p) in this identity refers to private saving. Because both equations equal Y, the right-hand sides must be equal:

$$C + I + G + X - M = C + S_p + T \quad [10\text{–}5]$$

Subtracting C from both sides and rearranging the terms yields

$$I = (T - G) + S_p + (M - X) \quad [10\text{–}6]$$

The first term on the right-hand side ($T - G$) is government saving, and the second term (S_p) is private saving. The third term ($M - X$) is both the current account deficit of the balance of payments and foreign saving. When a country's imports (M) exceed its exports (X), the difference must be financed by inflows of capital from abroad (foreign aid, bank loans, equity flows, etc.), which is foreign saving. The right-hand side of the equation is the sum of domestic and foreign saving, yielding

$$I = S_d + S_f \quad [10\text{–}7]$$

or the more traditional

$$I - S_d = M - X \quad [10\text{–}8]$$

Thus, as the difference between investment and domestic saving grows (or shrinks), the trade deficit grows (or shrinks) commensurately. The two balances must move together in tandem. One can think of this in many ways. As investment increases (without an increase in domestic saving), imports tend to increase and exports decrease to provide the materials needed for the investment project. A firm building a new factory requires steel, concrete, lumber, and machinery, which can come from either new imports or purchasing the goods domestically, leaving less for export. Alternatively, an export boom without a commensurate increase in imports generates profits that can contribute to domestic saving, closing the gap between investment and domestic saving.

Foreign saving includes both a public sector component (called *official foreign saving*) and a private component. Most **official foreign saving** is foreign aid provided

on **concessional terms** as either *grants* (outright gifts) or *soft loans,* meaning that they bear lower interest rates and longer repayment periods than would be available in private international capital markets. Industrialized-country governments and international agencies also provide some financing as loans on commercial terms (*hard loans*), particularly to middle-income countries.

Private foreign saving consists of four elements. *Foreign direct investment* (discussed at length earlier in this chapter) is made by nonresidents, typically but not always by multinational corporations, in enterprises located in host countries. Direct investment implies that funds go directly to an enterprise (as opposed to indirectly through a stock market), with the foreigner gaining full or partial control of the enterprise. *Portfolio equity* is the purchase by foreigners of host country bonds or stocks, without managerial control. Portfolio equity was a very important form of financing in the nineteenth and early twentieth centuries, but fell into disuse after World War II. It was revived in the 1990s, however, as rich-country investors showed interest in emerging stock and bond markets, especially in Asia and increasingly in Latin America. *Commercial bank lending* to developing country governments and enterprises supplanted portfolio investment in importance for a time but waned beginning in the 1980s when debt crises afflicted many developing countries. It expanded quickly in the early 1990s, then dropped sharply toward the end of the decade, as several developing countries experienced major financial crises. In some cases it had not fully recovered when the global financial crisis hit in the latter half of 2008. Finally, exporting firms from the industrialized countries, their commercial banks, and official banks (such as the U.S. Export-Import Bank) offer *export credits* (or credit guarantees) to firms in developing countries that import from the industrialized countries. These credits help the exporters promote sales by permitting delayed payment, often at commercial interest rates.

Capital flows to developing countries changed dramatically in several important ways between 1990 and 2008, as shown in Table 10–5. First, private capital flows grew extremely rapidly until 1996, growing from $82 billion to $281 billion in 1996. All the major categories of private finance expanded rapidly: long-term loans, short-term loans, portfolio equity, and foreign direct investment. But this trend changed abruptly in 1997 after the financial crises that struck Thailand, Indonesia, Korea, Russia, and several other countries. Private capital flows fell sharply for several years and then grew rapidly between 2000 and 2008 (and declined again in 2009 after the financial crisis that originated in the United States).

Second, and in sharp contrast to private flows, official development flows fell during the 1990s. In particular, net official lending dropped significantly and turned negative, meaning that developing countries were repaying old loans more than they were receiving new loans from donor governments. Whereas official development finance and private finance differed by a factor of two in 1990, private flows were 35 times larger in 2008. While these two patterns hold for developing countries as a whole, the patterns differ between low-income and middle-income countries. In particular, the large increase in private capital flows was concentrated in middle-

TABLE 10-5 Financial Flows from and to Developing Countries, 1990–2008 (billions, US$)

	1990	1996	2000	2008
Inflows	81.6	281.3	172.3	727.3
Official development finance				
Official loans (net)	26.4	3.8	−5.8	20.4
Private finance	55.2	281.3	178.1	706.9
Long-term debt flows (net)	16.6	82.5	7.4	124.2
Short-term debt flows (net)	13.2	37.4	−7.9	−16.3
Portfolio investment	3.4	32.9	12.4	15.7
FDI into developing countries	22.1	128.6	166.2	583.0
Outflows	53.5	200.5	226.6	1,105.0
Accumulation of reserves	32.4	90.4	46.8	447.3
FDI from developing countries	5.0	10.0	16.5	164.0
Other items/errors and omissions	16.1	100.1	163.3	493.7
Memo item: aggregate current account balance of developing countries	−28.1	−83.6	43.6	377.9

FDI, foreign direct investment.

Source: World Bank, *Global Development Finance 2005* (Washington, DC: World Bank, 2005) http://go.worldbank.org/581EDYH6R0; IMF, *International Financial Statistic* (Washington, DC: IMF); Annual Yearbook and World Bank, *Global Development Finance 2009* (Washington, DC: World Bank, 2009) http://go.worldbank.org/581EDYH6R0.

income countries. Private capital flows to low-income countries were small in 1990, and while they were larger after the year 2000, the increase accounted for a relatively small share of the increase in total private capital flows. Official flows to low-income countries were only slightly larger in 2008 than they were in 1990, particularly if the figures are adjusted for inflation; in middle-income countries, the picture is similar.

Third, whereas capital inflows to developing countries in aggregate grew rapidly, capital outflows grew even more rapidly. The aggregate current account deficit of developing countries (shown at the bottom of Table 10–5) moved from a deficit position to a surplus position in 2000. In other words, aggregate exports of all goods and services exceeded imports of goods and services, so developing countries as a whole were exporting capital to the richer countries. Correspondingly, total net capital flows (or net foreign saving, summing together all inflows and outflows) shifted from positive to negative. The main channel through which this occurred was a small number of developing countries running large balance of payments surpluses and accumulating foreign exchange reserves, especially China. In 2007 and 2008, China accumulated foreign exchange reserves at a rate of over $400 billion per year. Typically balance of payments surplus countries (along with other central banks in developing countries) hold their foreign exchange reserves in one of two ways: (1) as deposits in banks in the United States, Europe, or Japan, which is a form of saving available to finance new loans in those countries, or (2) by purchasing government securities, such as U.S. Treasury bills. China's central bank (the People's Bank

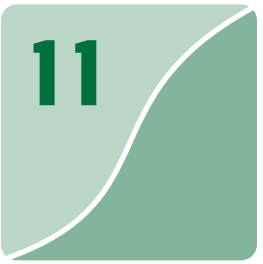
of China) became one of the largest purchasers of U.S. Treasury bills. In effect, the balance of payments surpluses generated by these countries are providing the saving necessary to finance the U.S. budget deficit. Historically and under normal circumstances, capital tends to flow from rich countries (which tend to have larger capital stocks) to low-income countries (where capital is scarce), but since 2000 that pattern has been reversed. However, this pattern is heavily concentrated in a small number of developing countries that are exporting a large amount of capital by accumulating reserves. The majority of low-income countries continue to be net importers of capital from rich countries.

SUMMARY

- Both public investment and private investment are central to economic growth and development. Public investment generally focuses on roads, ports, telecommunications facilities, schools, and health facilities. Private investment tends to be larger than public investment and provides the dynamism for the vast majority of new jobs, new technologies, and growth in economic output.
- Cost–benefit analysis provides a rigorous technique for evaluating both public and private investments and calculating the net present value over time of each investment. This technique can be used for commercial evaluation (to calculate private-sector financial benefits), economic evaluation (to measure the economic costs and benefits to society as a whole), or social evaluation (incorporating the impact of investments on important social goals).
- The key factors that influence investment at the macro level are similar to the variables discussed in Chapter 3 that promote economic growth. At the micro level the quality of the business environment is the main influence on the level and productivity of the investment. Governments that intervene excessively, thus delaying and raising the cost of business investment decisions, lead to rent seeking and corruption that further reduce the quantity and quality of investment.
- Foreign direct investment can generate important benefits, but not all FDI is alike. FDI focused on firms producing exports on competitive global markets appears to be most beneficial, creating jobs and bringing new technologies and ideas from abroad, among other spillovers. FDI for natural resources or for firms operating in protected domestic markets generally has had more mixed benefits and sometimes has been costly to recipient countries. Policies specifically aimed at attracting FDI, such as tax

holidays, are more controversial than those aimed at improving the general investment environment.

- Investment requires that there first be savings. Savings can be either by domestic individuals and organizations (corporations, governments) or by foreigners who then allow their savings to be transferred abroad through foreign aid, foreign direct investment, portfolio equity, or commercial lending.
- Domestic savings tend on average to be higher in high- and middle-income countries than in low-income countries, but there is enormous variation within each group.
- A determinant of the level of savings is the demographic makeup of a country. A country in the early stages of a demographic transition with a low death rate and a high birth rate will have a high rate of growth in the number and share of children not yet old enough to work, which may lower the savings rate. Countries further along in the demographic transition have lower birth rates and hence a higher share of the population is of working age (and not of retirement age), which can lead to higher rates of saving.
- Government saving makes an important contribution to total saving in many developing countries, although it is usually much smaller than private saving. An increase in government saving generally increases total domestic saving but typically by less than one for one, as it can be partially offset by a decline in private saving.
- Foreign saving is relatively small compared to domestic saving, but is often more controversial. The composition of foreign capital flows has changed markedly since the early 1990s. The amount of official finance flowing to low-income countries dropped sharply. By contrast, private finance grew even more quickly, at least until severe financial crises affected several countries in the late 1990s, but recovery and growth to even higher levels than in the past occurred through 2008. The increase in private capital flows has been concentrated in middle-income countries, with much less flowing to low-income countries. In a change from the normal pattern, in recent years some developing countries, particularly China, have been providing financing to rich countries.



Fiscal Policy

Beginning in late 2007, what had been a booming market for housing in the United States evolved instead into a depressed market with house prices falling and many homeowners forced into foreclosure. Financial instruments created by packaging large numbers of mortgages that were rated as safe investments suddenly dropped sharply in value, triggering losses not only for individuals but for major financial institutions such as Bear Stearns, Lehman Brothers, and the American International Group (AIG). The failure of these financial institutions in turn threatened the viability not only of other financial institutions but of the entire U.S. economy. The potential failure of many American financial institutions and the related sharp decline in economic activity in the United States in turn led to similar threats to financial institutions in the rest of the world, notably in Europe, and initiated a global recession that was the most severe recession worldwide since the Great Depression of the 1930s. Developing countries had not caused the global recession. Many had gone through a similar financial crisis in 1997–98, which had led to a major recession in several large developing economies, particularly in Asia. That crisis in turn led many of those economies to reform their financial systems in ways that protected them from the kind of financial meltdown being experienced in the high-income European countries and the United States. Developing countries, however, did feel the impact of the global recession as they saw their exports to the United States and Europe fall.

The magnitude of the recession in the United States and globally was so large that the normal mechanisms used to reverse the downturn, mainly monetary policies that lowered interest rates and increased the money supply, were not sufficient to end the

decline. Instead, first the United States, together with several major European economies, had to revert to large expenditures through the government budget to reverse the sharply falling demand for goods and services in the economy. Soon thereafter a number of developing economies also introduced their own stimulus packages in an effort to reverse falling export demand that threatened to drive their economies into recession only a decade after their last crisis. The largest of these developing country stimulus packages was China's. China initiated a program of nearly US$600 billion to be spent mainly on infrastructure, such as a national high-speed highway system, eight new north–south and east–west high-speed rail lines, and new and modernized ports and airports. The stimulus package kept Chinese gross domestic product (GDP) growth from falling despite a large drop in the demand for Chinese exports in 2009. The average rate of GDP growth in China, thanks to a large degree to these government expenditures, was 9.8 percent a year in 2009 and 2010.

When a government raises its expenditures or cuts taxes on individuals and corporations so that they can raise their expenditures, these actions are referred to as **fiscal policy**. When government efforts instead rely on the central bank to adjust the money supply and interest rates to raise or lower individual and corporate expenditures, this is referred to as the exercise of **monetary policy**.

In this chapter we focus on fiscal policy and begin by noting that the main purpose of fiscal policy is not to deal with economic crises, whether crises of inadequate demand leading to recessions as in 2009–10 or of excess demand leading to inflation. The first function of fiscal policy is to manage government expenditures used to fund whatever activities government policy makers and the people they represent think the government should support. The next closely related function of fiscal policy is to find a way of financing those government expenditures, either through the collection of taxes, issuance of government bonds, or in more desperate cases by printing money.

This chapter starts with a discussion of what expenditures are typically undertaken by governments and follows with an analysis of the kinds of taxes commonly used to pay for those expenditures. With this as background, we then return to the question of what happens when something goes wrong either in the public or the private sphere of the economy. What is it that a government can do to repair what has gone wrong?

In many respects these issues are similar in high-, middle-, and low-income economies, but there are significant differences. Developing countries often have to rely more on fiscal policy to curb inflation or recover from a recession because many of the instruments of monetary policy are not as effective or do not even exist. Underdeveloped markets in developing countries are also often a major reason why they rely on government to take responsibility for investments and other expenditures that in a more advanced economy would be handled by the private sector.

The limitations of markets in developing countries would seem to suggest that government expenditures would be a higher share of GDP than in high income countries. But, as the data in Table 11–1 indicate, the opposite is the case. Governments in

TABLE 11-1 Central Government Expenditure as a Share of GDP, 2009

COUNTRY	PERCENT OF GDP
Low-income economies	
Bangladesh	11.3
Cambodia	11.0
Ghana	17.9
Lower-middle-income economies	
Indonesia	15.8
El Salvador	21.6
Morocco	27.9
Upper-middle-income economies	
Colombia	19.4
Thailand	19.7
Mauritius	20.9
High-income economies	
South Korea	21.9
Denmark	42.5
United Kingdom	46.4

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

developing countries on average spend a smaller share of GDP than do high-income countries. A major reason is that high-income countries typically spend a larger share of GDP on social welfare measures (old-age pensions, poverty reduction, healthcare for all) largely because the people of these countries demand such expenditures for reasons of fairness and feel that at their relatively high incomes they can afford to do so. Developing country governments are more apt to focus on measures to increase investment for overall growth. There is, however, considerable variation in the share of government expenditures in GDP among countries at both ends of the income spectrum.

GOVERNMENT EXPENDITURES

At the most general level all countries rely on government for certain activities because of public goods and market failures discussed in Chapter 5. For private goods, such as rice, saris, or TV sets, the signals provided by unfettered competitive market mechanisms guide producers to satisfy consumer demand efficiently. For pure public goods, the market fails entirely. There are few examples of pure public goods; national defense and basic scientific research generally are cited as illustrations. The term *public good* refers to a good or service that exhibits two traits: **nonrival consumption** and **nonexcludability**.

In *nonrival consumption,* one person's use of a good does not reduce the benefits available to others. That being the case, no one has any incentive to offer to pay for the good: If it is available to one, it is available to all. *Nonexcludability* means that it is either impossible or prohibitively expensive to exclude anyone from the benefits once the good is available. In either case, the private market cannot provide the good: Market failure is total. For most public goods, the characteristics of nonrival consumption and nonexcludability are present but less pronounced, and the market can function only in an inefficient way. Examples include vaccination against contagious diseases, police protection, and mosquito abatement. Therefore, it is evident that the appropriate role of the public sector, to a significant degree, is a technical issue.

In practice in all countries, government expenditures designed to offset market failures of this sort account for only a portion of government expenditures even in the most laissez faire, small government–oriented societies. In democratic societies the people through their elected representatives generally want government to do more, such as provide pensions for the elderly and healthcare for the sick regardless of income. In authoritarian societies, the leadership often does much the same because failure to do so can threaten the leaderships' ability to govern.

CATEGORIES OF GOVERNMENT EXPENDITURES

The principal distinction between types of government expenditure is between **capital** or **public investment** expenditures also called **development expenditures**, on the one hand, and **recurrent expenditures** on the other. Development or public investment expenditures are those that, like investment in general, generate a stream of income or services into the future. Government typically takes on the task of building a network of highways, and some governments, particularly in developing countries, take responsibility for providing electric power or exploitation of natural resource wealth (for example, petroleum). Education and health expenditures can also be thought of as a form of investment undertaken by government although in practice government budgets often treat them as recurrent consumption expenditures. The quality of government development investments in these areas should be judged by the same criteria that are used to judge private investments, as was discussed in Chapter 10.

Recurrent expenditures are those expenditures that a government must make year after year. They include the salaries of government employees, the purchase of office equipment, and annual payments to families living in poverty and for pensions for retirees. Most of these categories of recurrent expenditures are self-explanatory, but one category is of particular importance in developing countries largely because it is often ignored. Capital investments to build a highway are often given a high priority but the expenditures needed to maintain the highway once built are often neglected, and the highway is soon filled with potholes. This results not only from the neglect of maintenance by developing country governments but also because foreign aid agencies will frequently support a capital project but will seldom support the

recurrent costs needed to keep the capital project functioning. The problem, however, is not unique to developing countries. The United States has neglected bridge and road maintenance in many parts of the country for decades and, at the time of the 2008–09 economic recession, faced the need for large maintenance and replacement expenditures, some of which were built into the stimulus program designed to get the country out of the recession.

Other common recurrent expenditures in developing (and high-income) countries are interest payments on government debt, consumer subsidies, and subsidies to loss-making state-owned enterprises. Government debt crises are the subject of Chapter 13, and here we note only that the problem with government deficits in developing countries on average is not that they are typically very large, although that is the case in some countries. The problem with chronic large deficits is that interest payments rise over time taking up an increasing share of the government recurrent cost budget either crowding out other worthwhile expenditures or leading to even larger government deficits. If the deficits are financed by borrowing abroad, foreign debt, typically in convertible currencies such as the dollar or the euro, accumulates and eats into the ability of the country to import and may eventually lead to a debt crisis. If the government borrows money domestically to cover the deficit, the interest cost in the budget will also rise. In developing countries, however, borrowing domestically often means borrowing directly from the central bank, and that is essentially the same as printing money and increasing the money supply. The reason developing countries frequently have to borrow from the central bank is that capital markets are underdeveloped and it often is not possible to sell government bonds to the general public or even to private companies. The net result is that deficits lead to money creation, and excessive money creation leads to inflation, which usually proves detrimental to the productive investment so central to economic growth.

Consumer subsidies are also not unique to developing countries. The political logic behind consumer subsidies is that certain commodities are considered essential, and government subsidizes these commodities so as to lower their price to the consumer and thus make them available to all regardless of income. In high-income countries the government typically has the ability to find out the true income of families and then target the subsidy so that it applies only to those below a predetermined poverty line—food stamps in the United States is a good example. In developing countries, however, government usually lacks the administrative capacity to discover the true income of families, although some middle-income nations are gaining this ability (see the discussion of Mexico's Oportunidades program in Chapters 6 and 8). When targeting is difficult, one option is to subsidize all sales of the commodity even though most of the subsidy will go to families that are not poor and could afford to pay the market price. General subsidies of this sort can be very expensive. When a country runs into debt and balance of payments problems and has to call for help to the International Monetary Fund (IMF), one of the first measures the IMF sets as a condition for a loan is elimination of these subsidies. Subsidies, however, are popular

with the public and eliminating them typically leads to demonstrations against the government and even to major riots that threaten the existence of the government.

Even in China, where administrative capacities at the local level are quite high by developing country standards, targeting subsidies to poor families is not possible; subsidies in the 1990s and in the first decade of the twenty-first century have instead gone to designated poor counties and are available to all those in the county, including many families that are not poor. An alternative approach tried in the 1980s in Sri Lanka was to ask families to voluntarily state whether they fell below the poverty line or not. The result was that well over half the population of the country said that they fell below the poverty line even though objective data made it clear that the share below the poverty line was only a fraction of that self-reported figure. In some cases there is not even a pretense that the subsidy is designed to alleviate poverty. Countries with large oil reserves, such as Venezuela and Indonesia, have subsidized refined fuel costs to households even though the heaviest users of such fuel are the well off who can afford automobiles and motor bikes.

Subsidies to cover the losses of state owned enterprises (SOEs) grew rapidly in a great many developing countries when they first achieved independence. This was particularly so in Africa. In Tanzania alone, the number of state enterprises increased 10-fold from 1965 to 1985. By 1980, SOEs were common, and often dominant, in manufacturing, construction, banking services, natural resource industries, and agriculture. Although SOEs typically were small-scale undertakings in most developing countries before 1950, many by the beginning of the twenty-first century were among the largest firms in their countries, and some, notably in China, are among the largest enterprises in their fields anywhere in the world.

Deficits in state enterprises averaged 4 percent of the GDP across all developing countries in the mid-1970s. The problem worsened in the early 1980s, particularly in such countries as Brazil, Costa Rica, the Dominican Republic, Ecuador, Egypt, the Philippines, Turkey, and Venezuela, where SOE deficits reached between 3 and 12 percent of the GDP. In all these countries, the rest of the public sector would have generated a fiscal surplus, excluding the net transfers to the state enterprises.[1] By the late 1980s and through the first decade of the twenty-first century, the drain of SOE subsidies on government budgets has led to efforts by many developing countries to eliminate these subsidies by privatizing their SOEs. International aid agencies frequently have made SOE privatization a condition for further assistance. Privatization, however, has been a slow process, because it has been actively resisted by SOE managers and workers, not to mention politicians who use SOE employment as a patronage machine for their supporters. In countries where a large part of the private sector is controlled by minorities, as is the case with the Chinese minority in Southeast Asia or the South Asian minority in East Africa, the majority population often fears that privatization is simply a way of turning even more of modern business over to the minority.

[1]World Bank, *World Development Report 1988* (New York: Oxford University Press, 1988), p. 171.

Finally transfers from central to subnational governments conventionally are treated as recurrent expenditures even though subnational governments (provinces, departments, counties, municipalities) may use the proceeds for capital formation, such as construction of schools and hospitals. Some countries, such as Indonesia, classify some transfers as capital spending and others as consumption, so international comparisons are difficult. In a few cases, subnational governments with small transfers have access to rich sources of revenue. This is true for Bolivia, where oil-producing provinces receive large oil royalties, and for Malaysia, where the state governments of Sabah and Sarawak have earned substantial tax revenues from the export of tropical timber. In most unitary states, such as Chile, government affairs at virtually all levels are run from the capital, and subnational units of governments have few responsibilities to go with their limited sources of local revenue.

There is often a compelling argument for moving decisions about expenditures to lower levels of government. Lower levels of government are typically closer to the people being served by both investment and recurrent expenditures and thus in principle are better able to design expenditures to best meet local needs. Problems arise, however, when local governments lacking adequate tax revenues under their control finance these project and recurrent expenditures through access to bank loans. The problem can be particularly acute where the local banks are owned or controlled by local politicians. The result can be projects driven more by the political needs of local politicians than the needs of the local population and an increasing often unsustainable local government debt. In the late 1990s local Argentinian governments borrowed heavily to cover local expenditures from both domestic and international sources running up debts neither they nor ultimately the central government had the capacity to repay. Argentina defaulted on this debt in 2002 and negotiated a large-scale debt restructuring in 2005–06. Similar although less severe problems existed in both Brazil in the 1990s, but the government took steps to strictly limit the ability of local governments to invest in this way (see Box 11–1).

In many places, including many federal countries, transfer of funds from the central government to subnational units is essential because the center has monopolized the most productive sources of tax revenue. In most cases, national governments impose income taxes because subnational governments lack the resources and skill required to administer such a complex tax and could not do so effectively anyway because the income tax base easily can migrate within a country. Similarly, in the past the major source of tax revenue for some developing countries, import duties, remained a central government resource because most countries have only a few serviceable ports. Very few developing countries today, however, receive a substantial portion of tax revenue from import duties, in part because import duties have been cut sharply over the years in the various international trade negotiations that have lowered tariffs for all countries.

We have just seen that there are few easy ways to reduce government recurrent expenditure to achieve higher public savings. The scope for doing so often was

BOX 11-1 REINING IN FISCAL DECENTRALIZATION IN BRAZIL AND CHINA

BRAZIL

Brazil has a federal system of government, and the states have a substantial role in capital and recurrent expenditures. The states in the past also had banks owned by their governments, and these banks contributed in a major way to large-scale profligate state spending over many decades. There were also problems involving borrowing by the central government from the central bank. In 1996, however, the central government took steps to increase the independence of the central bank and began a program to separate the banks from control by state governments. The first phase of this effort involved an effort to privatize state-owned banks in what was called the Program to Reduce State Involvement with Banking Activities. The central government helped with the privatization process, and many of the state banks were privatized and others turned into development agencies. Then in 2000 the government passed the Law of Fiscal Responsibility that prohibited government borrowing from banks it owned or from the central bank. All new borrowing by either the central government or state governments first had to be approved by the central bank and by the senate. In 2002 governments were further prohibited from borrowing within 180 days of the end of a government's term in office. This restrained the large expenditures during the final days of an administration often designed to pay off supporters knowing that the incumbent governor or president would not have to find ways of repaying the loans involved. There were also steps taken to reduce the banks' incentive for making such loans. These efforts have played an important role both in reducing wasteful government expenditures and in reining in the chronic inflation that had plagued Brazil for decades.

CHINA

China's provincial governments are not formally independent of the central government but they have had a substantial degree of autonomy in many economic areas. Formally the largest state-owned commercial banks are national not provincial banks, but in practice local governments and the local branches of the Communist Party have had a major say in appointments to head the local branches of these national banks. The result in the 1990s was that local governments borrowed heavily from the banking system to finance local projects and to support investments of provincial- and city-owned state industrial enterprises, which led to large-scale lending partly to prop up failing state-owned enterprises that fueled bursts of inflation. Inflation was stopped but only when the central government and the central bank imposed lending quotas on state-owned banks and their

branches. The government, however, did take steps to change the territory covered by the branches of the large commercial banks so that those territories did not coincide with that of the provincial and local governments, making it more difficult for local politicians to influence the selection of local bank managers.

During the large-scale investment surge that began with the 1997–98 financial crisis in Asia and elsewhere, the Chinese government began a massive state-led investment program that continued throughout the first decade of the twenty-first century, and local governments were pressured to carry out much of this investment program. Many of these investments were well thought out, but local governments lacked the tax base to pay for them and so borrowed heavily from the state-owned commercial banks and local state-owned banks. The result by 2011 was a local debt accumulation of over RMB 10 trillion, which was 25 percent of the gross domestic product (GDP). Given the weakness of the tax base of the local governments, there was a danger that the state-owned commercial banks would again be burdened by a large number of nonperforming loans requiring refinancing by the central government.[a]

[a]Anwar Shah, "Fiscal Decentralization and Macroeconomic Management," *International Tax and Public Finance* 13 (2006), 437–62.

greatest in the areas of military spending and subsidies to cover deficits of state-owned enterprises. Subsidies to state-owned enterprises, however, have been cut over the years either through privatization or by taking steps to make state enterprises competitive and profitable. Military expenditures in developing countries in the twenty-first century are still 2 percent of the GDP on average, with considerable variation across nations, which leaves some room for cutting but not a great deal. The main argument for cutting these latter expenditures is that military forces in developing countries are sometimes a greater danger to their own people than to potential enemies of the nation.

GOVERNMENT REVENUE AND TAXES

Government expenditures have to be funded or paid for. Most government expenditures are funded by revenue from taxes, including **direct taxes** (taxes on corporate or individual income and other payroll taxes such as Medicare and Social Security payments in the United States) or **indirect taxes** (sales taxes, import duties, and value-added taxes that are typically levied as a percentage of the value of a good being sold or imported), but there are other sources of funding. Mineral-rich countries often depend on royalties paid to the government by large petroleum and mineral mining companies, bonds are sometimes issued to cover the costs of a capital investment,

and fees are charged for many government-provided services. Revenue can also come through aid grants from other countries or in the form of loans from international lending agencies such as the World Bank.

As incomes per capita rise, the amount of taxes collected, expressed as a share of GDP, also tends to rise, although there is considerable variation among countries as the data in Table 11–2 indicate. The taxes reported in the table do not include what are often referred to as "social contributions"—that is, taxes paid for old-age support such as the Social Security and Medicare systems in the United States. Because high-income nations have more extensive transfer programs of this kind, they also collect more in taxes to finance them. If these taxes were included in the data, the share of taxes out of GDP would rise with country income level in an even more pronounced fashion.

The relationship between the kinds of taxes collected is also associated with per capita income. For countries with low incomes and in the early stages of development, there is typically much greater reliance on international trade taxes. For the three low-income economies in Table 11–2, at least one quarter of tax revenues is obtained from taxes on international trade. As incomes rise, reliance on taxes on international trade falls off significantly. Taxes on income, personal and corporate,

TABLE 11–2 Composition of Tax Systems by Type of Tax,* 2009

		COMPOSITION OF TAXES (PERCENT OF GDP)			
COUNTRY/ECONOMY	TOTAL TAXES	DOMESTIC INCOME TAXES	TAXES ON GOODS AND SERVICES	TAXES ON INTERNATIONAL TRADE	OTHER TAXES
Low-income economies					
Bangladesh	8.6	2.2	3.3	2.8	0.4
Cambodia	9.6	1.7	5.4	2.5	0.0
Ghana	12.5	4.2	5.4	2.9	0.0
Lower-middle-income economies					
Indonesia	11.4	5.7	4.8	0.3	0.6
El Salvador	12.5	4.7	6.8	0.9	0.0
Morocco	23.8	9.5	10.5	2.0	1.8
Upper-middle-income economies					
Colombia	11.8	4.5	5.5	0.9	0.8
Thailand	15.2	7.1	7.2	0.9	0.1
Mauritius	18.6	5.4	11.0	0.5	1.7
High-income economies					
South Korea	15.5	6.6	6.0	0.9	2.1
United Kingdom	26.0	13.3	10.1	0.0	2.5
Denmark	34.6	18.0	14.7	0.0	2.0

*Taxes refer to the central government only and are exclusive of social contributions–that is, to programs of old age support and so on.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

BOX 11-2 TAX RATES AND SMUGGLING: COLOMBIA

In the classic case from Colombia, the import duty rate on cigarettes was over 100 percent, and it was virtually impossible to purchase duty-paid cigarettes. At such high rates, import duty collections on cigarettes were nil and the market was flooded with smuggled foreign brands. In 1969, the duty rate was reduced to 30 percent. Cigarette smuggling on the poorly policed Caribbean coast of that country continued, but duty-paid packages began to appear in the mountainous interior, and duty collections soared. Smuggling profits possible under a 30 percent duty were no longer high enough to compensate smugglers for the risks of arrest.

provide about half of the tax revenue of Denmark, Korea and the UK; far less in Bangladesh and Cambodia.

In the sections that follow, we review the different kinds of taxes and explain why some taxes are favored over others at different stages of development.

TAXES ON INTERNATIONAL TRADE

Although reliance on the taxes of foreign trade has fallen sharply in recent decades, particularly in middle-income countries, tax revenue structures in developing countries historically have depended heavily on import duties. Many low-income countries remain markedly dependent on import duties. This is particularly so for countries like the Gambia, Lesotho, Côte D'Ivoire, Togo, and Bangladesh, where as much as one quarter to one third of the total government revenue comes from taxes on imports.

For most countries, attempts to raise revenues through higher duties are unfeasible and undesirable on economic grounds. Higher import duties intensify the incentive for smuggling or evading tariffs. Various studies have shown that, for countries with already high duty rates, the incentive to smuggle increases disproportionately with further increases. Therefore, a 10 percent rise in duty rates can result in an increase in smuggling activity by more than 10 percent (Box 11–2). In mountainous countries, such as Afghanistan and Bolivia, or archipelago countries, such as Indonesia and the Philippines, borders are especially porous to smuggled imports. Singapore during the first four decades after independence did not even report its trade with Indonesia largely because it did not want Indonesia to know how much trade between the two countries went through informal (smuggling) rather than official channels.[2]

[2]Bill Guerin, "Indonesia-Singapore Gap More Than Numbers," *Asia Times,* June 27, 2003. Available online at www.atimes.com/atimes/Southeast_Asia/EF27Ae03.html, accessed February 29, 2012.

Export taxes are constitutionally prohibited in the United States and extremely rare in other industrial countries. But export taxes are not uncommon in developing countries, particularly in tropical Africa and Southeast Asia. Taxes ordinarily are imposed on exports of raw materials and have been applied on timber (Ivory Coast and Liberia), tin (Malaysia), jute (Pakistan), and diamonds (Botswana) as well as on foodstuffs, such as coffee (Colombia), peanuts (the Gambia and Senegal), cocoa (Ghana), and tea (Sri Lanka).

Over the past few decades, export taxes have declined in importance as various economic organizations, notably the World Trade Organization (WTO), have put pressure on new members and some old ones to eliminate such taxes or at least not use them to promote particular industries. Some economic groupings such as the European Union and Mercosur ban export taxes. Export taxes often are imposed in the belief that they are paid by foreign consumers. That is, the taxes themselves are thought to be exported to consumers abroad, along with the materials. But the conditions necessary for exporting taxes on exports to foreign consumers rarely are present.

Export taxes also are employed to promote nonrevenue goals, including increased processing of raw materials within natural resource–exporting developing countries. This is done by imposing high rates of export tax on unprocessed exports (cocoa beans or logs) and lower or no rates of tax on processed items fabricated from raw materials (chocolate and plywood). In principle, this use of export taxes should increase the local value added on natural resource exports and thereby generate greater employment and capital income for the local economy. Unfortunately, in many cases, the result has been that a government gives up more in export tax revenues than its country gains in additional local value added, particularly when processed raw materials are exported tax free. One study documents several instances in Southeast Asia and Africa in which the additional value added gained in heavily protected local processing of logs into plywood typically was less than half the amount of export tax revenue that would have been collected had timber been exported in the form of logs.[3]

SALES AND EXCISE TAXES

A much more promising source of government revenue is indirect taxes on domestic transactions, such as sales and excise taxes. **Sales taxes**, including **value-added taxes**, are broad-based consumption taxes imposed on all products except those specifically exempted, usually food, farm inputs, and medicine. **Excise taxes** also are taxes on consumption, but these levies are imposed only on specifically enumerated items, typically tobacco, alcoholic beverages, gambling, and motor fuel. On average, developing countries depend on domestic commodity taxes for a substantial share,

[3]Robert Repetto and Malcolm Gillis, eds., *Public Policies and the Misuse of Forest Resources* (New York: Cambridge University Press, 1988), chap. 10.

ranging from a bit over a quarter (the poorest countries plus the United States) to over half of total revenues (Table 11–2).

Virtually every developing country imposes some form of sales tax. In most, the tax is not applied to retail sales because of the burdensome administrative requirements of collecting taxes from thousands of small retailers. In the past, as in Chile before 1970 and in some Indian states, the tax was imposed as a gross turnover tax collected at all levels of production and distribution, with harmful implications for efficiency, income distribution, and virtually every objective of tax policy. In developing countries, administrative problems are more tractable when the sales tax is confined to the manufacturing level: A much smaller number of firms is involved, and the output of manufacturers is far more homogeneous than sales of retailers or wholesalers. For these reasons many low-income countries have used either the single-stage or the value-added form of manufacturers' tax, usually exempting very small producers. This kind of sales tax, however, involves more economic distortions than either a wholesale or a retail tax, and for that reason as well as for revenue motives, more and more middle-income countries have turned to taxes at the retail level.[4]

Increasingly, these taxes have taken the form of one or another variant of the value-added tax (VAT), widely seen as the most effective method of taxing consumption yet developed. More than 60 countries, including all members of the European Union, have adapted the VAT since it was first adopted in its comprehensive retail form in Brazil in 1967.[5] More than 40 developing countries had adopted the tax by 1994, either to replace older, outdated forms of sales taxes or to allow them to reduce reliance on harder-to-collect income and property taxes. The typical VAT tax rate varies from around 10 percent in some of the poorest countries to 15 to 20 percent in most of the others.

Excise taxes might appear to represent an ideal source of additional tax revenue. These typically are imposed on sumptuary items having relatively inelastic demand. When the price elasticity of demand for such products is very low (as for tobacco products) or relatively low (as for alcoholic beverages), an increase in excise tax rates will induce little reduction in consumption of the taxed good. If price elasticity is as low as –0.2, not uncommon for cigarettes in part because smoking tobacco has been shown to be addictive, then an additional 10 percent excise tax on this product would yield an 8 percent increase in tax revenues. Moreover, it is a hallowed theorem in optimal tax theory that taxes levied on items with inelastic demand and supply involve the smallest losses in economic efficiency, or what is the same thing, the least excess burden. **Excess burden** is defined as a loss in private welfare above

[4]For a full discussion of the distortions involved in different forms of sales tax, see John Due and Raymond Mikesell, *Sales Taxation: State and Local Structure and Administration* (Baltimore: Johns Hopkins Press, 1983).

[5]Carl Shoup, "Choosing among Types of VAT," in Malcolm Gillis, Carl Shoup, and Gerry Sicat, eds., *Value-Added Taxation in Developing Countries* (Washington, DC: World Bank, 1990).

the amount of government revenue collected from a tax.[6] Further, many agree, with much justification, that consumption of both tobacco and alcohol should be discouraged on health grounds.

PERSONAL AND CORPORATE INCOME TAXES

Harried ministers of finance, perceiving slack in personal and corporate income taxes, often resort to rate increases in these taxes, with no change in the tax base. The results usually are disappointing, particularly for the personal income tax. Even in middle-income countries, only a small proportion of the population is covered by the personal income tax; in contrast, the vast majority of the adult population in the United States files income tax returns each year (although just under half of all U.S. households were required to pay federal income taxes as of 2011).[7] Therefore, few developing countries can rely heavily on the personal income tax for revenues. Whereas the personal income tax accounts for over 40 percent of all federal taxation in the United States (not including Social Security and Medicare payments), rarely does the personal income tax account for close to that much of total central government revenues in developing countries. In low- and middle-income nations, when personal income taxes are applied, they are paid largely by urban elites. Not only are these groups usually the most vocal politically but they have developed such a variety of devices for tax evasion and avoidance that rate increases stand little chance of raising additional revenues. (The same is often true in high-income economies.)

Rate increases for corporate taxes are usually not productive either. In only 16 of 82 developing countries did the corporation income tax account for more than 20 percent of total taxes in the 1980s. In 2006–08, corporate taxes and personal income taxes taken together accounted for more than 30 percent of all revenue in less than half of a sample of more than 60 developing countries. And, in many of those countries, corporate tax collections often originated with foreign natural-resource firms.

NEW SOURCES OF TAX REVENUES

If higher rates on existing taxes are unlikely to raise revenues much, a second option for increasing public savings is to tap entirely new sources of tax revenues. In many developing countries, whether by accident, design, or simply inertia, many sources of tax revenue may have been overlooked entirely. Many countries have not collected taxes on motor vehicle registrations; some have not used urban property taxes as a significant source of revenue; many have not applied corporation income taxes to the income of state-owned enterprises. Kenya, for example, did not seriously tax

[6]For a further discussion of optimal taxation and excess burden, see Figure 11–1 and Joseph Stiglitz, *Economics of the Public Sector* (New York: Norton, 1988).

[7]Rachel Johnson, Jim Nunns, Jeff Rohaly, Eric Toder, and Roberton Williams. "Why Some Tax Units Pay No Income Tax," Tax Policy Center, Urban Institute, July 2011.

farmland, and a few countries, such as Indonesia before its 1984 tax reform, did not collect personal income taxes on the salaries of civil servants. China began to experiment with the use of urban property taxes only in the first decade of this century.

CHANGES IN TAX ADMINISTRATION

A far more significant option for increasing tax revenues is to implement changes in tax administration that permit more taxes to be collected from existing tax sources, even at unchanged tax rates. The potential for increased revenues from such action is very large and seldom realized in virtually all developing countries. Shortages of well-trained tax administrators, excessively complex tax laws, light penalties for tax evasion, corruption, and outdated techniques of tax administration combine to make tax evasion one of the most intractable problems of economic policy in developing countries.

The case studies given in Box 11–3 discuss the magnitude of the problem in India and Bolivia, which is typical of many, perhaps most, other developing countries. These examples suggest that efforts to collect a greater share of the taxes due under the current law can increase revenues substantially. But the kinds of administrative reforms required are difficult to implement and especially difficult to sustain. Even Bolivia's successful reform raised its tax ratio to only 7 percent of GDP. Korea at the beginning of its rapid development period, a country noted for its efficient and determined administration, failed in the 1960s to reach its goal of increasing collections by 40 percent through more effective enforcement, although it was able to reduce underreporting of nonagricultural personal income from 75 percent to slightly under 50 percent. Administrative reform can help and is important at any stage of development; however, better tax administration tends to improve with economic development.

FUNDAMENTAL TAX REFORM

The final policy option available for increasing tax revenues is the most difficult to implement but the most effective when it can be done. Fundamental tax reform requires junking old tax systems and replacing them with completely new tax laws and regulations. Implementing tax reform engenders enormous technical and informational, not to mention political, difficulties in all countries. In general, governments resist genuine efforts to reform the tax structure until a fiscal crisis in the form of massive budgetary deficits threatens. Even during a fiscal crisis, it is difficult to mobilize a political consensus to allow unpopular tax measures to pass. Tax policies that protect favored groups and distort the allocation of resources did not just happen; more likely, they were enacted at the behest of someone, ordinarily the privileged and the powerful.

Probably, more has been said, to less effect, about tax reform than almost any topic in economic policy. This is no less true for the United States than for the scores of developing countries where major tax reform efforts have been mounted. That the process is painful and slow is evident from the experience of several countries: In the

BOX 11-3 TAX ADMINISTRATION IN INDIA AND BOLIVIA IN THE 1980S

INDIA

India during the 1980s provides one of the more egregious examples of poor tax administration. During the fiscal year 1981, for example, it was found that taxes were avoided on at least 40 and possibly 60 percent of potentially taxable income. Almost 70 percent of the taxpayers covered by a survey openly admitted bribing tax officials, and three quarters of the tax auditors admitted accepting bribes to reduce tax payments. The cost of a bribe was commonly known to be about 20 percent of the taxes avoided.

BOLIVIA

In Bolivia during the same period, the tax system was chaotic. More than 400 separate taxes were levied by the national, regional, and city governments. Taxpayer records were out of date, and tax collections were recorded more than a year after they had been paid. Of 120,000 registered taxpayers, one third paid no taxes at all, while 20,000 taxpayers who did pay were not on the register. The administration had all but ceased its data processing. Administration had gotten so bad that the government was helpless to collect taxes during the hyperinflation of the mid-1980s. Tax collections fell to only 1 percent of the GDP. Faced with this chaos, the Bolivian government reformed its tax administration, simplifying it drastically. This, together with price stabilization, enabled the government to collect revenues equivalent to over 7 percent of the GDP by 1990, a low ratio by world standards but a vast improvement for Bolivia.

Sources: Omkar Goswami, Amal Sanyal, and Ira N. Gang, "Taxes, Corruption and Bribes: A Model of Indian Public Finance" (201–13), in Michael Roemer and Christine Jones, eds., *Markets in Developing Countries: Parallel, Fragmented and Black* (San Francisco: ICS Press, 1991); Carlos A. Silvani and Alberto H. J. Radano, "Tax Administration Reform in Bolivia and Uruguay" (19–59) in Richard M. Bird and Milka Casanegra de Jantscher, eds., *Improving Tax Administration in Developing Countries* (Washington, DC: International Monetary Fund, 1992).

United States, the time lag between the birth of tax innovations (tax credit for child-care expenses, inflation proofing of the tax system) and their implementation usually is at least 15 years. If anything, the lag may be slightly shorter in developing countries.

In spite of the difficulties involved, some countries were able to carry through fundamental reforms in tax structure and administration before 1980. The classic example is Japan in the 1880s, when that society began its transformation to a modern industrial power. Korea implemented a major tax reform program in the early 1960s, as did Colombia and Chile in the 1970s and Indonesia in the 1980s (see Box 11–4).

BOX 11-4 INDONESIAN TAX REFORM

In the 1970s Indonesia was a major oil exporter and received more than half of its government revenue from royalties paid by the large oil companies. The increase in petroleum prices caused by the Organization of Petroleum Exporting Countries (OPEC) in 1973 and 1979 increased these royalties to an unprecedented share of the government budget. The Indonesian minister of finance at the time, however, knew that what goes up (in this case petroleum prices) can also come down and he wanted to put Indonesia's government finances on a more stable long term basis. Non-oil sources of tax revenue, however, were based on laws that in many cases went back to the Dutch colonial period and produced limited revenue for the government and did so inefficiently.

In 1981 the ministry of finance began a major effort to completely reform and modernize the Indonesian tax system. The goal of this effort was to create a broader base for the generation of tax revenue, increase the equity of the fiscal system, and generally bring the Indonesian system in line with international best practice. The process involved the complete rewriting of the tax laws, but it also involved fundamental changes in the way taxes were administrated and enforced together with the creation of a computerized tax information system among other major changes.

A team of 25 international experts, including economists, lawyers, accountants, and data processing experts, was put together to assist 10 Indonesian government officials in this endeavor. The role of the team was to help determine what aspects of best tax practice would work best in the Indonesian context. The experts then would present their findings for a full discussion of whether the expert recommendations would in fact work in the Indonesian context in meetings attended by the minister of finance and all of the ministry officials with relevant responsibilities in this area. When this process was completed and necessary adjustments to the recommendations made, the proposal was put first to the president of Indonesia and then to parliament, which passed the legislation in late 1983. Although Indonesia had an authoritarian government at that time, powerful rent-seeking individuals both inside and outside of the government could have had a major influence on the shaping of tax reform as happens in so many countries, but in this case the recommendations that were passed largely reflected international best practice, not the special interests of rent seekers.

The key reforms were the introduction of a low-rate simplified value-added tax together with major changes in the personal and corporate income tax codes. Nonpetroleum government revenues as a result of these reforms rose from

6 percent of Indonesia's GDP just before the passage of the reforms to 10 percent of GDP by the early 1990s. This increase in revenue had much to do with the fiscal stability that Indonesia enjoyed when petroleum prices and government royalties from petroleum fell in 1986.

Source: Michael Roemer and Joseph Stern, "Indonesia: Economic Policy Reform," (63–64), in Dwight H. Perkins, Richard Pagett, Michael Roemer, Donald Snodgrass, and Joseph Stern, eds., *Assisting Development in a Changing World: The Harvard Institute for International Development, 1980–1995* (Cambridge: Distributed by Harvard University Press, 1997).

After 1980, the pace of tax reform quickened notably, in both developing and industrial countries, with many similarities in the various reform programs. Throughout much of the postwar period, tax systems commonly were fine-tuned to achieve a wide variety of nonrevenue objectives. In particular, governments in developed and developing nations alike commonly sought substantial income redistribution through the use of steeply progressive tax rates. Also, complex and largely impossible to administer systems of tax incentives were widely used in attempts to redirect resources to high-priority economic sectors and promote foreign investment, regional development, and even stock exchanges.

While fine-tuning tax systems someday may yield the desired results, it requires, at a minimum, strong machinery for tax administration and traditions of taxpayer compliance. Within developing countries, at least, there has been growing recognition that such conditions seldom prevail. Governments increasingly have turned away from reliance on steeply progressive tax rates and complicated, costly tax incentive programs.

The 1980s through the beginning of the twenty-first century saw a worldwide movement toward an entirely different type of tax system, with a shift toward vastly simplified taxes imposed at much flatter rates and with much broader bases[8] and increasingly greater reliance on consumption rather than income taxation. Tax reform programs in Bolivia, Chile, Colombia, India, Indonesia, Jamaica, and Malawi exemplify most of these trends.[9]

Two aspects of this worldwide movement in tax reform are especially salient. First, the top marginal income tax rates of 60 to 70 percent were not uncommon from 1945 to 1979. But, since 1984, country after country has slashed the top marginal rate, often substantially. Table 11–3 shows this phenomenon in a selection of developing countries. During the same period, many industrial nations, ranging from Australia and Austria through the United Kingdom and the United States also cut the top rate sharply.

[8]Joseph A. Pechman, ed., *World Tax Reform: A Progress Report* (Washington, DC: Brookings Institute, 1988), p. 13.

[9]A number of these cases are discussed in Malcolm Gillis, ed., *Tax Reform in Developing Countries* (Durham, NC: Duke University Press, 1989).

TABLE 11–3 Countries Reducing Highest Rates of Income Tax, 1984–2009

COUNTRY	1984–85	2009
Low-income economies		
Tanzania	75	30
Uganda	70	30
India	62	30
Ghana	65	25
Lower-middle-income economies		
Papua New Guinea	50	42
Pakistan	60	20
Philippines	60	32
Sri Lanka	55	35
Indonesia	45	35
Peru	50	30
Guatemala	42	31
Jamaica	58	25
Upper-middle-income economies		
Botswana	75	25
Brazil	60	27.5
Costa Rica	50	15
Colombia	49	33
Thailand	65	37
Argentina	45	35
Mexico	55	28
Malaysia	55	27
Chile	56	40

Sources: George J. Yost III, ed., *1994 International Tax Summaries, Coopers & Lybrand International Tax Network* (New York: Wiley, 1994); Glenn P. Jenkins, "Tax Reform: Lessons Learned," in Dwight H. Perkins and Michael Roemer, eds., *Reforming Economic Systems in Developing Countries* (Cambridge: Harvard Institute for International Development, 1991); Denise Bedell, "Personal Income Tax Rates," *Global Finance*, available at www.gfmag.com/tools/global-database/economic-data/10442-personal-income-tax-rates.html#axzz1YDeg5nMb, accessed February 2012.

In many of these cases, deep cutbacks in the highest tax rates were accompanied by reforms involving a very substantial broadening of the income tax base, through the reduction of special tax incentives, abolition of tax shelters, and the like. This pattern was especially notable in Bolivia, Colombia, Indonesia, Jamaica, and Sri Lanka, so that, even with a rate reduction, higher-income groups often ended up paying a higher proportion of total taxes than before.

Reasons for the worldwide shift toward lower tax rates on broader income tax bases are not difficult to find. First, income taxes imposed at high marginal rates have proven difficult or impossible to administrate, even in wealthy countries such as the United States. With high marginal tax rates, the incentives to evade taxes (through concealment of income) or avoid taxes (by hiring expensive legal talent to devise tax shelters) are very high. Second, the growing mobility of capital across international boundaries has meant that the risk of capital flight from a particular country increases

when that country's top rates of income tax exceed those prevailing in industrial nations, where tax rates have been falling.[10] Third, the operation of the income tax systems of such developed nations as the United States and Japan has placed downward pressure on the tax rates everywhere. This is because of the *foreign income tax credit,* wherein a country like the United States allows foreign income taxes to be credited (subtracted) from U.S. taxes due on income repatriated from abroad. This credit could be used, however, up to only the amount of tax payable at U.S. rates, which was reduced from 50 to 28 percent in 1986. Finally, high marginal rates of income tax did not prove to be particularly efficacious in correcting severe inequalities in income distribution in either rich or poor countries.

The second striking feature of recent tax reforms worldwide has been the steadily growing number of countries adopting the value-added tax. Several factors account for the popularity of the VAT: The two most important are its reputation as a money machine and its administrative advantages, relative to other forms of sales taxes and income taxes.[11] The record of the VAT in generating large amounts of revenue quickly and in a comparatively painless fashion has given it a reputation as an efficient way to generate money. Although this reputation stems largely from the experience in European countries, the record in developing countries does lend some support to the alleged revenue advantages of the VAT. In Indonesia, the 4 percent share of the value-added tax in GDP in 1987 was nearly three times the share garnered in 1983 by the taxes it replaced. Notwithstanding the marked revenue success of the VAT in nations such as Brazil, Chile, and Indonesia, its reputation as a money machine for many other countries appears to have been at least slightly overstated.

In the 1980s, it was common to hear the claim that the VAT was largely self-administering. This is not so, but the tax-credit type of VAT has three principal advantages over single-stage retail and nonretail sales taxes in limiting the scope for evasion. First, the VAT is self-policing to some extent because underpayment of the tax by a seller (except, of course, a retail firm) reduces the tax credit available to the buying firm. Even so, firms that also are subject to income taxes have incentives to suppress information on purchases and sales to avoid both the value-added and income taxes. Also, this possible advantage of the VAT is diminished when evasion at the final (retail) stage of distribution is endemic. Second, cross-checking of invoices enables the tax administration to match invoices received by purchasers against those retained by sellers. The cross-check feature is a valuable aid in audit activities but no substitute for a true, systematic audit. Third, that a large share of the VAT is collected before the retail level is an advantage particularly because, in most developing countries, an abundance of small-scale retail firms do not keep adequate

[10]For discussion of the implications for taxation of growing international mobility of financial and physical capital, see Dwight R. Lee and Richard B. McKenzie, "The International Political Economy of Declining Tax Rates," *National Tax Journal* 42, no. 2 (March 1989), 79–87.

[11]For a full discussion of these reasons, see Alan A. Tait, *Value-Added Tax* (Washington, DC: International Monetary Fund, 1988), chap. 1.

records. In sum, the administrative advantages of the VAT are very real, if sometimes exaggerated by enthusiastic proponents.[12]

TAXES AND INCOME DISTRIBUTION

For decades, developed and developing countries alike have sought to use the fiscal system, particularly taxation, to redress income inequalities generated by the operation of the private market. Social philosophers from John Stuart Mill and eminent nineteenth-century Chilean historian Francisco Encina to twentieth-century philosophers such as John Rawls have sought to establish a philosophical basis for income redistribution, primarily through progressive taxes. No scientific basis is used to determine the optimal degree of income redistribution in any society. And across developing countries different views prevail as to the ideal distribution of income. But, in virtually all countries, the notion of **fiscal equity** permeates discussions of budgetary operations. In the overwhelming majority of countries, fiscal equity is typically defined in terms of the impact of tax and expenditure policy on the distribution of economic well-being. Progressive taxes, those that bear more heavily on better-off citizens than on poor ones, and expenditures whose benefits are concentrated on the least advantaged are viewed as more equitable than regressive taxes and expenditures.

On the tax side of the budget, the materialistic conception of equity requires that most taxes be based on the **ability to pay**. The ability to pay can be measured by income, consumption, wealth, or some combination of all three. Clearly, individuals with higher incomes over their life spans have a greater ability to pay taxes, quite apart from the moral question of whether they should do so. Indeed, the redistribution impact of taxation almost always is expressed in terms of its effects on income. However, philosophers since the time of Thomas Hobbes, a seventeenth-century English philosopher, have argued that consumption furnishes a better index of ability to pay than income; in this view, tax obligations are best geared to what people take out of society (consume) rather than what they put into society (as measured by income).

In practice, developing countries have relied heavily on these two measures of ability in fashioning tax systems. Personal and corporate income taxes employ income as the indicator; sales taxes and customs duties are indirect assessments of taxes on consumption. At a minimum, equity usually is assumed to require the avoidance of regressive taxes whenever possible. A number of tax instruments have been employed to secure greater progressivity in principle, if not in practice; all suffer from limitations to one degree or another.

[12]For a succinct summary of some of these issues, see John F. Due, "Some Unresolved Issues in Design and Implementation of Value-Added Taxes," *National Tax Journal* 42, no. 4 (December 1990), 383–98.

PERSONAL INCOME TAXES

The most widely used device for securing greater progressivity has been steeply progressive rates under the personal income tax. In some countries in some periods, nominal or legal marginal income tax rates have reached very high levels, even for relatively low incomes. Thus, for example, tax rates applicable to any income in excess of $1,000 in Indonesia in 1967 reached 75 percent, largely because tax rates were not indexed to rapid inflation; in Algeria in the 1960s, all income in excess of $10,000 was subject to marginal tax rates of nearly 100 percent; Tanzania imposed top marginal rates of 95 percent as late as 1981. Although, in most developing countries, marginal income tax rates are considerably lower than the preceding examples and, as is apparent from Table 11–3, have been falling, some countries still attempt to impose rates in excess of 50 percent, such as Angola, Gabon, and Cuba, but most countries with marginal rates at this level in the twenty-first century are high-income European states with large social welfare programs.

If the tax administration machinery functioned well and capital were immobile among countries, the pattern of actual tax payments of high-income taxpayers would resemble the legal, or theoretical, patterns just described. In fact, in most countries, **effective taxes** (the taxes actually collected as a percent of income) fall well short of theoretical liabilities. Faced with high income tax rates, taxpayers everywhere tend to react in three ways: (1) They evade taxes by concealing income, particularly capital income not subject to withholding arrangements; (2) they avoid taxes by altering economic behavior to reduce tax liability, whether by supplying fewer labor services, shipping capital to tax havens abroad, or hiring lawyers to find loopholes in the tax law; and (3) they bribe tax assessors to accept false returns.

For all these reasons, the achievement of substantial income redistribution through progressive income taxes has proven difficult in all countries, including the United States and the three Scandinavian nations where tax rates were long among the world's most progressive. Tax avoidance is the favored avenue for reducing tax liability in the United States, where use of the other methods can result in imprisonment. But, where tax enforcement is relatively weak, particularly where criminal penalties for evasion are absent and tax officials deeply underpaid, tax evasion and bribery are used more commonly. The scope for substantial redistribution through the income tax is, therefore, even more limited in developing countries than in the United States or Sweden.

Notwithstanding these problems, a significant share of the income of the wealthiest members of society is caught in the income tax net in many developing countries. Revenues from personal income tax collections in countries such as Colombia, South Korea, and Chile have been as high as 15 percent of total taxes and in a few others have run between 5 and 10 percent of the total. In virtually all developing countries, the entirety of such taxes is collected from the top 20 percent of the income distribution. This means, of course, that the very presence of an income tax, even one imposed at proportional rather than progressive rates, tends to reduce income inequality.

TAXES ON LUXURY CONSUMPTION

In view of the difficulties of securing a significant redistribution through income taxes, countries may employ heavy indirect taxes on luxury consumption as a means of enhancing the progressivity of the tax system. Efforts to achieve this goal usually center on internal indirect taxes, such as sales, and on customs duties on imports, but not excises on tobacco and alcohol.

Several developing countries have found that, provided tax rates are kept to enforceable levels, high rates of internal indirect taxes on luxury goods and services, coupled with lower taxes on less income-elastic items, can contribute to greater progressivity in the tax system. For revenue purposes, countries typically impose basic rates of sales taxes on nonluxuries at between 4 and 8 percent of manufacturers' values. This is equivalent to retail taxes of between 2 and 4 percent because taxes imposed at this level exclude wholesale and retail margins. Food, except that consumed in restaurants, almost always is exempted from any sales tax intended to promote redistributive goals. In developing countries, the exemption of food by itself renders most sales taxes at least faintly progressive, given the high proportion (up to 40 percent in many middle-income countries and over 50 percent in low income countries) of income of poor households spent on food. Sales taxes involving a limited number of luxury rates of between 20 and 30 percent at the manufacturers' level have been found to be workable.

The redistributive potential of sales tax rates differentiated in this way, however, is limited by the same administrative and compliance constraints standing in the way of the heavier use of income taxation in developing countries. While sales taxes are not as difficult to administer as income taxes, they do not collect themselves. A manufacturer's sales tax system employing three or even four rates may be administratively feasible in most countries, even when the highest rate approaches 40 percent. Rates much higher than that or reliance on a profusion of rates in an attempt to fine-tune the tax lead to substantial incentives and opportunities for tax evasion. Jamaica had over 15 rates before 1986, and Chile had over 20 from 1960 to 1970. In recognition of these problems, Indonesia adopted a flat-rate manufacturers' tax in 1985: The tax applied at a rate of 10 percent on *all* manufactured items and imports. The tax nevertheless was slightly progressive because it did not apply to items that did not go through a manufacturing process, including most foodstuffs consumed by low-income families.

Although the use of internal indirect taxes, such as sales taxes, can contribute to income redistribution goals without causing serious misallocation of resources, the same cannot be said for the use of customs duties. Sales taxes are imposed on all taxable goods without regard to national origin, including goods produced domestically as well as abroad. Tariffs apply only to imported goods. Virtually all countries, developed and developing, use customs duties to protect existing domestic industry. Developing countries in particular employ customs duties as the principal means of

encouraging domestic industry to produce goods that formerly were imported. This strategy, called **import substitution**, is examined at length in Chapter 18.

Deliberate policies to encourage import substitution through the use of high protective tariffs, under certain conditions, might lead to results sought by policy makers. But accidental import substitution arises when tariffs are used for purposes other than protection, and this is unlikely to have positive results. Many countries, as already pointed out, use high tariffs to achieve heavier taxation of luxury consumption. Often heavy tariffs are imposed on imported luxury items for which there is no intention of encouraging domestic production. Therefore, many Latin American and some Asian countries have levied customs tariffs of 100 to 150 percent of value on such appliances as electric knives, hair dryers, sporting goods, and mechanical toys. For most countries, these items are clearly highly income elastic and apt candidates for luxury taxation.

But efforts to tax luxuries through high customs duties lead to unintended, and almost irresistible, incentives for domestic production or assembly of such products. In virtually all countries, save the very poorest, alert domestic and foreign entrepreneurs have been quick to seize on such opportunities. By the time local assembly operations are established, they usually can make a politically convincing case that the duties should be retained to enable local production to continue, even when the value added domestically is as low as 10 percent of the value of the product. Such operations, if subject to any local sales taxes, usually succeed in being taxed at the basic tax rate, usually 5 to 10 percent. By relying on tariffs for luxury taxation, the government ultimately forgoes the revenues it previously collected from duties on luxury goods, as well as severely undermining the very aims of luxury taxation.

If, instead, higher luxury rates on imports are imposed under a sales tax collected on both imports and any domestic production that may develop, unintended import substitution can be avoided. The use of import tariffs for luxury taxation—indeed, for any purpose other than providing protection to domestic industry—is one illustration of the general problem of using one economic policy instrument (tariffs) to achieve more than one purpose (protection, luxury taxation, and revenue). Reliance on import duties for revenue is subject to the same pitfalls just discussed. If it is desired to increase government revenue from imports, a 10 percent sales tax applied both to imports and any future domestic production yields at least as much revenue as a 10 percent import duty, without leading to accidental protection.

CORPORATE INCOME AND PROPERTY TAXES: THE INCIDENCE PROBLEM

Income taxes on domestic corporations and property taxes often are mentioned as possible methods for securing income redistribution through the budget. Corporate income ultimately is received, through dividends and capital gains, almost exclusively by the upper 5 to 10 percent of the income distribution. Also, ownership of

wealth, which in many lower-income countries largely takes the form of land, tends to be even more concentrated than income. But, to a greater extent in developing than in developed countries, efforts to secure significant fiscal redistribution through heavier taxes on domestic corporations and property are limited both by administrative and economic realities.

Administrative problems bedevil efforts to collect income taxes from domestic firms to at least as great an extent as for income taxes on individuals. Hence, in many countries where corporate taxes on local firms have been important, as much as two thirds to three quarters of nonoil corporate taxes flow from state-owned firms, not from private firms owned by high-income individuals. Taxes on land should be subject to less severe administrative problems because it is an asset that cannot be hidden easily. However, land valuation for tax purposes has proven difficult even in Canada and the United States. It is more difficult in developing countries. Few developing countries have been able to assess property at anything approaching its true value.

Economic realities hinder efforts to achieve greater progressivity in the tax system through heavier use of corporate and land taxes because of the tendency for taxation to unintentionally burden groups other than those directly taxed. This is the **incidence problem**. The incidence of a tax is its ultimate impact, not who actually pays the tax to the government but whose income finally is affected by the tax when all economic agents have adjusted in response to the tax. The point of incidence is not always the point of initial impact. Taxes on domestic corporations may reduce the income of capitalists, who in turn might shift their investment patterns to reduce taxation. The income of the workers they employ and the prices charged to consumers may be affected as well. Ultimately, all taxes are paid by people, not by things such as corporations and property parcels.

The implications of incidence issues may be illuminated by a simple application of incidence analysis to the corporation income tax. Consider a profit-maximizing company that has no significant monopoly power in the domestic market. If taxes on the company's income are increased in a given year, then after-tax returns to its shareholders in that year are reduced by the full amount of the tax. In the short term, the incidence of the tax clearly is on shareholders. Because shareholders everywhere are concentrated in higher-income groups, the tax is progressive in the short run. If capital were immobile, unable to leave the corporate sector, the long-term incidence of the tax also would rest on shareholders and the tax would be progressive in the long run as well.

But, in the long run, capital can move out of the corporate sector. To the extent that capital is mobile domestically but not internationally, the corporate tax also is progressive in the long run. Returns on capital remaining in the corporate sector are reduced by the tax. Untaxed capital owners employed outside the corporate sector also suffer a reduction in returns, because movement of capital from the taxed corporate sector drives down the rate of return in the untaxed sector. Because the corporate tax reduces returns to capital throughout the economy, all capital owners suffer,

including owners of housing assets, and in a closed economy, the long-run incidence again is progressive.

However, few if any developing economies are completely closed; indeed, capital in recent years has become much more mobile internationally. To the extent that capital can move across national borders and higher returns are available in other countries, domestic capital migrates to escape higher corporate taxes. But as capital leaves an economy, ultimately output is curtailed and the marginal productivity of workers falls. Prices of items produced with domestic capital rise. In this way, an increase in corporate taxes may be borne by domestic consumers, who pay higher prices for the reduced supply of corporate-sector goods. Domestic workers, whose income is reduced when production is curtailed, may also bear a part of the burden of the corporate tax. Hence the corporate tax may be regressive (worsen the income distribution) in the long run. Although under other plausible conditions, an increase in the corporation income tax may result in greater relative burdens on capitalists, this scenario is sufficient to illustrate that often the intentions of a redistributive tax policy may be thwarted by all the workings of the economy. Policy makers cannot be sure that all taxes imposed on wealthy capital owners ultimately are paid by them.

The foregoing discussion suggests that, whereas some tax instruments may achieve income redistribution in developing countries, the opportunities for doing so are limited in most countries, a conclusion supported by a large number of empirical studies. With few exceptions, these studies show that the failure to administer personal income taxes effectively—the failure to utilize the limited opportunities for heavier taxes on luxury consumption, overreliance on revenue-productive but regressive excise taxes, and inclusion of food in sales taxes—significantly reduces the redistribution impact of tax systems. Tax systems in developing countries tend to produce a burden roughly proportional across income groups, with some tendency for progressivity at the very top of the income scale. The very wealthy pay a somewhat greater proportion of their income in taxes than the poor, but the poor still pay substantial taxes: at least 10 percent of their income in many cases studied. Of course, in the absence of such efforts, the after-tax distribution of income may have been even more unequal. This suggests that, although difficult to implement and often disappointing in results, tax reforms intended to reduce income inequality are not futile exercises and they may prevent taxes from making the poor worse off.

The limits of tax policy suggest that, if the budget is to serve redistribution purposes, the primary emphasis must be on expenditure policy. Where redistribution through expenditures has been a high priority of governments, the results generally have been encouraging. The effects of government expenditure on income distribution are even more difficult to measure than those of taxes. But both the qualitative and the quantitative evidence available strongly indicate that, in developing countries, budget expenditures may transfer very substantial resources to lower-income households, in some cases as much as 50 percent of their income. Some of the measures using government expenditures to attack poverty were discussed in Chapter 6.

ECONOMIC EFFICIENCY AND THE BUDGET

SOURCES OF INEFFICIENCY

On the expenditure side of the budget, *social cost-benefit analysis* can be deployed to enhance efficiency (reduce waste) in government spending. On the tax side, promotion of economic efficiency is more problematic.

All taxes lead to inefficiencies to one degree or another. The objective, therefore, is to minimize tax-induced inefficiencies consistent with other goals of tax policy. In most developing societies, this objective largely reduces the necessity of identifying examples of waste engendered by taxes and purging them from the system. If a particular feature of a tax system involves large efficiency losses, called *excess burden* in public economics, and at the same time contributes little or nothing to such other policy goals as income redistribution, then that feature is an obvious candidate for abolition. A full discussion of those elements of tax systems that qualify for such treatment is properly the subject of an extended public finance monograph. We can do little more here than indicate some of the principal examples.

A major source of inefficiency in taxation is excessive costs of tax administration. In some countries and for some taxes, these costs have been so high that they call into question the desirability of using certain taxes for any purpose. This is true for certain kinds of narrow-based stamp taxes that have been used in Latin America to collect government revenue on the documentation of transfer of assets, rental agreements, checks, and ordinary business transactions. Many stamp taxes cost more to administer than they collect in revenue.

In some countries, even broad-based taxes have had inordinately high costs of collection. Sales taxes in Chile and Ecuador in the 1960s before the reform of their tax systems cost $1 in administration for every $4. Administrative costs to collect net taxes of $100 in most Organization for Economic Co-Operation and Development (OECD) countries in the early twenty-first century in contrast range from $0.42 to $2.08, with the United States at $0.47 at the low end and Japan at $1.50 at the higher end.[13] And, because taxes on capital gains are so difficult to administer everywhere, including North America, the cost of collecting this component of income taxes often exceeds the revenue in developing countries.

Many developing countries offer substantial tax incentives to encourage investment in particular activities and regions. Many of these, particularly income tax holidays for approved firms, have proven very difficult to administer and few have

[13]These are the average rates for 2000–02 and are taken from OECD, *Tax Administration in OECD Countries: Comparative Information Series* (Paris: Organization for Economic Cooperation and Development, 2004).

led to the desired result.[14] Given persistently pressing revenue requirements in most countries, granting liberal tax incentives may have no effect other than requiring higher rates of tax on taxpayers who do not qualify for incentives. It is a dictum of fiscal theory that economic waste (inefficiency) arising from taxation increases by the square of the tax rate employed, not proportionately. Therefore, it is not difficult to see that unsuccessful tax incentive programs involve inefficiencies for the economy as a whole that are not compensated for by any significant benefits. Largely for this reason, Indonesia abolished all forms of tax incentives in a sweeping tax reform in 1984.

Finally, some features of major tax sources involve needless waste. From the earlier discussion of the use of import duties for luxury tax purposes, it is clear that this is often a major source of inefficiency. The use of progressive tax rates in a corporation income tax is another example. Progressive rates of corporate tax, where they cannot be enforced, do little to contribute to income redistribution; and where they can be enforced, they lead to several kinds of waste. Two of the most important are fragmentation of business firms and inefficiency in business operations. The incentive for fragmentation is evident: Rather than be subject to high marginal rates of taxes, firms tend to split into smaller units and lose any cost advantages of size. Where high progressive rates are employed for company income, the tax takes a high proportion (say, 70 percent) of each additional dollar of earnings, so the incentive to control costs within the firm is reduced. For example, for a firm facing a marginal tax rate of 70 percent, an additional outlay of $1,000 for materials involves a net cost to the firm of only $300, because at the same time taxes are reduced by $700.

NEUTRALITY AND EFFICIENCY: LESSONS FROM EXPERIENCE

Experience around the world, both in developed and developing countries, seems to indicate that, in societies where efficiency in taxation matters, this objective is best pursued by reliance on taxes that are as neutral as possible. A **neutral tax** is one that does not lead to a material change in the structure of private incentives that would prevail in the absence of the tax. A neutral tax system, then, is one that relies, to the extent possible, on uniform rates: a tax on all income at a flat rate or a sales tax with the same rate applied to all goods and services. A neutral tax system cannot be an efficient tax system.

An **efficient tax system** involves a minimum amount of excess burden for raising a required amount of revenue, where the *excess burden* of a tax is the loss in total welfare, over and above the amount of tax revenues collected by the government. Figure 11–1 demonstrates how the excess burden of, say, a commodity tax is greater the more elastic the demand or the supply of the taxed item. Figure 11–1a depicts the inelastic case, good A, whereas panel b shows the elastic case, good B. Constant

[14]See, for example, Arnold C. Harberger, "Principles of Taxation Applied to Developing Countries: What Have We Learned?" in Michael Boskin and Charles E. McLure Jr., eds., *World Tax Reform: Case Studies of Developed and Developing Countries* (San Francisco: ICS Press, 1990).

marginal costs (MC) are assumed in both cases and at the same level for both goods to portray more starkly the contrasting results achieved in those cases. Before the tax is imposed on either good, the equilibrium price and quantity are P_a and Q_a for good A and P_b and Q_b for good B. Now, we impose a tax rate (t) on both goods. The new equilibrium (post-tax) magnitudes are P_{at} and Q_{at} for good A and P_{bt} and Q_{bt} for good B. For good A, the total amount of government revenue is the rectangle $P_aP_{at}cd$. The total loss in consumer surplus arising from the tax is the trapezoid $P_aP_{at}ce$. The excess of the loss in consumer surplus over the amount of government revenue is the conventional measure of efficiency loss from a tax, or excess burden. For good A, the excess burden is the small triangle *cde*. By similar reasoning, the excess burden in the case of good B is the larger triangle *fgh*. We see that taxes of equivalent rates involve more excess burden when imposed on goods with elastic demand.

We can see that efficient taxation requires neither uniformity nor neutrality but many different tax rates on different goods, with tax rates lower for goods with elastic demand and higher for goods with inelastic demand. This is known as the **Ramsey rule**, or the **inverse elasticity rule**. The problem is that a tax system under this rule would be decidedly regressive: The highest taxes would be required on foodstuffs, drinking water, and sumptuary items. Taxes would be lower on items such as clothing, services, and foreign travel, which tend to be both price and income elastic.

The principle of neutrality in taxation is not nearly as intellectually satisfying a guide to tax policy as efficient taxation. Nevertheless, neutral taxation is to be preferred as one of the underlying principles of taxation, along with equity, until such

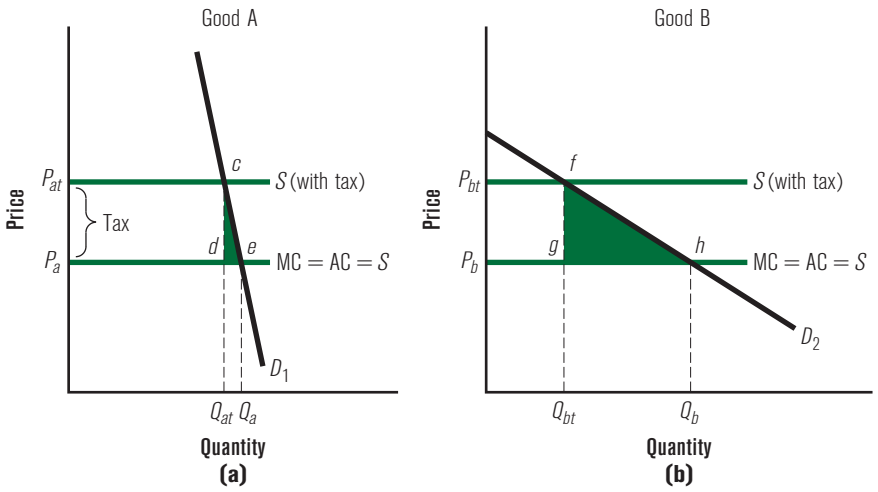


FIGURE 11–1 Taxation and Efficiency: Excess Burden of Commodity Taxes with Constant Marginal (MC) and Average Costs (AC), under Competition

(a) Case 1: Inelastic demand. (b) Case 2: Elastic demand. The shaded area in each case represents the excess burden of equal tax rates imposed on different goods. The greater the elasticity, the greater the excess burden.

time as analysts are able to identify optimal departures from neutrality—and uniformity in tax rates—in real-world settings. More intellectually satisfying tax policies must also wait until such time as administrative capacities are equal to the task of operating necessarily complicated structures of efficient or optimal taxes.

There is a paradox here. Neutral, uniform-rate taxes are less suited for efficiency goals than perfectly administered efficient taxes. Yet neutral tax systems are more likely to enhance efficiency in the economy than are efficient or optimal systems because neutral systems with uniform rates can be administered most easily and are much less vulnerable to evasion. This is not to say that neutrality has ever been or should be the overriding goal of tax policy. Governments often undertake very deliberate departures from neutral tax treatment of certain sectors or groups of society to achieve other policy goals. But, in real-world settings, these departures involve costs, not only in terms of tax administration but often in both equity and efficiency terms as well.

SUMMARY

- The expenditure side of the government budget is made up of capital or development items together with a wide variety of current use expenditures that are called *recurrent costs*. Developing country governments and donor aid programs frequently waste capital by neglecting the recurrent costs needed to maintain that capital.
- Developing countries rely mainly on indirect taxes (sales or value-added taxes and customs duties) rather than income and capital gains taxes because the latter are extremely difficult and expensive to administer in developing countries and thus produce less revenue.
- Tax reform in developing countries generally involves simplifying the tax system (one or a small number of different tax rates) and introducing taxes such as the value-added tax that have some self-enforcement properties. Greater complexity in the tax code leads to greater difficulty and greater corruption in administering the code. Simplified tax codes also reduce the distortions in the economy introduced by taxes.
- Taxation on the basis of an individual's ability to pay (higher-income people pay a larger share of their income in taxes) is a fundamental principle of equity. A good tax system is thus progressive rather than regressive, but progressive tax systems are extremely difficult to administer, particularly in developing countries, except in limited areas such as the exemption of food from sales taxes. When one looks at the actual incidence of taxes on high- and low-income earners, taxes that are designed to be progressive often turn out to be the opposite, and tax reform needs to take this into account.

12 Financial Development and Inflation

A country's **financial system** consists of a variety of interconnected financial institutions, both formal and informal. A central bank lies at the core of the organized financial system and is responsible for the control of the money supply and the general supervision of organized financial activity. Virtually everywhere, and particularly in developing countries, the commercial banking system is the most visible and vital component of the formal financial system, as acceptor of deposits and grantor of shorter-term credit. Other elements of the formal financial system include savings banks; insurance companies; and in a growing number of middle-income countries, pension funds and investment banks specializing in long-term credit as well as emerging stock exchanges. In high-income countries since the early 1990s there was the creation of a new set of institutions called *hedge funds* that played an important role in the global economic crisis of 2008–09.

Standing behind these financial institutions are a variety of government and sometimes private institutions whose role is to regulate the financial system. Typically separate agencies exist to regulate the commercial banks and the stock market, but the nature of regulation of the financial system is constantly changing. The need for regulation arises first and foremost from the fact that financial institutions exist primarily to manage the financial assets of people and companies other than the owners and managers of the financial institutions themselves. Regulation is designed to ensure that the financial assets of the public are managed responsibly without excessive risk. In the 1990s and during the first years of the twenty-first century the trend in high-income countries in North America and Europe was to reduce and even eliminate many of the regulations governing the financial system in the belief that the

market would be an adequate regulator on its own. Many financial companies also developed new financial instruments that were designed to provide a service, such as insurance, that would not be subject to the standard regulations governing insurance contracts (these new instruments were called *credit default swaps*). This trend toward further liberalization of financial markets came to an abrupt end in 2009 and 2010 as the global financial crisis made it clear that markets on their own could not adequately control risky behavior of financial institutions, and a wide variety of new regulations were put in place.

The financial crisis of 2008–09 started in the United States and spread quickly to Europe, but developing countries were soon enveloped in the consequences of the global recession. The primary difference between developing and high-income countries in this context was that the developing country financial systems were simpler and made much less use of the more exotic financial instruments than did high-income countries. Commercial banks dominated the financial system in developing countries, investment banks if they existed at all were smaller and weaker, and insurance companies were in their infancy. The existence of financial systems depends on trust that they will manage a depositor's or an investor's funds well and that trust is built on a history of responsible behavior, on an effective regulatory system, and on the transparency of the operations of the financial organizations themselves. It takes decades or longer to develop trustworthy institutions and trust can be lost quickly as happened in 2008–09, but in developing countries this trust and the related effective regulation was largely confined to the commercial banks.

The limited scope of the formal financial sector in developing countries, however, left many people and small companies without ready access to credit or even to a reliable place to put their savings. In recent decades this has led to the creation of new **micro-credit institutions** designed to reach these left-out individuals, but long before formal micro-credit institutions were created, there was in virtually all developing countries an informal financial sector existing alongside the formal one. Informal financial institutions include pawnshops, local moneylenders, trade credit, and other arrangements involving the borrowing and lending of money such as intrafamily transfers and cooperative credit. In very low income countries, or even in some middle-income countries, the informal financial sector may rival the formal system in size. In almost all cases the informal sector is unregulated, although governments do interfere in its operation from time to time.

Financial policy embraces all measures intended to affect the growth, utilization, efficiency, and diversification of the financial system. In North America and western Europe, the term *financial policy* ordinarily is used as a synonym for *monetary policy*, the use of monetary instruments to reduce price instability caused by fluctuations in either internal or external markets. In the United States, these instruments include open-market operations, changes in legal reserve requirements of commercial banks, and shifts in central bank (Federal Reserve System) lending (rediscount) rates to commercial banks (these terms are explained later in this chapter). In developing

countries, the term *financial policy* typically has a much broader meaning. Monetary policy is part of financial policy but so are measures intended to encourage the growth of savings in the form of financial assets, develop money and capital markets, and allocate credit among different economic sectors. Regulation to control risk, as in the 1997–98 and the 2008–09 financial crises, is also a component of financial policy.

THE FUNCTIONS OF A FINANCIAL SYSTEM

The financial system provides four basic services essential for the smooth functioning of an economy. First, it provides a medium of exchange and a store of value, called *money*, which also serves as a unit of account to measure the value of the transactions. Second, it provides channels for mobilizing savings from numerous sources and channeling them to investors, a process called *financial intermediation*. Third, it provides a means of transferring and distributing risk across the economy. Fourth, it provides a set of policy instruments for the stabilization of economic activity. When these functions are managed poorly the result on a macro level can be a financial crisis leading to a global recession, inflation, or a debt crisis. The management of inflationary pressures will be discussed in this chapter and debt crises are the subject of Chapter 13.

MONEY AND THE MONEY SUPPLY

An economy without money as a **medium of exchange** is primitive. Trade between individuals must take the form of high-cost, inefficient barter transactions. In a barter economy, goods have prices, but they are expressed in the relative prices of physical commodities: so many kilos of rice for so many liters of kerosene, so many meters of rope for so many pairs of sandals, and so on. Trading under such circumstances involves onerous information costs.

Few societies have ever relied heavily on barter because of the high costs implicit in this means of exchange. At some point, prices of goods and services begin to be expressed in terms of one or more universally accepted and durable commodities, like gold and silver or even beads and cowrie shells. The rise of commodity money diminishes the transaction and storage costs of trade but still involves problems of making exchanges across space and time. Gold and silver prices fluctuate, and the commodities, therefore, are not fully reliable as **units of account**. As specialization within an economy increases, financial instruments backed by commodities appear. In the last century, with the rise of central banking all over the world, currency evolved into *fiat* money, debt issued by central banks that is legal tender. It is backed not by commodities of equivalent value but only by the full faith and credit of the central bank.

As markets widen and specialization proceeds apace, a need arises for still another financial instrument, **transferable deposits**. In the normal course of development, *checking* or *demand deposits* (deposits that may be transferred to any economic agent at the demand of the depositor) appear first and ordinarily bear little or no interest. Rising levels of economic activity, however, create increasing needs for transaction balances; individuals will always maintain some balances in demand deposits to meet these needs but tend to economize on the levels of such deposits if no interest is paid on them. With further monetization still another financial instrument begins to grow in importance, *time deposits*, which also are legally transferable on demand but sometimes only after stated periods. Time deposits involve contractual interest payments; higher interest rates induce people to hold greater amounts of deposits in this form.

While checking (demand) and time deposits are **liabilities** (or debts) of commercial banks, they are **financial assets** for the individuals who hold them. Both demand and time deposits are known as *liquid financial assets*. Unlike *nonfinancial assets* that also can be held by households and businesses (inventories, gold, and land), demand and time deposits can be quickly and conveniently converted into their currency equivalents. *Currency*, by definition, is the most liquid of all assets. The concept of liquid financial assets is an important one in any discussion of financial policy in developing countries. For most developing countries, the movement of savers in and out of liquid financial assets may be the prime factor behind the success or failure of a financial policy. We will see that long-term shifts from tangible, or nonfinancial, physical assets to financial assets, particularly liquid assets, bodes well not only for economic growth but also for economic stability.

A country's **money supply** may be defined as the sum of all liquid assets in the financial system. While not all economists agree about what constitutes a liquid financial asset, most vastly prefer this money supply concept to those commonly employed in early postwar monetary analysis. Formerly, the money supply was conventionally defined as the sum of only two liquid financial assets: currency in circulation outside banks (C) plus demand deposits (D), which together are known as M1 (*narrow money*). However, it later became clear that, because depositors tend to view time and savings deposits (T) as almost as liquid as demand deposits, the former also should be included in any workable concept of money supply, called M2 (*broad money*). Finally, for high-income countries, specialized deposit-taking financial institutions have arisen and offer an array of options to savers other than those available in commercial and savings banks. The liabilities of these specialized institutions (O) are included in M3 (*total liquid liabilities*), the broadest measure of money. Thus

$$\text{M1} = C + D \qquad [12\text{—}1]$$

$$\text{M2} = \text{M1} + T \qquad [12\text{—}2]$$

$$\text{M3} = \text{M2} + O \qquad [12\text{—}3]$$

For most low-income countries and many middle-income countries, liquid financial assets constitute by far the greatest share of outstanding financial assets. But, as income growth continues and the financial system matures, less-liquid financial assets assume progressively greater importance. These include primary securities such as stocks, bonds (issued both by government and firms), and other financial claims on tangible (physical) assets that are convertible into currency equivalents with only some risk of loss to the asset holder and are hence less liquid than demand or time deposits.

The evolution of financial activity follows no set pattern across countries. Differing economic conditions and policies may result in widely divergent patterns of financial growth. Nevertheless, as per capita income rises, money increases as a ratio to gross domestic product (GDP). Table 12–1 shows these patterns for broad money (M2).

TABLE 12–1 Broad Money (M2) as a Percentage of Gross Domestic Product (GDP), 1980 and 2008

	PERCENT OF GDP	
COUNTRY/ECONOMY	1980	2008
Low-income countries	19	36
Ethiopia	–	31
Tanzania	–	28
India	32	70
Bangladesh	13	55
Kenya	30	40
Nigeria	24	30
Ghana	16	24
Honduras	23	51
Lower-middle-income countries	28	54
Pakistan	39	44
Bolivia	18	55
Cameroon	21	19
Philippines	22	59
Sri Lanka	28	33
Indonesia	13	36
Peru	17	31
Egypt	52	84
Upper-middle-income countries	29	79
Colombia	24	33
Argentina	29	25
Mexico	27	26
Malaysia	71	118
Korea, Republic	29	62
High-income countries	72	113
Saudi Arabia	13	48
Japan	138	207
United States	68	83

Source: World Bank, *World Development Indicators 2011* (Washington, DC: World Bank, 2011), available at http://databank.worldbank.org/ddp/home.do.

For all income classes shown in the table, the ratio of M2 to the GDP rose substantially from 1980 through 2008. This reflects both economic growth and changing policies. Looking at country income aggregates, the ratio rises as one moves from low- to middle- to high-income economies. But the variance among countries is notable. In 2008, Kenya, a low-income economy, had a higher monetization ratio (40 percent) than either Sri Lanka (33 percent) or Mexico (26 percent) both middle-income nations. Clearly, income level is not the only determinant of financial growth.

FINANCIAL INTERMEDIATION

As financial structures become increasingly rich and diversified in terms of financial assets, institutions, and markets, the function of money as a medium of exchange, store of value, and unit of account tends to be taken for granted, except in situations of runaway inflation. As financial development proceeds, the ability of the financial system to perform its second major function, financial intermediation, grows as well. The process of financial intermediation involves gathering savings from multitudinous savers and channeling them to a much smaller but still sizable number of investors. With a few exceptions, households are the main net savers in developing countries. At early stages of economic development, a preponderant share of intermediation activities tends to be concentrated in commercial banks. As development proceeds, new forms of financial intermediaries begin to appear and gradually assume a growing share of the intermediation function. These include investment banks, insurance companies, pension funds, and securities markets.

Financial intermediation is best seen as one of several alternative technologies for mobilizing and allocating savings. The fiscal system discussed in Chapter 11 furnishes another alternative, and we see in Chapter 14 that reliance on foreign savings constitutes still another. Further, we will observe in this chapter that inflation has also been employed as a means of mobilizing resources for the public sector. Indeed, a decision to rely more heavily on financial intermediation as a means of investment finance is tantamount to a decision to rely less heavily on the government budget, foreign aid and foreign investment, and inflation to achieve the same purpose.

TRANSFORMATION AND DISTRIBUTION OF RISK

Another major service provided by a well-functioning financial system is the transformation and distribution of risk. All economic activities involve risk taking, but some undertakings involve more risk than others. Individual savers and investors tend to be risk averse; the marginal loss of a dollar appears more important to them than the marginal gain of a dollar. But the degree of risk aversion differs among individuals. When risk cannot be diversified (or pooled) across a large number of individuals, savers and investors demand greater returns (or premiums) for bearing risk and activities involving high risk tend not to be undertaken. But high-risk activities

may well offer the greatest returns to the economy as a whole. A well-functioning financial system furnishes a means for diversifying, or pooling, risk among a large number of savers and investors. The system may offer assets with differing degrees of risk. Financial institutions that specialize in assessing and managing risks can assign them to individuals having different attitudes toward and perceptions of risk. The term *well-functioning financial system* needs to be emphasized here. Early in the twenty-first century, highly sophisticated financial systems in high-income countries, notably the United States, managed risk poorly and this led to the financial crisis and global recession of 2008–09.

STABILIZATION

Finally, the financial system provides instruments for the stabilization of economic activity in addition to those available under fiscal policy and direct controls. All economies experience cyclical changes in production, employment, and prices. Governments often attempt to compensate for these fluctuations through policies affecting the money supply. Because unemployment in developing countries rarely is of the type that can be cured by monetary expansion, the use of financial policy for stabilization purposes generally focuses on efforts to control inflation. As the economic crises that began in Asia in 1997 and in the United States in 2008 demonstrated, however, the financial system itself can become a major source of instability in the country. From the beginning of the twenty-first century, as a result, financial policy also has focused on more than simply whether the supply of money is growing too rapidly or too slowly. Policy makers have had to devote much attention to ensuring the integrity and stability of banks and other nonbank financial institutions to avoid a severe recession brought on by financial panic. In the discussion that follows, we focus first on the use of the financial system to control inflation and then return to the issue of how the financial system should be designed and managed to avoid panic and recession.

INFLATION

From the 1980s to the present **price inflation**, defined as a sustained increase in the overall price level, generally was regarded as a malady that in its milder forms was annoying but tolerable and in its moderate form corrosive but not fatal. Runaway inflation, also known as **hyperinflation**, however, always has been recognized as severely destructive of economic processes, with few offsetting benefits. Whereas a number of influential thinkers in the 1950s and 1960s advocated some degree of moderate inflation (for example, inflation rates between 8 and 12 percent) as a tool for promoting growth, few adherents of this view remain. Many others did not

actively advocate inflation but tended to have a higher threshold of tolerance for a steadily rising general price level than now is common.

INFLATION EPISODES

Inflationary experiences vary widely among developing countries and generalizations are difficult to make. Nevertheless, postwar economic history offers some interesting national and regional contrasts in both susceptibility to and tolerance for different levels of inflation. The period before the early 1970s was one of relative price stability in developing countries. In the southern cone of Latin America, however, particularly in Argentina, Brazil, and Chile, **chronic inflation** (prices rising 20 to 50 percent per year for three years or more) was an enduring fact of economic life for much of the 1950s through the 1980s and even beyond. The experience of these countries indicates that a long period of double-digit inflation does not necessarily lead to national economic calamity in all societies. In all, about two dozen countries experienced chronic inflation from 1950 to 2008, as shown in Table 12–2, but only a few, including Tajikistan, Venezuela, and Uzbekistan, have experienced chronic inflation throughout much of the first decade of the twenty-first century.

However, a tolerable rate of inflation in one country may constitute economic trauma in another. This may be seen more readily by considering the often progressive inflationary disease, acute inflation. **Acute inflation**, defined here as inflation between 50 and 200 percent for three or more consecutive years, was experienced by 18 countries listed in Table 12–2 over the postwar period, in some cases more than once per country. For Brazil, the progression from chronic to acute inflation did not result in any noticeable slowing of that country's relatively robust economic growth, whatever it may have meant for income distribution. In Ghana, on the other hand, a decade of acute inflation coincided with a decade of decline in GDP per capita. Although it may be tempting to attribute economic retrogression in Ghana at the time to acute inflation, it is more likely that the same policies that led to sustained inflation, not the inflation itself, were responsible for declines in living standards there.[1] Since the middle of the 1990s acute inflation has been rare.

Although acute inflation has proven toxic to economic development in some settings and only bothersome in others, **runaway (hyper) inflation** almost always has had a devastating effect. Inflation rates in excess of 200 percent per year represent an inflationary process that is clearly out of control; 22 countries have undergone this traumatic experience since 1950. One major bout of runaway inflationary experience occurred in three Latin American countries: Bolivia and Argentina in 1985, Argentina again in 1988–90, and Peru in 1988–91. In Bolivia, the annual rate of inflation over

[1]For a diagnosis of the causes of the Ghanian economic decline after 1962, see Michael Roemer, "Ghana, 1950 to 1980: Missed Opportunities," and Yaw Ansu, "Comments" (201–30), both in Arnold C. Haberger, ed., *World Economic Growth* (San Francisco: ICS Press, 1984).

TABLE 12-2 **Inflation Outliers: Episodes of Chronic, Acute, and Runaway Inflation among Developing Countries, 1948–2008 (average annual rates)**

	CHRONIC INFLATION (20–50 PERCENT, 3 YEARS)		ACUTE INFLATION (50–200 PERCENT, 3 YEARS)		RUNAWAY INFLATION (200+ PERCENT, 1 YEAR)	
COUNTRY	YEAR(S)	RATE	YEAR(S)	RATE	YEAR(S)	RATE
Angola			1997–2003	143	1993–96	858
Argentina	1950–74	27	1977–82	147	1976	443
			1986–87	111	1983–85	529
					1988–91	1,400
Armenia					1990–97	700
Azerbaijan					1990–96*	590
Belarus			1997–2002	87	1990–96*	715
Bolivia	1979–81	33	1952–59	117	1983–86	1,132
Brazil	1957–78	36	1979–84	108	1985	227
					1986	145
					1987–93	831
Chile	1952–71	29			1973–76	308
	1978–80	36				
	1983–85	26				
Colombia	1979–82	26				
	1988–92	28				
Dominican Republic			1988–92	51		
Ecuador	1997–2001	49				
Georgia					1990–96*	2,279
Ghana	1986–90	32	1976–83	73		
			1995–97	75		
Indonesia					1965–68	306
Kazakhstan					1990–96*	605
Kyrgyz Republic					1990–96*	256
Latvia					1990–96*	111
Lithuania					1990–96*	179
Malawi	1998–2002	40				
Mexico			1982–88	70		
Myanmar	2001–03	29	1992	116		
Nicaragua	1979–84	33			1985–91	2,130
Paraguay			1951–53	81		
Peru	1975–77	32	1950–55	102	1988–91	1,694
			1978–87	85		
Romania			1991–93	132		
			1997–92	56		
Russia	1996–99	42			1993–95	376
Sierra Leone			1983–92	81		
South Korea			1950–55	95		
Tajikistan	2000–08	22			1990–96*	394
Turkmenistan					1990–96*	1,074
Tanzania	1980–89	27				
Turkey	1981–87	38	1978–80	69		
			1988–2002	70		

(*Continued*)

TABLE 12–2 *Continued*

Uganda	1990–92	37	1981–89	101		
Ukraine					1993–95	1,167
Uruguay	1948–65	26	1965–68	83		
	1981–83	34	1972–80	68		
Uzbekistan	2000-08	29	1984–92	76		
Venezuela	1987–92	40				
	1993–98	43				
	2002–08	28				
Zaire/Congo	1981–82	36	1976–80	68	1991–92	2,987
			1983–90	68	1992–95	3,206
					1996–2002	218
Zambia	1985–87	44	1988–92	113		
			1993–96	71		
Zimbabwe					2002–07	5,305†

*For a number of the former Soviet republics, the gross domestic product (GDP) price deflator data are given for the entire period 1990–96, not broken down by year.
†Data for Zimbabwe are not available after 2007. Some estimates place Zimbabwe's annual inflation at over 30,000 in the late 2000s.

Sources: International Monetary Fund, *International Financial Statistics Yearbook 1994* (Washington, DC: International Monetary Fund, 2004) and International Monetary Fund, *International Financial Statistics Yearbook*, 2003 (Washington, DC: International Monetary Fund, 2003); and World Bank, *World Development Indicators 2010*, at http://databank.worldbank.org/ddp/home.do accessed June 2011.

a period of several months in 1985 accelerated to a rate of nearly 4,000 percent. In Argentina, the monthly rate of price increases was 30.5 percent in June 1985 alone; on an annual basis that would have been an inflation rate of 2,340 percent. In both Bolivia and Argentina for much of 1985, workers had little choice but to spend their paychecks within days of receipt, for fear that prices would double or triple over the next week. In Peru, hyperinflation in 1989 gave birth to publications devoted only to the tracking of inflation (see Box 12–1).

In the 1990s, hyperinflation was experienced by virtually all the new republics formed after the breakup of the Soviet Union. In the first part of the 1990s, inflation reached a rate of nearly 400 percent a year in Russia over a period of several years and over 1,000 percent a year in the Republic of Georgia and Turkmenistan. In the latter two countries, civil war put heavy demands on government expenditures while reducing the ability of those governments to collect taxes, leading them to finance their activities by printing money. In Africa in the 1990s, civil war in Angola and the Congo also led to inflation rates of over 1,000 percent per year. Since the year 2000, however only the Congo (briefly) and Zimbabwe (throughout the first decade of the twenty-first century) have experienced runaway inflation. In the Congo the ongoing war in the eastern part of the country is the main cause. In Zimbabwe runaway inflation has been the result of across the board mismanagement of the economy throughout that decade.

BOX 12-1 HYPERINFLATION IN PERU, 1988–90

The economic and social havoc wrought by hyperinflation is difficult to comprehend for those who have not lived through the experience. The Peruvian hyperinflation, which began in 1988 and continued until 1991, provides some rueful examples.

The inflationary process was triggered by large budgetary deficits and sustained by subsequent ongoing deficits, virtual economic collapse, and steadily rising inflationary expectations. Peru already had experienced two serious bouts with acute inflation since 1950, but the pace of inflation in 1989 was the highest in the nation's history: 28 percent per month, or about 2,000 percent per year. Moreover, inflation accelerated in the first six weeks of 1990, as prices rose by 6 percent per week, or about 1 percent per day. From January 1989 to December 1990, the value of the Peruvian currency (the inti) on the free market fell from 1,200 intis per dollar to 436,000 intis per dollar.

This hyperinflation may turn out to be one of the best documented in history: In 1988, Richard Webb, an internationally respected Peruvian economist, began to publish a magazine devoted essentially to helping producers and consumers cope with the chaos associated with runaway inflation. The magazine, called *Cuanto?* (*How Much?*) appeared monthly. It not only provided details on price developments for a large number of commodities and services but managed to extract what little humor there is in a situation in which the price of a movie ticket rises while people are waiting in line to buy it or taxi drivers must carry fare money in a burlap bag because the domestic currency collected in fares each evening is too bulky to fit in a wallet.

But precious little is funny about hyperinflation. In Peru, it was a story of government employees going without pay for weeks at a time, of indices of poverty nearly doubling from 1987 to 1989, further impoverishing the poorest 40 percent of the population. It was a story of precipitous decline in gross domestic product (GDP) and the rise of pervasive black markets in everything from dollars to gasoline to cement. It was a time when, on each payday, laborers rushed to the market to buy their weekly food supplies before they were marked up overnight. It was a tale of wide variations in price rises, in which prices for such items as pencils and chicken increased by more than 25 times from February 1989 to February 1990, but prices of light bulbs and telephone services increased "only" 10-fold.

Imagine life in Lima, the capital city, in the first few weeks of 1990, for a middle-income family trying to survive. For the first 40 days of the year, increases in the price of dying outpaced the price of living: The cost of funerals rose 79 percent, while house rent rose by 56 percent and the price of restaurant meals and haircuts increased by 44 percent. By mid-1990, the economic paralysis of Peru was virtually complete. Peru's hyperinflation ended in 1992 (although inflation remained acute at 75 percent) as government reduced its budget deficit to 1 percent of GDP.

The countries listed in Table 12–2 brought on inflation in three different ways. In one group (including Argentina, Chile, Ghana, Indonesia, Peru, Russia, and Ukraine), large budget deficits relative to GDP were financed by borrowing from the central bank. In a second group (Paraguay in the early 1950s and Brazil and Uruguay before 1974), inflation was caused by a massive expansion of credit to the private sector. And in Nicaragua, Sierra Leone, Uganda, Angola, Congo, Zimbabwe, and several of the nations created out of the former Soviet Union, political strife or civil war exacerbated the fiscal and monetary causes of inflation. Whatever the initial impetus to inflation, once it begins to accelerate, the public begins to expect inflation to continue, and this leads to even-higher, more-sustained price increases.

For developing countries as a group, Table 12–3 shows that, on average, inflation was only 13 percent a year until the oil crisis began in 1973 but jumped to 21 percent a year during the period of rising oil prices until 1981. Inflation then accelerated to more than 30 percent a year even as oil prices began falling and exceeded 50 percent a year after oil prices collapsed in 1986. In the 1990s, these high rates of inflation continued through the first half of the decade and then slowed markedly after 1995. The principal exceptions were the states of the former Soviet Union, where triple-digit inflation in many cases continued to the end of the century and in a few cases a few years more. The industrial countries, in contrast, had much lower inflation throughout and the highest price increases coincided with the rise in oil prices.

MONETARY POLICY AND PRICE STABILITY

We saw from Table 12–3 that inflation largely was conquered in Asia during the 1970s but accelerated in Latin America and Africa during and especially after the oil-price boom of the 1970s. In the 1990s, however, attempts to control inflation were more serious and widespread with the notable exception of the republics formed out of the collapse of the Soviet Union. Even in these former Soviet republics, inflation by the beginning of the twenty-first century was generally down to single digits. The only case of runaway inflation during the first decade of the twenty-first century, as already noted, was Zimbabwe, where inflation was the result of a general collapse of the economy and of a government fiscal policy in the context of a ruling elite desperately trying to hold onto power. Controlling inflation in the midst of a civil war and the collapse of government functions requires first ending the war and reestablishing a functioning government. To control inflation under more normal circumstances, monetary policy is the principal instrument used to achieve price stability.

TABLE 12–3 **Inflation by Regional Groupings, 1963–2008 (percent per annum)**

REGION	1963–73	1973–81	1981–92	1992–95	1995–98	1998–01	2001–08
World	26.3	13.8	17.9	19.2	6.8	4.2	4.5
Industrial countries	24.6	10.3	13.0	2.5	1.9	1.8	2.1
Developing countries	12.9	20.7	44.1	47.9	12.6	7.1	8.5
Africa	24.9	17.3	27.0	40.2	12.1	8.2	9.1
Asia	13.5	28.8	16.8	11.6	6.9	2.4	3.9
Middle East*	24.2	16.6	16.0	15.0	9.5	5.5	12.6*
Latin America and Caribbean	18.4	43.7	162.8	216.5	15.5	8.4	9.6

*The Middle East gross domestic product deflator series ends in 2007.

Sources: International Monetary Fund, *International Financial Statistics Yearbook 1994* (Washington, DC: International Monetary Fund, 1994); International Monetary Fund, *International Financial Statistics, 1998* (Washington, DC: International Monetary Fund, 1998); International Monetary Fund, *International Financial Statistics, 2003* (Washington, DC: International Monetary Fund, 2003); and International Monetary Fund, *International Financial Statistics 2010* (Washington, DC: International Monetary Fund, 2010).

MONETARY POLICY AND EXCHANGE-RATE REGIMES

Appropriate use of monetary policy in controlling inflation depends critically on the type of exchange-rate regime used by a country. Exchange-rate regimes form a continuum with fixed (pegged) exchange rates at one end and floating (flexible) exchange rates at the other. Under a **fixed-exchange-rate** system, a country attempts to maintain the value of its currency in a fixed relation to another currency, say, the U.S. dollar: The value of the local currency is **pegged** to the dollar. This is done through intervention by the country's monetary authorities in the market for foreign exchange and requires the maintenance of substantial **international reserves** (reserves of foreign currencies), usually equivalent to the value of four or more months' worth of imports.

Consider the case of Thailand, where from 1987 to early 1997 the Thai currency, the baht, was fixed at an exchange rate close to 25 baht to US$1. Because the exchange rate, if left to its own devices, would change from day to day to reflect changes in both the demand for and supply of exports and imports and capital flows, to defend the peg, the government must be prepared to use the country's international reserves to buy or sell dollars at an exchange rate of 25 to 1 to keep the exchange rate from moving. If, for example, a poor domestic rice harvest caused the nation to increase its food imports, the baht–dollar exchange rate would tend to rise (the baht would depreciate, as its dollar value falls) in the absence of any net sales of dollars from Thailand's international reserves. To sustain a fixed exchange rate, of course, the country must have sufficient foreign exchange reserves to keep on buying baht at that fixed rate. In 1997, because of large capital outflows, Thailand actually ran out of foreign exchange and had to abandon its support of the pegged rate.

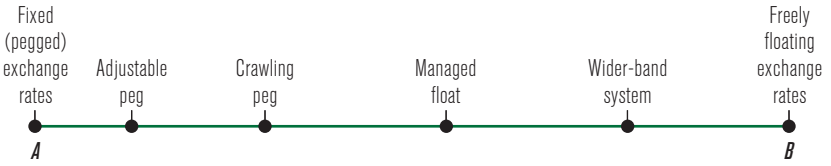


FIGURE 12–1 Continuum of Prototypes of Exchange-Rate Regimes

As one moves from point *A* to point *B*, both the frequency of intervention by domestic monetary authorities and the required level of international reserves tend to be lower. Under a pure fixed-exchange-rate regime (point *A*) authorities intervene so that the value of the currency vis-à-vis another, say, the U.S. dollar, is maintained at a constant rate.

Under freely **floating rates**, the authorities simply allow the value of local currency vis-à-vis foreign ones to be determined by market forces. Between the two ends of this continuum lie a number of intermediate options (Figure 12–1).[2] Closest to the floating-exchange-rate option is the *wider band* system, by which the exchange rate of a country is allowed to float or fluctuate within a predefined band of values, say, between 23 and 27 baht to US$1. But when conditions threaten to push the value of the currency beyond the band, the authorities intervene by buying or selling local currency as appropriate to stay within the band. Further along the continuum away from floating rates is the *managed float*, where the authorities are committed to defend no particular exchange rate, but they nevertheless intervene continuously at their discretion. A country with steadily shrinking international reserves, for example, might allow the value of its currency to depreciate against the value of other currencies—that is, allow the exchange rate to rise against other currencies.

Two other systems are closely related hybrids of fixed and floating rules. The *crawling peg*, used over a long period by Brazil, Colombia, and Indonesia, involves pegging the local currency against some other currency but changing this in gradual, periodic steps to adjust for any differential between the country's inflation rate and the world inflation rate. Closest to a fixed-exchange-rate system is the *adjustable peg*, involving a commitment by the monetary authorities to defend the local exchange rate at a fixed parity (peg), while reserving the right to change that rate when circumstances require.

Two very rigid forms of pegged exchange rates that a small number of countries have adopted are currency boards and dollarization. With a *currency board*, the government issues domestic currency only when it is fully backed by available foreign exchange reserves at the given exchange rate. Currency in circulation increases when additional foreign exchange becomes available (say, through increased export receipts) and decreases when foreign exchange becomes scarcer (say, through an

[2]For a full discussion of these and other types of exchange-rate regimes, see John Williamson, *The Open Economy and the World Economy* (New York: Basic Books, 1983), pp. 238–41; or Anne O. Krueger, *Exchange Rate Determination* (New York: Cambridge University Press, 1983), pp. 123–36.

increase in imports or capital outflows). This system ensures that the country will not run out of foreign exchange. However, the main instrument of adjustment becomes domestic interest rates, which increase when foreign exchange (and domestic currency) becomes scarcer, and decline when foreign exchange becomes more available. Hong Kong, Bulgaria, Argentina, Brunei, Djibouti, Estonia, and Lithuania all have or have had currency boards.

With *dollarization*, one country adopts another country's currency, as Panama did many years ago when it adopted the U.S. dollar as its currency. Most economists believe that currency boards and dollarization are appropriate in only a very limited number of developing countries that are small, very open to trade, and not vulnerable to large commodity price swings. In Hong Kong, the currency board system works reasonably well, although high interest rates and hence slower growth were required to prevent large currency outflows during and immediately after the Asian financial crisis. In Argentina at the beginning of the twenty-first century, the inability of the central government to rein in excessive local government spending caused pressures on reserves that ultimately made it impossible to maintain the peg to the dollar that was essential to the continuance of its currency board.

The currencies of all the major industrial countries have floated vis-à-vis one another since the early 1970s, with occasional intervention by national monetary authorities to prevent very sharp swings in rates. Most developing countries have adhered to either the adjustable-peg or the crawling-peg system, although an increasing number, particularly in Africa, have been adopting floating-rate systems as part of stabilization programs. Because, in practice, both pegged systems operate for particular periods like fixed-rate regimes, for our analysis, we focus most of our attention on monetary policy issues arising under fixed exchange rates in small, open economies.

SOURCES OF INFLATION

In open developing economies with fixed exchange rates, the rate of monetary expansion no longer is under the complete control of domestic monetary authorities. Rather, countries with fixed exchange rates may be viewed as sharing essentially the same money supply, because the money of each can be converted into that of the others at a fixed parity.[3] Under such circumstances, the stock of money (M), by definition, is the sum of two components: the amount of domestic credit of the banking system that is outstanding (DC) and the stock of international reserves of that country (IR), measured in terms of domestic currency. The money supply, therefore, has a domestic component and an international component, so we have

$$M = DC + IR \qquad [12\text{–}4]$$

[3]This section draws substantially on syntheses of monetary and international economics by Arnold C. Harberger. See his "A Primer on Inflation" and "The Inflation Syndrome," papers presented in the Political Economy Lecture Series, Harvard University, March 19, 1981.

Changes in the domestic money stock can occur either through expansion of domestic credit or by monetary movements that lead to changes in international reserves. That is,

$$\Delta M = \Delta DC + \Delta IR \qquad [12\text{–}5]$$

Under fixed exchange rates, a central bank of any small country can control *DC*, the domestic component, but has only very limited control over *IR*, the international component. Under such circumstances, developing countries that attempt to keep the rate of domestic inflation below the world inflation rate (through restrictive policies on domestic credit) are unable to realize this goal. If, fueled by monetary expansion abroad (growth in the world money supply), world inflation initially is running in excess of domestic inflation, the prices of internationally traded goods rise relative to those of domestic, nontraded goods.[4] Imports fall, exports rise, and the balance of payments moves toward surplus and causes a rise in international reserves. Therefore, the foreign components of the money stock rise. This is tantamount to an "importation of money" and eventually undoes the effort to prevent importation of world inflation. Again, a small country on fixed exchange rates can do little to maintain its inflation rate below that of the rest of the world. For very open countries with few restrictions on the movement of goods and capital into and out of the country, the adjustment to world inflation can be very rapid (less than a year). For less-open countries with substantial restrictions on international trade and payments, the process takes longer, but the outcome is inevitable under fixed exchange rates.

The fact that financial policy for stabilization in countries with fixed exchange rates is heavily constrained by international developments is sometimes taken to mean that changes in the domestic component of the money stock have no impact on prices in economies adhering to fixed exchange rates. On the contrary, excessive expansion in money and credit surely will result in domestically generated inflation that, depending on the rate of expansion, for a time, can be well in excess of world inflation rates. However, such a situation cannot continue for long, as excess money creation spills over into the balance of payments via increased imports and leads to a drain on international reserves and, ultimately, an inability to maintain the fixed exchange rate. As reserves dwindle, the country no longer can defend its exchange rate and devaluation becomes inevitable.[5] Inflation, therefore, can be transmitted to small, open economies through the working of the world economy or generated by domestic developments.

A growing number of developing countries have floating exchange rates (point *B* on the continuum in Figure 12–1). A floating exchange-rate regime allows countries

[4]This is but one of several mechanisms that led to changes in international reserves sufficient to thwart efforts by developing countries to insulate themselves from world inflation.

[5]Import controls are frequently used to stem the drain of reserves and avoid devaluation for a time. But import controls engender another set of distortions and inefficiencies, explored in Chapter 18, that eventually require more drastic measures, including devaluation.

to insulate themselves from world inflation. Under such a system, the rise in world prices attendant on world inflation would initially favor exports from the country and discourage imports. As a consequence, the current account of the country's balance of payments improves, international reserves rise, and the exchange rate soon appreciates (fewer baht are required to buy dollars, for instance). The appreciation in the country's exchange rate cancels out external price increases and prevents the importation of world inflation.

Under any exchange-rate regime, domestically generated inflation may result from excessive increases in domestic credit from the banking system to either the public or the private sector. Budget deficits of the central government, for example, must be financed by borrowing, but the embryonic nature of money and capital markets in most developing countries generally means that governments facing deficits ordinarily must resort to borrowing from the central bank. Borrowing from the central bank is equivalent to direct money creation via the printing press. The result is a direct addition to the reserve base of the monetary system, an increase in so-called high-powered money. It is important, however, to recognize that not all budgetary deficits are inflationary. We have seen that a growing economy is characterized by a growing demand for liquid assets, including money. Moderate budgetary deficits year after year, financed by the central bank, can help satisfy this requirement without leading to inflation. In general, the money stock may expand at least as fast as the growth in real income, with little or no inflationary consequences.

Earlier we saw that liquid assets normally are between 40 and 60 percent of GDP in low- and lower-middle income economies (with wide variations), equivalent to roughly three to six months of income. Therefore, the public generally is willing to hold this much in money balances. A deficit of 2 percent of GDP financed by money creation adds only marginally to the money supply and easily may be accepted by the public. But a deficit of 8 percent of GDP increases the stock of money by an amount equal to one more month of income, an amount the public may be unwilling to hold (unless nominal interest rates on deposits are greatly increased). The excess spills over into higher prices.

Use of bank credit to finance government deficits has not been the only source of inflationary monetary expansion in developing countries. Sometimes excessive growth of credit to the private sector has played a more significant role in domestically generated inflationary processes. Nevertheless, as a general rule, inflation rates that are much in excess of world inflation usually have been traceable to budgetary deficits.

For countries attempting to maintain fixed exchange rates, efforts to avoid price increases in excess of world inflation must be a matter primarily of fiscal policy, not monetary policy. If budget deficits are not held to levels consistent with world inflation, even very deft deployment of monetary policy instruments are unable to prevent rapid inflation, devaluation, or both. There still is a role for monetary policy in developing countries, but that role must be largely passive. Resourceful use of monetary policy can help by not making things worse and moderating strong

inflationary pressures until the budget can be brought under control, provided the latter is done fairly quickly.

Monetary factors are causes of inflation in both fixed- and floating-exchange-rate countries. In the case of fixed exchange rates, both world monetary expansion and domestic monetary expansion generate inflation; in flexible-exchange-rate countries, inflation arises from domestic monetary sources. But, thus far, no mention has been made of so-called nonmonetary causes of inflation. It seems plausible that internal and external shocks, such as those arising from widespread crop failure in the domestic economy or a drastic increase in prices of imported energy, could have important effects on inflation in countries suffering such shocks. This is true, but the mechanism whereby nonmonetary factors may initiate or worsen inflation needs to be clearly portrayed.

Nonmonetary disturbances indeed may precipitate policy reactions that lead to domestic monetary expansion large enough to accommodate higher relative prices of food or oil and large enough to cause inflation. In the absence of accommodating monetary expansion in the face of such shocks, inflation can be contained, but at some cost. Failure to allow the money supply to expand to accommodate higher relative prices of important goods leads to increases in unemployment that most governments find unacceptable. Governments in such cases usually attempt to allow monetary expansion sufficient to avoid unwanted consequences for employment. But it is important to remember that, however advisable monetary accommodation may be on social and employment grounds, expansion in the money stock is required to fuel inflation, whatever external or internal factors may precipitate the expansion. This truth, known for centuries, is often incorrectly interpreted to mean that nonmonetary factors cannot "cause" inflation. They can, but only through an expansion of the national or international stock of money or both.

CONTROLLING INFLATION THROUGH MONETARY POLICY

The array of available instruments for anti-inflationary monetary policy in developed countries include (1) open-market operations, in which the central bank can directly contract bank reserves by sales of government securities;[6] (2) increases in legal reserve requirements of banks, so that a given volume of reserves support a lower stock of money (and reduce the credit expansion multiplier as well); (3) increases in rediscount rates, so that commercial bank borrowing from the central bank becomes

[6]Open-market operations are used as an instrument of monetary policy in countries with well-developed financial markets. When the Federal Reserve System in the United States or a central bank in Europe wants to curtail the growth of the money supply, it sells government securities (bonds, bills) on the open market. When a buyer pays for the securities, the effect is to reduce directly the reserves of the banking system because the funds are transferred from commercial bank deposits or household cash holdings to the account of the Federal Reserve. When the Federal Reserve wants to expand the money supply, it buys securities on the open market and thus directly adds to bank reserves.

less attractive; and (4) moral suasion, by which the exhortations of monetary authorities are expected to lead to restraint in bank lending policies.

For virtually all developing countries, the first instrument (open market operations) is not available for inflation control. Securities markets typically are absent or not sufficiently well developed to allow the exercise of this powerful and flexible instrument, although some emerging market economies can now use this tool. The other three monetary policy instruments are employed, with varying degrees of success, in developing countries. In addition, developing countries often resort to two other tools employed only infrequently in developed countries: (5) credit ceilings imposed by the central bank on the banking system and (6) adjustments in allowable nominal rates of interest on deposits and loans.

RESERVE REQUIREMENTS

All central banks require commercial banks to immobilize a portion of their deposits in the form of legal reserves that may not be lent to prospective customers. For example, legal reserve requirements for Indonesian and Malaysian banks in the late 1970s were expressed as 30 percent of deposits in domestic currency in the former and 20 percent of all deposits in the latter. Banks in Malaysia were required to add 20 units of currency to reserves for every 100 units of deposits. These figures are not too far out of line with legal reserve requirements in many industrial nations, where reserve ratios of 15 percent for demand deposits and 5 percent for time deposits are common.

Increases in reserve requirements can be used to help moderate inflation. An upward adjustment in reserve requirements works in two ways: It reduces the stock of money that can be supported by a given amount of reserves, and it reduces the money multiplier. The first effect induces banks to contract credit outstanding; the second reduces the growth in the money stock possible from any future increment to reserves.[7] Changes in legal reserve requirements are usually employed only as a last-ditch measure, although China raised reserve requirements several times during the 2000–10 period, when price increases were still modest but rising. Even small changes in the required ratio of reserves to deposits can have a very disruptive impact on commercial bank operations, unless banks are given sufficient time to adjust.

CREDIT CEILINGS

In some countries, such as Indonesia from 1947 to 1983; China in 1994 to 1996 and during the first part of the twenty-first century; and at various other times in Malaysia, Sri Lanka, and Chile, credit ceilings have been used as supplementary instruments

[7]In its simplest form the money multiplier (m) can be expressed as

$$m = (c + 1)/(c + k)$$

where c = the ratio of currency outside banks to deposits and k = the ratio of reserves to deposits. If k is raised, then m falls.

of inflation control. The International Monetary Fund (IMF) often requires countries seeking balance-of-payments support to adopt credit ceilings as a prerequisite for assistance. General ceilings of domestic credit expansion represent a useful method of controlling growth in the domestic components of the money supply. Credit ceilings, however, do not allow full control of money-supply growth in developing countries operating under fixed-exchange-rate regimes because the monetary authorities have no control over foreign components of the money supply. Nevertheless, general credit ceilings sometimes can be usefully deployed in combating inflation in countries not experiencing major imbalances in external payments. Unfortunately, ceilings work the least well where they are needed the most because countries attempting to deal with chronic inflation are usually those experiencing the most destabilizing changes in their international reserve positions. General credit ceilings are unlikely to have much effect on inflation unless the government simultaneously takes steps to reduce the budgetary deficits that, except in major oil-exporting countries, typically are the root causes of chronic, acute, and especially runaway inflation.

Countries often supplement general credit ceilings with specific ceilings on lending to particular sectors of the economy. Indonesia attempted to fine-tune credit controls in this way from 1974 to 1983, with poor results. The system of ceilings was so detailed and cumbersome that domestic banks were unable to come close to exhausting the ceilings. Excess reserves rose. The banks had little choice but to place their excess reserves in deposits overseas, primarily in banks in Singapore. Many domestic firms in Jakarta were then forced to seek credit from Singapore banks, which held well over a billion dollars of deposits from Jakarta banks that might have been lent to domestic firms at a lower rate in the absence of credit ceilings. China, in the first part of the twenty-first century, placed ceilings on credit to real estate developers in what was believed to be an overheated real estate market but with only limited impact on investments in this area.

INTEREST-RATE REGULATION AND MORAL SUASION

In most industrial countries, the central bank can influence interest rates by varying the *rediscount rate* charged on central bank loans to commercial banks that require additional liquidity. Because the rediscount rate is central to commercial banks' operations, it is important in determining the market rate of interest on both deposits and loans. As more developing countries adopt financial reforms that free interest rates from central bank control, they are better able to use the rediscount rate as a tool for influencing market interest rates.

In developing countries that controlled rates on loans and deposits, the controlled rates have been instruments of anti-inflationary packages. The use of such interest rate adjustments have been common in Latin America, and increases in deposit rates and loan rates were major elements in the stabilization programs of South Korea and Taiwan in the mid-1960s and Indonesia in both 1968 and 1974. The

objective in each case was twofold: to stimulate the demand for liquid assets and to discourage the loan demand for marginal investment projects by private-sector borrowers. The extent to which such measures can be successful depends on the interest elasticity of the demand for liquid assets and the interest elasticity of the demand for loans. In most of the cases just cited, particularly in the three Asian countries, both sets of elasticities evidently were sufficiently high because the stabilization packages succeeded to a large degree. Interest rate increases, however, were less effective in China in the 1990s and during 2000–09, largely because many firms, particularly state-owned or -controlled enterprises, are not very sensitive to the impact of higher interest costs on their investment plans.

Moral suasion by the monetary authorities, sometimes called *open-mouth operations* or *jawbone control*, is practiced no less extensively in developing than in developed countries. Warnings and exhortations to commercial banks to restrict lending or to encourage them to focus lending on particular activities have been quite common in Ghana. They also were used at various times in Malaysia, Singapore, Brazil, and elsewhere, sometimes before the imposition of credit ceilings and often to reinforce pressure on banks to adhere to ceilings. In both developed and developing countries, however, moral suasion has proven credible only when accompanied by forceful use of more tangible instruments of monetary control.

INTERNATIONAL DEBT AND COMBATING RECESSIONS

Our discussion to this point has focused on the role of monetary instruments in controlling domestic inflation because that was the dominant problem facing monetary and fiscal policy makers in developing countries since the end of World War II. In this chapter, however, we have not dealt with another major problem that has played an important role in developing countries, the role of accumulating debt obligations by many countries, particularly in Latin America and Africa. Borrowing from other countries and from international institutions such as the World Bank is also a way of financing a development program when domestic financial resources are inadequate, but it can lead to accumulating international debt to a level that the borrowing country can no longer afford. Because these international debt issues are both important and complex, we devote the next chapter to understanding the origins and solutions to developing country debt problems.

Until recently most developing countries did not have to face a problem common in high-income countries in recessions: the lack of sufficient aggregate demand to keep a country's resources of capital and labor fully employed. Developing countries frequently experienced recessions, but they were typically the result of a fall in the price of their natural resource exports or some other balance-of-payments problem. Increasing demand in these circumstances led not to the employment of unused resources but to either inflation or a worsening balance-of-payments deficit. But in 2008–09, countries that had come to rely on exports of manufactures to Europe and

North America suddenly experienced a sharp decline in the demand for their products, as a severe global recession led to a fall in incomes in high-income countries.

In the United States a standard method for getting out of a recession in previous years had been to rely on the central bank to lower interest rates, which would then lead to an increase in investment in a number of sectors, notably in housing. Even in mild recessions, however, interest rates in developing countries are seldom effective in generating increases in investment demand, largely because financial markets in these areas are undeveloped and many borrowers are not very sensitive to small changes in interest charges. In the case of the global recession of 2008–09 lowering interest rates was not very effective in either high- or middle-income countries. The only alternative for the countries affected was to turn to fiscal policy and to introduce what were called *stimulus packages,* large government expenditure programs paid for by borrowing mostly from the central bank. Among developing countries China had the largest stimulus package of well over US$500 billion if all of the components are included, and similar smaller packages were tried in Vietnam and elsewhere in Southeast Asia. These packages were made possible in part because the countries involved had built up large reserves of foreign exchange and thus did not have to fear that the increased demand would lead to large balance-of-payments deficits that could not be covered from existing foreign exchange reserves. China had enormous reserves of over $2 trillion (over $3 trillion in 2011). Many other countries in Asia, in the aftermath of the 1997–98 financial crisis that had been due in part to a lack of sufficient reserves, had built up their foreign exchange reserves as well. In the 1997–98 financial crisis, the most affected countries could not have introduced large stimulus packages to offset the impact on incomes and employment because their reserves were inadequate for even existing balance-of-payments requirements.

FINANCIAL DEVELOPMENT

Up to this point we have discussed the general role of financial systems in economic development and then turned to how financial systems can be mismanaged in a way that undermines stability in the economy and threatens growth in incomes and employment. High inflation by itself has been a major reason why the economic performance of many developing countries has suffered, and preventing inflation has often been the major focus of policy makers both within developing countries and in the international financial agencies. Controlling inflation is a major tenet of Washington Consensus–style reforms. But there is more to financial development than controlling inflation. The development of a financial system that carries out its major functions efficiently and fairly is an important and integral part of the development process.

SHALLOW FINANCE AND DEEP FINANCE

Policies for **financial deepening** seek to promote growth in the real size of the financial system, the growth of financial assets at a pace faster than income growth. In all but the highest-income developing countries, private-sector financial savings predominantly take the form of currency and deposits in commercial banks, savings and loan associations, postal savings accounts, and in some countries, mortgage banks. For most developing countries, growth in the real size of the financial system is reflected primarily in growth in the share of liquid assets in GDP. In contrast, under **shallow finance**, the ratio of liquid assets to GDP grows slowly or not at all over time and typically falls: The real size of the financial system shrinks. Countries able to mobilize large volumes of government or foreign savings can sustain high growth rates even under shallow finance policies, although even these countries may find financial deepening attractive for reasons of employment and income distribution. But, for countries where mobilization of government savings is difficult and foreign savings scarce or unwanted, deep finance may be essential for sustained income growth. This is because growth in the share of liquid assets in GDP provides an approximate indication of the banking system's ability to increase its lending for investment purposes. We will see that the hallmark of a deep financial strategy is avoidance of negative real interest rates; shallow finance, on the other hand, typically involves sharply negative real interest rates.

Growth in the real size of the financial system enhances its capacity for intermediation, gathering savings from diverse private sources and channeling these savings into productive investment. The need for financial intermediation arises because savings endowments do not necessarily correspond to investment opportunities. Individuals with the greatest capacity to save are usually not those with the entrepreneurial talents required for mounting new investment projects. Except in very simple, rudimentary economies, mechanisms are required to channel savings efficiently from savers to entrepreneurs.

If a country is restricted to self-financing by individuals or companies, many productive opportunities involving high private and social payoffs will never be seized because the resources of even the small number of very wealthy are not unlimited. Innovative, smaller-scale investors are not the only groups that fare poorly where financial intermediation is poorly developed; savers are penalized as well. Let us first examine the case in which even the most basic financial intermediaries, commercial banks, are absent. Under these circumstances, the domestic options open to savers are limited to forms of savings, such as acquisition of gold and jewelry, purchase of land and consumer durable goods, or other relatively sterile forms of investment in physical assets. Alternatively, wealthier savers may ship their savings abroad. The common feature of all such investments is that the resources devoted to them are inaccessible to domestic entrepreneurs who would adopt new technology, begin new firms, or expand production in existing enterprises.

All developing countries, therefore, have financial institutions, however embryonic, to serve as intermediaries between savers and investors, even where these intermediaries are limited to commercial banks that accept checking (demand) and time (savings) deposits from savers, to relend to prospective investors for a short term. Intermediation flourishes under deep finance, but under strategies of shallow finance intermediation is constricted and the financial system can contribute little to further the goals of economic growth. Later, we see that shallow finance may have unintended effects on employment and income distribution as well.

SHALLOW FINANCIAL STRATEGY

Shallow financial policies have a number of earmarks: high legal reserve requirements on commercial banks, pervasive nonprice rationing of credit, and most of all, sharply negative real interest rates. Countries rarely, if ever, have consciously and deliberately adopted strategies of shallow finance. Rather, the repression of the financial system flows logically from certain policies intended to encourage, not hinder, investment.

In developed and developing countries alike, policy makers often have viewed low nominal rates of interest as essential for the expansion of investment and controlled interest rate levels tightly. So long as the supply of investment funds is unlimited, low interest rates foster all types of investment activities, as even projects with low returns appear more attractive to investors. In accordance with that observation and in the belief that low interest rates are particularly essential to assist small enterprises and small farmers, governments often have placed low ceilings on nominal interest rates charged on all types of loans. These low ceilings are quite apart from special credit programs involving subsidized credit for special classes of borrowers. Because financial institutions ultimately must cover costs (or else be subsidized by governments), low legal ceilings on nominal loan rates mean low nominal interest rates on deposits as well.

As long as inflation is held in check, low ceilings on nominal loan and deposit interest rates may not retard growth, even when these ceilings are set below the opportunity cost of capital. The United States over the period 1800–1979 managed rather respectable rates of income growth, even in the presence of a set of archaic usury laws and other interest rate controls that (particularly before 1970) often involved artificially low, administered ceilings on interest rates. Even so, throughout most of the period before 1979, real interest rates in the United States remained positive; periods in which real interest rates were sharply negative were intermittent and confined to wartime (1812, 1861, 1917–18, and 1940–46).[8]

Usury laws and other forms of interest rate ceilings also have been common in developing countries. Financial officials in many developing countries, observing

[8]Steven C. Leuthold, "Interest Rates, Inflation and Deflation," *Financial Analysis Journal* (January–February 1981), 28–51.

gross imperfections in financial markets, have concluded that the market should not be permitted to determine interest rates. Monopoly (or oligopoly) power in financial markets, particularly in commercial banking, provides ample scope for the banks and other lenders to exercise market power in setting interest rates on loans at levels higher than the opportunity cost of capital.

There are ample observations of gross imperfections in financial systems in developing countries. Barriers to entry into banking and finance often allow a few large banks and other financial institutions to possess an inordinate degree of control over financial markets and thus exercise monopoly power in setting interest rates. Often these barriers are a direct result of government policies. Governments have prohibited new entrants into the field, adopted such stringent financial requirements for entry that only the very wealthy could amass the needed capital, or reserved permission for entry to political favorites who were attracted to banking and finance largely by the monopoly returns available when entry was restricted.

In this way, one set of government policies, entry restrictions, helps give rise to the need for extensive controls on prices charged by financial institutions. Typically, these controls take the form of interest rate ceilings imposed to limit the scope of monopoly power in the financial system. Controls by themselves do not necessarily lead to shallow finance. Rather, a combination of rigid ceilings on nominal interest rates and inflation impedes financial development and ultimately retards income growth.

Few economists believe that steeply positive real interest rates are essential for healthy growth in the real size of the financial system. The Chilean experience with very high real interest rates from 1981 to 1983 strongly suggests the opposite. There is no widely accepted answer to the question of what the required level of real interest rates is for a steady development of the financial system. The required real rate differs across countries in different circumstances. In some, financial growth may continue even at 0 or mildly negative real interest rates; for others, moderately high positive real rates of between 3 and 5 percent may be essential.

Apart from a few Latin American countries and Indonesia, most developing countries were able to keep rates of inflation at or below 5 to 6 percent before 1973. Inasmuch as nominal deposit rates typically were between 3 and 5 percent, real interest rates tended to be slightly positive or only mildly negative. When inflation accelerated in many developing countries after 1973, because few countries made more than marginal adjustments in nominal deposit rates, real interest rates turned significantly negative in a number of nations, as Table 12–4 shows for 1980. Negative interest rates endured in a few African and Latin American countries throughout the 1980s but, as the table suggests, are much less common today.

When real interest rates turn significantly negative, the maintenance of low nominal rates for promoting investment and income growth becomes counterproductive. Inflation taxes on liquid financial assets bring real growth in the financial system to a halt. Sharply negative real rates lead to shrinkage in the system, as the demand for liquid assets contracts. Contraction in the financial system results in a reduction in

TABLE 12–4 Real Lending Interest Rates, 1980s and 2008–09

COUNTRY	1980S	2008–09
Low-income countries		
Ethiopia	−18	−17
Tanzania	−18	7
India	4	4
Bangladesh	−5	8
Kenya	1	8
Nigeria	−4	19
Honduras	12	14
Lower-middle-income countries		
Bolivia	2	15
Philippines	0	6
Sri Lanka	−1	10
Indonesia	22	6
Peru	−19	18
Egypt	−4	1
Upper-middle-income countries		
Brazil	–	37
Hungary	1	6
Colombia	9	8
Malaysia	5	13
Korea	−5	2
High-income countries		
United Kingdom	−3	−1
Japan	2	3
United States	6	2

Source: World Bank, *World Development Indicators 2011* http://databank.worldbank.org/ddp/home.do, accessed June 2011.

the real supply of credit and thus constricts investment in productive assets. Under such circumstances, nonprice rationing of investible resources must occur and can take many forms. In most developing countries, only those borrowers with either the highest-quality collateral or the "soundest" social and political connections or those willing to make the largest side payments (bribes) to bank officers are successful in securing finance from the organized financial system. These criteria do not yield allocations of credit to the most productive investment opportunities.

Negative real interest rates make marginal, low-yielding, traditional types of investment appear attractive to investors. Banks and financial institutions find such projects attractive as well because they may be the safest and the simplest to finance and involve the most creditworthy borrowers. Satisfying the financial requirements of such investors constricts the pool of resources available to firms, with riskier projects offering greater possibilities for high yields. In addition, in the presence of substantial inflation, interest rate ceilings discourage risk taking by the financial institutions

themselves because under such circumstances they cannot charge higher interest rates (risk premia) on promising but risky projects. Also, negative real interest rates are inimical to employment growth because they make projects with relatively high capital-output ratios appear more attractive than if real interest rates were positive. This implicit subsidy to capital-intensive methods of production reduces the jobs created for each dollar of investment, even as the ability of the financial system to finance investment is shrinking.

Negative real rates of interest tend to lower the marginal efficiency of investment in all the ways described. In terms of the Harrod-Domar model described in Chapter 4, shallow financial strategies cause higher capital-output ratios. Consequently, growth in national income and, therefore, growth in savings tend to be lower than when real rates are positive. Therefore, shallow finance retards income and employment growth even if the interest elasticity of savings is 0. And if savings decisions are responsive to real interest rates, then shallow finance has even more serious implications for income growth, as the ratio of private savings to the GDP also contracts.

DEEP FINANCIAL STRATEGY

Deep finance as a strategy has several objectives: (1) mobilizing a larger volume of savings from the domestic economy—that is, increasing the ratio of national savings to the GDP (where the interest elasticity of savings is thought to be positive and significant); (2) enhancing the accessibility of savings for all types of domestic investors; (3) securing a more efficient allocation of investment throughout the economy; and (4) permitting the financial process to mobilize and allocate savings to reduce reliance on the fiscal process, foreign aid, and inflation.

A permanent move toward policies involving positive real interest rates or, at a minimum, avoidance of sharply negative real rates is the essence of deep finance. In turn, this requires either financial liberalization that allows higher nominal rates on deposits and loans, curbing the rate of inflation, or some combination of both.

Given the difficulties involved in securing quick results in reducing inflation to levels consistent with positive real rates of interest, the first step involved in a shift from shallow to deep financial strategies ordinarily is to raise the ceilings on nominal rates for both deposits and loans. In extreme cases of acute inflation, the initial step has involved raising ceilings on nominal deposit rates to as much as 50 percent in Argentina and Uruguay in 1976 and to nearly 200 percent in Chile in 1974 (where real interest rates nevertheless remained negative until 1976). As the real rate moves toward positive levels, savers strongly tend to increase their holdings of liquid assets; this allows a real expansion in the supply of credit to investors. Available evidence suggests that countries that attempt to maintain modestly positive real interest rates over long periods tend to be among those with the highest rates of financial growth.

Where finance is deep, inflation tends to be moderate; therefore, savers are not subject to persistently high inflation taxes on liquid-asset holdings. That being the case, they are less inclined to shift their savings into much more lightly taxed domestic assets such as gold, land, or durable goods and foreign assets such as currencies or land and securities. Rather, financial resources that otherwise may have been used for these purposes flow to the financial system, where they are more accessible to prospective investors. Nonprice rationing of credit, inevitable under shallow finance, diminishes as well, and the capacity of the financial system to identify and support socially profitable investment opportunities expands. Higher-risk, higher-yielding investment projects stand a far better chance of securing finance under deep than shallow finance. Growth prospects are enhanced accordingly.

Investment finance problems, however, do not end with the provision of a growing real flow of short-term credit. As economies move to higher levels of per capita income, the pattern of investment shifts toward longer horizons. Longer-term investment requires longer-term finance. Commercial banks everywhere are ill-suited for providing substantial amounts of long-term finance, given that their deposits primarily are of a short-term nature.

Therefore, as financial and economic development proceeds, the need for institutions specializing in longer-term finance rises accordingly: insurance companies, investment banks, and ultimately equity markets (stock exchanges) become important elements in financial intermediation. The type of well-functioning commercial bank system that tends to develop under deep finance almost always is a necessary condition for the successful emergence and long-term vitality of institutions specializing in longer-term investment finance.

Earlier we observed that entry into the financial field is rarely easy, and other factors also often lead to gross imperfections in financial markets. Under such circumstances, many developing-country governments have found intervention essential to develop financial institutions specializing in longer-term finance. Intervention may take the form of establishment of government-owned development banks and other specialized institutions to act as distributors of government funds intended as a source of longer-term finance, as in Indonesia and Pakistan. In Mexico and Colombia, governments provided strong incentives for private-sector establishment of long-term financial institutions. Other governments sought to create conditions favorable for the emergence of primary securities (stocks and bonds) markets, the source par excellence for long-term finance. In cases in which these measures have been undertaken in the context of financial markets with strong commercial banking systems (Hong Kong, Singapore, Brazil, and Mexico), efforts to encourage long-term finance have met with some success. In cases in which commercial banking has been poorly developed as a consequence of shallow finance (Ghana, Uruguay before 1976) or the government has sought to "force feed" embryonic securities markets through tax incentives and other subsidies (Indonesia before 1988, Kenya, Turkey), the promotional policies have been less effective.

INFORMAL CREDIT MARKETS AND MICRO CREDIT

The discussion of financial development to this point has dealt with modern credit institutions, the formal market. But in many developing countries, as noted at the beginning of this chapter, **informal credit markets** coexist with modern financial institutions. These markets arise in many forms. In rural India, village moneylenders make loans to local farmers who have no access to commercial banks. In Ghana and other West African countries, market women give credit to farmers by paying for crops in advance of harvest, and they assist their customers by selling finished goods on credit. In South Korea, established lenders actually made loans on the street outside modern banks; this justifies their designation as the *curb* (or *kerb*) *market*. In much of rural Africa wealthy family members make loans to less fortunate kin, and all over the developing world there are cooperative arrangements to raise funds and share credit among members. Even in modern economies, pawnbrokers and others give credit outside the formal credit system.

Informal credit generally is financed by the savings of relatively wealthy individuals, such as local landowners, traders, family members who have moved into lucrative jobs or businesses, and the pooled efforts of cooperative societies. But informal lenders also may have access to the formal banking system and borrow there, to relend to customers with no access to banks. How can they do this if the banks cannot? First, because they know their borrowers so well and may have familial, social, or other ties to them, informal lenders face lower risks than distant, large banks that might loan to the same borrowers. Loan recovery rates are higher (usually much higher than found in large banks in developing countries) because those who borrow in informal markets know that the availability of loans in the future depends on repaying current loans. Second, they also face lower administrative costs in making loans. Of course, moneylenders charge very high interest rates, and this is a third reason they coexist with banks, which are often prevented by law from charging rates high enough to cover the risks and costs of loans in small amounts to very small firms and low-income borrowers.

Beginning in the 1970s, micro-credit institutions and micro-finance institutions more broadly were developed to reach populations in developing countries that had no access to credit, except in informal markets in which interest charges frequently reached annual rates of 100 percent. The best known of these new institutions was the Grameen Bank, founded in 1976 by Muhammad Yunus. Building on innovations such as Grameen, the movement to provide micro credit gained momentum and spread around the world to Africa, Latin America, and Asia where billions of people had no access to reasonable credit. The expansion in micro credit was then supplemented with programs to provide institutions designed to facilitate savings and other financial transactions by the rural poor in developing countries. **Micro finance** was the term applied to this broader set of financial transactions.

The basic idea of micro credit and micro finance was that the poor could be safe credit risks if they were organized appropriately. The Grameen Bank began by

providing micro loans to poor women in the village of Jobra, Bangladesh. Today, Grameen covers virtually all of Bangladesh, with more than 2,500 branch offices and more than 20,000 employees. The Grameen Bank reports that they have over 8 million borrowers in Bangladesh; about 97 percent of whom are women, most are landless, and all are poor. Loans are made to individual women, but only through local groups that provide social pressure for repayment. Loans for micro enterprises average 28,000 taka (about US$400) and carry an annual interest rate of 20 percent. Loans are also given for other purposes, including for housing and education, at 8 and 5 percent, respectively. There is even a special program targeted at beggars, which offers very small loans at 0 interest. Today, Grameen maintains that its loans are not heavily subsidized, neither by donors nor the government of Bangladesh. The recovery rate on Grameen loans is impressive; it exceeds 97 percent and the bank reports it has earned a profit every year since it was founded. The Grameen Bank is more than a financial institution. Its loans require recipients to accept certain "social disciplines," such as cleanliness and family planning, and the bank provides such services as advice on home construction and access to education for some borrowers.[9]

A major problem with many efforts to promote micro credit, however, was that many of the institutions set up to provide such credit depended on subsidies, often quite large subsidies, either from international aid agencies or from private donations from the well to do. This was true of the Grameen Bank earlier in its history. In Indonesia in the 1980s, a government bank, the Bank Rakyat Indonesia (BRI), followed a different path. It began to provide full banking services, both loans and savings deposit facilities, to farmers, traders, and other small-scale borrowers, through their branches in over 3,000 villages. Loans are provided for either working capital or investment purposes. Today, loans range from 25,000 to 25 million rupiahs (from about $3 to $3,000), with repayment periods as short as three months and as long as two years. BRI programs—*kupedes* for the lending program and *simpedes* for the savings program—replaced a government-promoted effort that depended on heavy government subsidies and had a high rate of loan default. These new loans, in contrast, were commercially profitable without any subsidies right from the beginning.

BRI charged market rates of interest, around 30 percent a year, on its loans and paid attractive rates, about 12 percent, on deposits. These rates changed during the financial crisis of 1997–98, largely because of the high inflation in that period, but the basic rate-setting principles remained the same. In the absence of inflation, rates of 30 percent on loans may seem high, but they were considerably below the rates charged by informal money lenders. At these rates, the BRI was able to attract sufficient savings deposits to more than finance its loan program. Several million savers and borrowers mostly in rural areas gained access to the formal banking system for the first time through these BRI programs. Roughly 97 percent of the loans were repaid on time. In addition, the program became a major generator of profits for the

[9]World Bank, *World Development Report 1989*, p. 117; (New York: Oxford University Press, 1989) A. Wahid, *The Grameen Bank: Poverty Relief* (Boulder, CO: Westview Press, 1993).

BRI, something that could not be said about many of its credit programs for large-scale producers and borrowers during the 1997–98 financial crisis.[10] These loans have continued to be a major source of BRI profits in the years since the crisis.

Micro credit is sometimes promoted by its supporters as the answer to rural and urban poverty in developing countries, but as we have emphasized throughout this book, there is no single answer to how to lift the world's poor out of poverty. Micro credit and micro finance have helped poor households in Bangladesh, Indonesia, and elsewhere, but there is considerable debate over their impact on poverty (Box 12–2). While we know that some programs have done well, we do not know whether

[10]Richard H. Patten and Jay Rosengard, *Progress with Profits: The Development of Rural Banking in Indonesia* (San Francisco: ICS Press, 1991); Marguerite Robinson, "Rural Financial Intermediation: Lessons from Indonesia," Development Discussion Paper 434, Harvard Institute for International Development, October 1992.

BOX 12–2 DOES MICRO CREDIT REDUCE POVERTY?

To determine whether micro-credit loans reduce poverty, the experience of borrowers might be compared with nonborrowers. If borrowers, on average, have a lower poverty rate than do nonborrowers, we might conclude that micro credit was an effective instrument for reducing poverty, and the impact of micro credit on poverty reduction could be quantified. But this approach is problematic. We face a classic problem of *identification*. What if borrowers are systematically different from nonborrowers in some unobservable way? Maybe those who obtain micro-credit loans are more entrepreneurial or have a higher tolerance for risk. Their lower poverty rates may have more to do with these attributes than with the loans they received. Borrowers of micro credit might have sought out other types of informal credit, making it difficult to isolate the marginal impact of micro credit on poverty reduction.

Problems of identification like these can be addressed by randomized controlled trials (RCTs), which compare the impact of an intervention between an otherwise similar treatment and control group. Researchers at the Abdul Latif Jameel Poverty Action Lab (J-PAL) located at the Massachusetts Institute of Technology (MIT) conducted an RCT on micro finance in 2005 in the city of Hyderabad, India. Working with an Indian micro-finance firm, Spandana, branches were opened in half of 104 randomly selected slum neighborhoods; in the other half, branches were not opened. Spandana loans are similar to those pioneered by the Grameen Bank. They are offered to women who form groups with 6 to 10 members who are jointly responsible for the loans. Initial loans are for 10,000 rupees (about $200 at market exchange rates and $1,000 in terms of purchasing power parity [PPP]), with an annual percentage rate of 24 percent. Then 15

to 18 months after the introduction of the new branches, detailed household surveys were administered to almost 7,000 households living in the treatment and control neighborhoods to determine the impact of the micro-credit loans.

Because micro-finance firms other than Spandana were available throughout Hyderabad, residents of both treated and untreated neighborhoods had access to micro-credit loans. In neighborhoods where Spandana opened a new branch, 27 percent of residents took out a loan compared to 19 percent in untreated areas. This suggests that the treatment had an effect on borrowing rates. But it also indicates that between 70 and 80 percent of households did not take advantage of Spandana or any other micro-finance loans. The overwhelming majority of households either were not credit constrained or were unwilling to assume the risks associated with taking out a loan. This suggests that it may be a lack of profitable opportunities to invest borrowed funds rather than a lack of cheaper credit that constrains most poor households.

In comparing treated and untreated neighborhoods, the J-PAL study found that 7 percent of households started a new business in neighborhoods where Spandana opened a new branch compared to only 5.3 percent of households in the control areas. Access to micro credit therefore had a significant impact on the formation of new businesses. About one third of Spandana borrowers started a new business, and about one fifth spent their loans on equipment for existing businesses. The other half of borrowers used their loans for current consumption, including on food or medicine, to buy a durable for household use (for example, a television set), or to pay down other debts. In the period covered by the survey, there was no statistically significant increase in total consumption expenditures between treatment and control areas. In other words, micro credit did not reduce poverty rates over this time period.

As the J-PAL study notes, 15 to 18 months may be too short a time to assess the long-term impact of micro credit. The new businesses started as a result of Spandana loans may need more time to earn profits and lift families above poverty. In the short term micro credit did not reduce poverty, nor was it a recipe for improving the schooling or health of children or for empowering women. Of course, the J-PAL study is only one study, and as with all RCTs we must be careful in drawing general conclusions from any one experiment. Evidence from a variety of nonexperimental approaches finds many benefits of micro-credit loans and other micro-finance products. The authors of the Hyderabad study themselves are measured in their conclusions, noting that micro credit "may not be the 'miracle' that is sometimes claimed on its behalf, but it does allow households to borrow, invest and create and expand businesses."

Sources: Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan, "The Miracle of Microfinance? Evidence for a Randomized Evaluation," Massachusetts Institute of Technology, June 2010; Kathleen Odell, "Measuring the Impact of Microfinance: Taking Another Look," Grameen Foundation Publication Series, 2010.

the hundreds and thousands of such programs around the world justify the large outlays of outside aid that they have used. About half of the estimated $12 billion committed to micro-finance organizations worldwide came as grants or concessional loans from aid agencies and donors. These resources might have been spent in other ways and had more of a poverty impact.

Micro credit often has its biggest impact when it helps individuals with limited means to start a new business and gradually expand that business. But there are many other barriers in addition to the lack of credit that can inhibit the development of new businesses and undermine the value of the credit. Many individuals borrow saying the loan is for a small business but use the money for consumption that leaves them with a debt that they may have difficulty repaying. Some borrow funds from one micro-finance organization to pay off loans they secured from another micro-finance organization. Such practices, clearly, are not sustainable for either the borrower or the lender.

Research indicates that profitable micro finance on the BRI model is not a complete substitute for subsidized micro finance in reaching the poorest parts of the population. Profitable micro finance reaches large numbers of people who previously had no access to credit or appropriate vehicles for savings, but many of the programs designed to reach the poorest parts of the population in poor countries and provide other nonfinancial services along with credit also reach people that the for-profit programs do not. There is clearly a role for both kinds of credit facilities.[11]

[11]For a lengthy analysis of these and many other related issues see Beatriz Armendariz and Jonathan Morduch, *The Economics of Microfinance* (Cambridge: MIT Press, 2005); Robert Cull, Asli Demirguc-Kunt, and Jonathan Morduch, "Microfinance Meets the Market," *Journal of Economic Perspectives*, 23, no. 1 (2009), 167–92. Another source that evaluates micro finance is *David Roodman's Microfinance Open Book Blog* (http://blogs.cgdev.org/open_book).

SUMMARY

- The money supply is made up of the liquid assets of an economy but the degree of liquidity of particular assets varies, leading to different, more precise definitions of the money supply.
- The rate of inflation varies greatly among developing economies. Inflation, in effect, is a tax on those who hold money balances, and a moderate rate of inflation can sometimes increase government savings and investment and not harm growth, but the higher the rate of inflation, the more people shift away from liquid assets, thus undermining the development of the financial system and harming economic growth.
- Exchange rate management is an essential component of any effort to control inflation. Exchange rate systems range from fixed exchange rates,

by which the local currency is pegged to the dollar or some other currency, to floating rates that move with market forces. In small countries with fixed-rate systems, worldwide inflation is transferred rapidly to the country with the fixed exchange rate. With floating rates, inflation arises mainly from domestic, not international, sources.

- A variety of additional mechanisms for controlling inflation are at the disposal of the central bank, all of which must, in one way or another, reduce the growth rate of the money supply. Mechanisms such as open-market operations are generally more efficient than mechanisms such as credit ceilings, but most developing countries are not in a position to conduct open-market sales and purchases of government bonds. Credit ceilings and increased bank reserve requirements are less efficient but are effective in lowering the growth of the money supply and inflation.
- Financial panics have been present throughout the last several centuries, but countries have learned how to eliminate some of the reasons for financial panic. As the experience in the late 1990s and 2007–09 of several countries around the world demonstrated, however, there are reasons why panics can still occur.
- The nominal rate of interest is the agreed-on rate between lenders and borrowers, but the real rate of interest, the nominal rate adjusted for inflation, most influences whether individuals are willing to hold liquid assets or not.
- Positive real interest rates are necessary for financial deepening, defined as the rising ratio of liquid assets to GDP. Negative real interest rates have the opposite effect. Financial deepening generally supports growth, although growth sometimes still occurs in its absence.
- As development proceeds and the economy becomes more complex and the need for long-term investment financing increases, economies require a wider variety of financial institutions. Stock markets and bond markets are one source of such long-term finance but so are insurance companies and development banks, and the latter are often supported by government for the explicit purpose of providing longer term finance.
- The formal financial system particularly in developing countries tends not to reach small businesses and individuals with limited financial resources. Informal financial markets exist to provide financing for such potential borrowers, but the interest charges on such loans tend to be very high. To counter this problem a variety of more formal efforts in the form of micro-credit institutions have arisen, providing more reasonable credit terms. The expansion of micro finance since the 1970s has been rapid, but there is still much we do not know about its overall impact on poverty.

13 Foreign Debt and Financial Crises

Beginning in 2007 (and continuing into 2011), much of the world was beset by the largest financial crisis since the Great Depression of the 1930s—a crisis originating in the United States and other advanced economies. Debt crises in several countries have played a central role. In 2010, and again in 2011, Greece, a member of the European Union was on the verge of being unable to meet its debt obligations, and there was fear the problem would spread to Portugal, Italy, and Spain. In 2010, the other members of the European Union, notably Germany, pledged up to $1 trillion to help Greece and, if necessary, the others meet their debt obligations. For the countries in crisis, devaluation of the currency, a policy often used to offset some of the negative effects of having to cut back on expenditures to repay debt, was not available because Greece, Italy, Spain, and Portugal use the common currency of the European Union, the euro. Our focus in this chapter, however, is on debt problems in developing countries, thus we will not deal further with the details of the debt problems of southern Europe and the threat these problems pose for the euro. Nonetheless, it is important to note that the global financial crisis that began in the advanced economies has had a severe impact on developing countries as both trade and financial flows have fallen rapidly in the late 2000s. The Mexican experience provides an informative case with which to begin our discussion of debt and financial crises in developing countries.

In the two decades following World War II, Mexico, like most developing countries, had few transactions with international capital markets. The country borrowed relatively little from international banks, and foreign investors only occasionally took a stake in Mexican firms. And, while it borrowed some funds from official lenders

(governments and international organizations such as the World Bank), Mexico's total borrowing remained small. At the time, international capital markets were relatively undeveloped compared to today, and they were only just beginning to recover from the disruptions of the two world wars and the Great Depression. But things began to change in the late 1960s and early 1970s. Capital market instruments became more sophisticated, and the costs of international transactions fell. Bankers and other investors started to look for opportunities outside their traditional Western markets. Official lending increased as well. Mexico, along with many other countries, was eager to expand investment and accelerate growth and began to look abroad for financing. In the decades since then, Mexico's foreign borrowing has had tremendous impact, both positive and negative, on its economic growth, macroeconomic stability, job creation, and even political dynamics.

Most of Mexico's foreign financing in the early years was in the form of borrowing, especially from private creditors, as opposed to foreign direct investment. Mexico was suspicious of foreign investors and wanted to maintain ownership control of its enterprises, so borrowing was more palatable than foreign direct investment. Foreign commercial banks were more than willing to lend money to the Mexican government, state-owned companies (such as PEMEX, the state oil company), and other businesses. In 1970, Mexico's foreign borrowing from private creditors was relatively small, amounting to $350 million, or about 1 percent of the gross domestic product (GDP). But by 1980, it was borrowing nearly $15 billion a year from private creditors, about 8 percent of the GDP. More than half the debt was short term, meaning that it had to be repaid (and usually reborrowed) within one year.

At first, the borrowing seemed to help: Mexico's GDP per capita rose 4.2 percent per year between 1972 and 1981, compared with 2.6 percent during the previous seven years. But it turned out that Mexico had borrowed too much too quickly: In August 1982, it shocked the world by declaring that it could not service its debts. The country plunged into a deep economic crisis that led to a sharp reduction in investment, negative growth rates, deep strains on government finance, and considerable economic hardship for millions of Mexicans. Mexico's crisis quickly cascaded to many other developing countries that had borrowed heavily abroad, and the consequences of these crises are still playing out today in some countries. Foreign lending to Mexico plummeted after the crisis before recovering in the early 1990s, only to plunge again when Mexico faced another payments crisis in 1994 and 1995. Since the mid-1990s, Mexico's net foreign borrowing has been negative in most years, meaning that it has been repaying international lenders more than it has been borrowing.

Globally, debt flows to developing countries have followed a pattern similar to Mexico's, swinging back and forth from large inflows to large outflows and then back again. Borrowing from private commercial banks became the dominant form of international capital flows to developing countries, especially middle-income countries, in the 1970s and 1980s, surpassing official borrowing in almost every year. But debt flows fell sharply in the 1980s after the debt crises (Table 13–1). In the early

TABLE 13–1 Debt Flows to Developing Countries, 1980–2008 (billion US$)

	1980	1990	1996	2000	2008
Private debt finance	70.9	29.8	119.9	−0.4	107.9
Commercial Banks	28.7	3.2	30.7	−5.8	123.0
Bonds	1.1	1.1	49.5	17.5	10.5
Other	11.5	12.3	2.3	−4.3	−9.3
Short-term debt flows (net)	29.5	13.2	37.4	−7.9	−16.3
Official debt finance	20.5	26.4	2.8	4.9	−20.4
Bilateral	13.1	11.0	−10.4	−6.9	2.4
Multilateral	7.4	15.4	13.1	11.8	18.0
Total	91.4	56.2	122.7	4.5	128.3

Sources: World Bank, *Global Development Finance 2005* (Washington, DC: World Bank, 2005) and World Bank, *Global Development Finance 2009* (Washington, DC: World Bank, 2009); and International Monetary Fund, *International Financial Statistics*, (Washington, DC: IMF, 2009).

1990s, private debt flows grew sharply again, quadrupling in size in just six years from $30 billion in 1990 to $120 billion in 1996. But then a series of financial crises struck several emerging markets around the world, starting with Mexico in 1995; then hitting Thailand, Indonesia, Malaysia, and Korea in 1997; Russia and Brazil in 1998; and Turkey and Argentina in 2001. Debt flows collapsed, and by 2001, the $120 billion inflow of five years earlier had completely disappeared. But once the crises abated, borrowing increased again, and private debt flows rebounded to $72 billion just two years later in 2003. Official borrowing was much smaller and less volatile than private borrowing, but it fell steadily during the 1990s, and by 2003, developing countries as a whole were repaying more to official creditors than they were borrowing. For many of the poorest developing countries, grants, rather than loans, were the dominant source of foreign financing.

Even among the largest borrowers, foreign borrowing has always been small compared to domestic capital flows, in most years accounting for perhaps one fourth or less of available investable funds and only occasionally amounting to more than 3–4 percent of GDP. From the perspective of economic growth models that emphasize the role of capital formation in the growth process (such as the Harrod-Domar and Solow models), private debt flows are only modestly important. But foreign debt plays a more complex role than the basic numbers indicate. The economic impact of debt flows is magnified by their effect on macroeconomic stability, the exchange rate, interest rates, government budgets, and domestic financial institutions and their potential to trigger widespread financial crises.

This chapter explores the advantages, disadvantages, and risks associated with foreign borrowing. It describes various forms of debt, the patterns of debt flows to developing countries over the last several decades, and the basic conditions under which debt is sustainable. It examines in some detail the developing-country debt crises of the 1980s, the ongoing debt problems faced by some low-income countries,

and the emerging-market financial crises of the 1990s. It then explores the main lessons that have emerged from these crises for debt management in the future. We also probe some of the tough questions that arise when countries default or no longer can pay their debt service: How much should be forgiven? How should the burden of bad loans be shared between creditors and borrowers? And what conditions, if any, should be placed on countries seeking debt forgiveness?

ADVANTAGES AND DISADVANTAGES OF FOREIGN BORROWING

Foreign debt gained a bad reputation after the crises of the 1980s, the late 1990s, and the Greek and Irish debt problems of 2010–11, in many ways deservedly so. However, prudent borrowing has been an important part of the development strategy of many countries. The United States borrowed heavily in the middle and late nineteenth century to finance its westward expansion (especially the railroads). Most countries in western Europe relied on foreign borrowing at one time or another.

From a national perspective, borrowing permits a country to invest more than it can save and import more than it can export. If the additional funds finance productive investment, they should yield sufficient returns to pay the interest and principal on the initial foreign inflows. As we saw in Chapters 3 and 4, because capital is relatively scarce in low-income countries, these countries have the *potential* to realize higher rates of return on investment and more rapid economic growth than richer countries, providing the foundation for lending from rich to low-income countries. Under these circumstances, foreign borrowing can help support growth and development at the same time as it yields attractive returns to the lender.

For countries interested in augmenting domestic saving with foreign capital, borrowing brings several advantages and disadvantages relative to foreign direct investment and other forms of capital. When a domestic firm borrows abroad, there are no controversies over foreign ownership, profit repatriation, tax holidays, and the like that arise with foreign direct investment (FDI). In addition, foreign borrowing can be undertaken much more quickly and easily than can FDI. Repayments are limited to the terms of the loan, so if an investment is highly profitable, the home country can keep the profits rather than repatriate them abroad. To simplify only slightly, as long as the rate of return on investment projects exceeds the interest rate on the debt, foreign borrowing can be a very sensible strategy to augment domestic savings, add to investment, and accelerate growth. In addition, borrowing can play a critical stabilization role for countries buffeted by balance of payments shocks. When a country's export prices suddenly collapse or its import prices rise, some temporary borrowing can help bridge the financing gap until prices rebound or the economy can adjust to the change, thus helping ease the costs of adjustment.

However, there are downsides as well. Debts must be paid, even when a project goes bad. Too much borrowing of the wrong kind or for the wrong purposes can leave developing countries vulnerable to sudden capital withdrawals and financial crises. Short-term debt, in particular, can very quickly switch from rapid inflow to rapid outflow, which can cause sudden plunges in exchange rates and skyrocketing interest rates and wreak havoc on banks, private companies, and government budgets.

Borrowing too much to finance consumption or poorly conceived investments also can lead to trouble. Borrowing is much easier politically than raising taxes to pay for spending, especially if governments do not think they will still be in office when it is time to repay. If governments borrow to finance monuments, lavish office buildings, or fleets of expensive cars, there will be no future income stream to repay the debts. But even where governments borrow to finance investments that initially seem sound, sudden drops in export prices, rising interest rates or oil prices, or an unforeseen world recession can turn a worthwhile project into a financial loss. Governments strapped with large debt service payments sometimes face painful choices between repaying foreign lenders as promised, or maintaining spending on health, education, or other important domestic programs.

DEBT SUSTAINABILITY

How much aggregate debt can a country take on before it begins to get into trouble? There is no simple answer. A country's debts are sustainable if they can be serviced without resort to exceptional financing (for example, a special bailout by friendly donors) and without a major future adjustment in the country's income and expenditure. But every country's situation is different. A country's ability to repay depends on a wide array of factors: the size of its debt, its trade and budget deficits, the interest rate on its debt, the mix of loans and grants it receives, its vulnerability to shocks (for example, natural disasters such as droughts or hurricanes or a fall in export prices), and the rate of growth of GDP, exports, and government tax revenues.

At a simple level, a country's debt sustainability depends on (1) how much it owes and (2) its capacity to make the required payments. But neither of these is quite as straightforward as they first seem. There are two ways to think about how much the country owes: the total stock of debt and the amount of payments due in a particular year. The usual measure of the total debt stock is the sum of the face value of all debts outstanding, but this can be misleading if a significant portion of debt is subsidized. For example, the World Bank's International Development Association (IDA), the part of the World Bank that provides finance for low-income countries, provides loans with an interest rate of just 0.75 percent and 40 years to repay. Obviously a $100 million loan extended on these terms creates a much smaller burden than a $100 million loan that must be repaid in 5 years with a market-determined

interest rate of 7 percent. Under these circumstances it makes much more sense to calculate the net present value (NPV) of the debt. A loan extended on normal market terms would have an NPV equal to 100 percent of its face value, a grant would have NPV equal to 0, and subsidized loans would have an NPV in between, depending on the interest rate and the maturity structure (that is, the schedule to repay the loan's principal).

Debt service is the amount due for principal and interest payments in a given year. Policy makers need to know how much is due each year, but looking at just one year can be misleading because it leaves out how much is due the next year or the year after, when a big payment might be due. Thus an accurate projection of how much debt service is due when is an important component of assessing debt sustainability.

What about the second part of sustainability, a country's capacity to pay? Three measures of capacity are most common: GDP, exports, and government tax revenue. GDP is the broadest measure of the economic resources available to repay the debt. The larger a country's productive capacity and corresponding income, the greater its ability to repay debt. But if a country is repaying foreign debt, the real constraint may not be total income, but the ability to earn the dollars (or other foreign exchange) needed to repay the loans denominated in foreign currency. When Thailand, Indonesia, and Korea could not fully service their debts in late 1997, the problem was not a lack of sufficient income, but a lack of available foreign exchange. Under these circumstances, exports earned each year or the stock of foreign exchange reserves held by the central bank may be more suitable indicators of debt servicing capacity. Exports and reserves focus attention on the **external transfer problem**, the challenge of generating sufficient foreign exchange to transfer to external (foreign) creditors. But foreign exchange is not always the binding constraint either. If the government is the major debtor, an important indicator of repayment capacity is its ability to generate tax revenue that can be used to repay the debt. The importance of tax revenues points to the **internal transfer problem**, the government's challenge of raising enough revenue from households and firms to enable the government to repay the nation's debts.

DEBT INDICATORS

Analysts typically turn these broad measures into ratios, with the numerator containing either debt or debt service and the denominator showing a measure of the capacity to repay. The most common debt sustainability indicators include the following:

- *Debt/GDP*, perhaps the broadest measure of debt sustainability, compares total debt to the economy's total capacity to generate resources to repay. A closely related measure uses the net present value: *NPV debt/GDP*.

Where most debt is on market terms (for example, commercial bank loans), the difference between these two measures is small; to the extent debts are contracted on concessional (subsidized) terms, the two differ. Although each country is different, history suggests that debt distress tends to be more likely in countries where the NPV debt/GDP ratio exceeds 30 to 50 percent.[1]

- *Debt/exports* (and NPV debt/exports) compares the total debt to the capacity to generate foreign exchange. A wide range of debt/export ratios might be compatible with sustainability, reflecting the tendency for exports to vary more widely than the GDP. Analysts tend to place threshold levels for individual countries anywhere between 100 and 300 percent, most frequently around 200 percent.
- *Debt/revenue* (or NPV debt/revenue) is most relevant when the government is the largest debtor and there are concerns about its ability to generate tax revenue to repay the loans. Analysts cite ratios of anywhere between 140 and 260 percent as thresholds for individual countries.
- *Debt service/exports* has the great advantage of focusing attention on the amount owed in a single year relative to the export earnings available to make the payments, but it tells less about the overall burden of debt over time. Concern about debt distress tends to grow when this ratio exceeds 20 to 25 percent.
- *Debt service/revenue* focuses on the government's ability to generate tax revenues to make payments due in a single year. The higher this ratio, the more tax revenue that must be devoted to making debt payments and the less available for other government expenditures on health, education, infrastructure, or other purposes. History suggests that debt distress tends to appear once this ratio exceeds 10 to 15 percent.
- *Short-term foreign debt/foreign exchange reserves* focuses on the amount of debt due to be repaid within the next year compared to the available amount of foreign exchange reserves. Analysts suggest that a country is vulnerable to a rapid withdrawal of capital and a financial crisis when this ratio approaches 1:1.

These ratios reflect two broadly different ways that countries can face debt difficulties: insolvency and illiquidity. An **insolvent borrower** lacks the net worth to repay outstanding debts out of future earnings. An **illiquid borrower** lacks the ready cash to repay current debt-servicing obligations, even though it has the net worth to repay the debts in the long term. An illiquid debtor may need cash to make immediate payments, but still has the capacity to repay the debt over time, while an insolvent

[1]International Monetary Fund, "Debt Sustainability in Low-Income Countries: Proposal for an Operational Framework and Policy Implications," February 3, 2004. Available at www.imf.org/external/np/pdr/sustain/2004/020304.htm, accessed March 2012.

debtor does not have the income or assets to repay. The debt/GDP ratio in effect is a measure of overall solvency, indicating the value of debt relative to aggregate economic resources (in theory, ideally we would like a measure of overall wealth—the value of all assets—not just annual income, but this is extremely difficult to measure for a country). The debt service/exports, debt service/revenue, and short-term debt/reserves indicators are measures of liquidity, indicating whether a country has the capacity to make the payments due this year.

These ratios appear at first to be simplistic and mechanical, but each captures and depends on important broader features of both the debt and the economy. To see this, we examine two of the ratios in more depth: debt/exports and debt/GDP.

To explore the debt/export ratio, recall that foreign saving (F) equals the difference between imports and exports of goods and nonfactor services ($M - X$), or roughly the current account of the balance of payments. For simplicity, assume all foreign saving is in the form of borrowing. Therefore, the increase in debt in any year is

$$\Delta D = iD + M - X \qquad [13\text{–}1]$$

where ΔD represents the change in the debt stock and i is the average interest rate. For simplicity, assume that X and M grow at the same exponential rate, g_X. In that case, the stock of debt also grows exponentially at the same rate, and in the long run, the ratio of debt to exports settles at

$$D/X = a/(g_X - i) \qquad [13\text{–}2]$$

where a is the ratio of the current account deficit to exports, $(M - X)/X$, and is a constant, assuming that imports and exports grow at the same rate.[2]

Equation 13–2 tells us that the long-run ratio of debt to exports depends on the size of the current account deficit, the growth rate of exports, and the interest rate. If exports are growing faster than the average rate of interest ($g_X > i$), a country can continue to import more than it exports, meaning that a can remain positive. This should make intuitive sense: Borrowing to cover the gap between imports and exports is sustainable so long as exports are growing more than enough to cover interest payments.

Consider a numerical example. If the current account deficit as a share of exports were 8 percent ($a = 8$), the average interest rate were 5 percent, and exports were growing 9 percent a year, then the ratio of debt to exports would settle at 2 (or 200 percent). If exports grew faster than 9 percent, the debt-export ratio would fall. If, on the other hand, export growth fell to 5.5 percent, the debt/export ratio would explode

[2]Equation 13–1 is solved to yield equation 13–2 by letting $\Delta D = g_X D$ and substituting into equation 13–1. Then equation 13–1 becomes $g_X D = iD + M - X$, from which equation 13–2 can be readily derived (*hint:* start by dividing each side by X). The result is given by Albert Fishlow, "External Borrowing and Debt Management" (220–21) in Rudiger Dornbusch and F. Leslie C. H. Helmers, eds., *The Open Economy: Tools for Policymakers in Developing Countries* (New York: Oxford University Press, 1988).

to 1,600 percent. If it fell below the interest rate (either because export growth fell or world interest rates rose), the numerator a also would have to turn negative; that is, the current account deficit would turn to a surplus. The country could no longer run a deficit and borrow to make up the difference; instead it would have to run a surplus of exports over imports and use the balance to repay debts. This is precisely what has happened to many low-income, high-debt countries, with enormous adverse consequences in some cases.

Turning to the debt/GDP ratio, instead of examining imports and exports, a similar calculation can be made from the investment-saving perspective. Because we know from Chapter 10 that the current account balance $M - X$ is equal to the investment-saving balance $I - S_d$, equation 13–1 can be converted to the following:

$$\Delta D = iD + I - S_d = iD + vY - sY = iD + (v - s)Y \qquad [13–3]$$

In this equation, $Y =$ GDP and v and s are the investment and saving shares of GDP, respectively. As with the trade balance perspective, assume that debt and GDP grow at the same exponential rate, g_Y. Then the long-run equilibrium ratio of debt to GDP is

$$D/Y = (v - s)/(g_Y - i) \qquad [13–4]$$

In this case, if investment exceeds saving by 1 percentage point, the growth rate of GDP (expressed in current dollars) is 7 percent a year, and i averages 5 percent, the debt/GDP ratio settles at 0.5 (or 50 percent). But if GDP slips to a long-run rate of 5.5 percent, the ratio balloons to 200 percent.

Thus a country with poor overall economic performance, reflected in low export or GDP growth rates, is far more likely to get into trouble than a country with better performance. Policy makers must take these factors into account in determining how much a country can borrow and on what terms. Strong debt management, however, goes beyond these basics and includes allowing for the risk of a sudden fall in export receipts, higher interest rates, or other adverse shocks. And, as we shall see, the terms on which the debt is obtained (especially the maturity structure) can have a profound impact on vulnerability to a crisis.

FROM DISTRESS TO DEFAULT

Many developing countries have seen their debt ratios climb to uncomfortable levels at one time or another during the last two decades. Sometimes, even countries with good economic management can get into debt difficulties when they face drops in export prices or unexpected increases in import prices, say, from an increase in world oil prices. As debt-service payment burdens grow and it becomes more difficult to service the debts, countries face several important questions and trade-offs. To what extent should they continue to raise taxes and cut spending to service debts?

At what point do they try to renegotiate the terms of these loans, or in more extreme circumstances, consider outright default?

Most countries rightly see debt agreements as legal contracts in which they have an obligation to fully repay. Most want to avoid default because it can have significant negative consequences, just as filing for bankruptcy can have adverse consequences for a company. A country defaulting on its debts is likely to have much more difficulty borrowing in the future, at least for some period of time, until their prospects brighten. When creditors begin to lend again, they are likely to charge higher interest rates to compensate for the higher risks. Moreover, the process of renegotiation and restructuring can be time-consuming and costly and is something most finance officials would rather avoid. And most countries do not want the stigma and bad publicity that a default might generate.

At the same time, in extreme situations policy makers may see the costs of continuing to service their debts as outweighing the cost of default. Political leaders are willing to ask their citizens to undertake only so much austerity to repay foreign creditors. Moreover, sometimes, it might be fully appropriate for creditors to bear some of the cost of bad loans, especially if they pushed hard to provide funds for questionable projects. To the extent that creditors were partially responsible, they should absorb some of the costs through smaller repayments or even forgiveness of some of the remaining loan. Defaults are not uncommon for either private companies or public entities and certainly are not limited to developing countries (Box 13–1). Defaults occurred frequently in many Latin American and European countries during the nineteenth century and the 1930s, and Greece threatened to default in 2010 before being bailed out by the European Union, only to find itself once again facing a potential default in 2011 and 2012. Several U.S. states defaulted on their debts in the nineteenth century.

In addition, it is possible that continued repayment might undermine a country's ability to make future payment to such an extent that both the debtor *and* the creditor would be better off with some debt forgiveness. This situation is known as a **debt overhang**: The debt creates such a drag on growth that it undermines the ability of the country to make repayments.[3] Inflows of foreign financing should have a positive impact on growth, but debt service has a negative impact. As debt service grows, the negative impact becomes larger, and with debt service large enough, the overall impact on growth could be negative.

To see how this could happen, consider the following scenario. A government starts to raise taxes to repay the debt, and private companies and individuals begin to anticipate even higher taxes in the future, so they reduce investment. The growing

[3]Paul Krugman, "Financing versus Forgiving a Debt Overhang," *Journal of Development Economics* 29 (1988), 253–68; Jeffrey Sachs, "The Debt Overhang of Developing Countries," in R. Findlay, G. Calvo, P. Kouri, and J. Braga de Macedo, eds., *Debt, Stabilization and Development: Essays in Honor of Carlos Diáz Alejandro* (Oxford: Basil Blackwell, 1989).

BOX 13-1 A SHORT HISTORY OF SOVEREIGN LENDING DEFAULT

In 1979, then-Citicorp chairman Walter Wriston famously pronounced that "countries don't go bankrupt." While it may technically be true that a country cannot go bankrupt, the assertion that they cannot default runs counter to the historical record. Default by governments is a practice as old as the concept of credit. Throughout history, there are many examples of countries that have refused to pay their bills or unilaterally written off debts incurred by previous governments.

The first such recorded default occurred in the fourth century B.C.E. when 10 of 13 Greek city-states with debts to the Delos temple walked away from their contractual obligations. Not long after, the island of Chios announced publicly that payments on its unsustainable debt would cease until economic conditions improved. Default in ancient times often took the form of currency depreciation, rather than a declaration of bankruptcy. For example, over the course of the three Punic Wars (241–146 B.C.E.), Rome reduced the metallic content of its monetary unit from 12 ounces to 0.5 ounce, in a series of de facto government defaults.

The practice of governmental default continued through the Middle Ages and into modern times. An eighteenth-century French minister of finance contended that "each government should default at least once every century, in order to restore equilibrium." In the nineteenth century, as the practice of lending abroad became more common, government default increased, and most European nations at least partially defaulted on their debt commitments. Some defaulted multiple times, with Spain and its seven recorded defaults leading the way. The record in developing nations was similar: Every Latin American nation without exception defaulted during the nineteenth century.

While the U.S. government avoided outright default, many individual states defaulted during this period. Some defaulted on civil war debts, while others did so on bonds issued to failed enterprises, usually railroad or bank endeavors. Arkansas and Florida each defaulted three times during the nineteenth century. And at the conclusion of the Spanish-American War in 1898, the U.S. government repudiated the debts that had been incurred by Cuba while under Spanish rule.

Germany's reparation obligations after World War I led to protracted debt difficulties that nurtured German grievances in the early 1930s. In the wave of defaults that accompanied the Great Depression of the 1930s, the international capital market collapsed, thereby leading to further widespread default. International capital flows fell sharply during the Depression and World War II,

so sovereign defaults were rare in the late 1940s and the 1950s. But the international capital markets gradually revived in the 1960s and 1970s, and the debt crises of the 1980s saw a return to the sovereign defaults that had been common in the nineteenth century.

Adapted from Nancy Birdsall and John Williamson (with Brian Deese), *Delivering on Debt Relief: From IMF Gold to New Aid Architecture*, Center for Global Development and Institute for International Economics, (Washington, DC: Peterson Institute of International Economics and Center for Global Development, April 2002).

debt service burden makes lenders more reluctant to provide new finance, as they begin to fear the possibility of default. As debt burdens grow very large, the government begins to face perverse incentives against undertaking stringent adjustment measures. Economic reforms may help the country avoid default, but local citizens bear the cost of adjustment while creditors gain the benefits. Worse policies might lead to poorer economic performance but could lead to a larger debt write-off. Thus the debt burden becomes a disincentive for the very economic reforms that might be most needed. Empirical evidence on the existence of a debt overhang is mixed, but the idea has been influential in thinking about the rationale for debt restructuring and write-offs in developing countries.

Finally, while debt agreements are legal contracts, under some circumstances, governments may see the agreements as illegitimate, making default an easier option to contemplate. Debts taken on by a military dictator who took power through a coup and stole the money may not be seen as legitimate obligations by a succeeding democratically elected government. We discuss these **odious debts** later in the chapter.

THE 1980s DEBT CRISIS

The questions about default and debt restructuring are much more than hypothetical possibilities: They became very real issues for many developing countries during the last several decades. During the 1970s a large number of developing countries, especially in Latin America, borrowed extensively and accumulated large amounts of debt, so large that, by 1983, long-term debt owed by developing countries to commercial banks had grown from $19 billion to $307 billion, 16 times larger than it had been in 1970 (Table 13–2). Total debt stocks grew by a factor of 10 between 1970 and 1983, then doubled again in the 10 years that followed. But the rapid growth of debt in the 1970s and early 1980s led to trouble for some countries that had accumulated too much debt too fast. In the first half of 1982, eight countries had to reschedule their

TABLE 13–2 Long-Term Debt, All Developing Countries, 1970–2009 (billion US$)

STOCKS	1970	1983	1993	2003	2009
All sources	61	618	1,316	1,960	2,759
Official creditors	33	218	697	805	764
Private creditors	28	400	619	1,155	1,995
Of which, commercial banks	19	307	304	580	136

Sources: World Bank, *Global Development Finance 2005* (Washington, DC: World Bank, 2005); and World Bank, *Global Development Finance 2011* (Washington, DC: World Bank, 2011).

debt payments (that is, negotiate new terms to stretch out repayment). In August 1982, Mexico stunned global markets by declaring that it could not make its debt payments, signaling the beginning of a much broader series of debt crises affecting dozens of countries. From 1983 to 1987, more than $300 billion of debt repayments had to be rescheduled. Now, almost 30 years later, the effects of these crises are still reverberating in some countries, and the lessons learned are crucial for future economic management in developing countries. That much is clear from the fact that the total stock of developing country debt had grown to exceed $2.7 trillion by 2009, nearly $2 trillion of which was owed to private creditors (Table 13–2). (It is interesting to note, however, that the role of commercial banks among private creditors has declined substantially over time.)

CAUSES OF THE CRISIS

The crisis resulted from several things going wrong at once: adverse international economic shocks outside of the control of the debtor countries, poor domestic economic management, and bad lending decisions by international banks.[4]

INTERNATIONAL ECONOMIC SHOCKS In 1973, the Organization of the Petroleum Exporting Countries (OPEC) announced that it would restrict oil production, sparking a very sharp increase in world oil prices. Import values and trade deficits

[4]A huge literature examines the 1980s debt crisis. We cite some of the key works here. Jeffrey D. Sachs, ed., *Developing Country Debt and Economic Performance*, 4 vols.: Vol. 1, *The International Financial System*; Vol. 2: *Country Studies: Argentina, Bolivia, Brazil, Mexico*; Vol. 3 (with Susan Collins): *Country Studies: Indonesia, Korea, the Philippines, Turkey*; Summary vol.: *Developing Country Debt and the World Economy* (Cambridge, MA: National Bureau of Economic Research, 1989–1990). William Cline, *International Debt: Systematic Risk and Policy Response* (Washington, DC: Institute for International Economics, 1984); William Cline, *International Debt Reexamined* (Washington, DC: Institute for International Economics, 1995); Ishrat Husain and Ishac Diwan, eds., *Dealing with the Debt Crisis* (Washington, DC: World Bank, 1989); Joseph Kraft, *The Mexican Rescue* (New York: Group of Thirty, 1984).

immediately increased for oil importers around the world (including developing countries), slowed global economic growth (hampering export markets for many developing countries), and created volatility in a wide range of other commodity markets. In 1979, OPEC again cut back its production, leading to even larger increases in oil prices.

At the same time, the U.S. economy was facing a combination of slow growth and accelerating prices. With the economy under pressure, in August 1971, President Richard Nixon released the dollar from the gold standard that had determined its value since the end of World War II. This move led to a sharp deprecation of the dollar and an increase in currency volatility and uncertainty around the world. Large budget deficits emanating from spending on the Vietnam War and the sharp hike in world oil prices added to the pressure. By the end of the 1970s, U.S. inflation reached double digits, and interest rates rose very sharply, reaching 16 percent in 1981. The increase in interest rates substantially increased the interest burden for developing country debtors, whose original loans were written with variable interest rates. For a group of 15 countries in Central and South America, the rise in interest rates between 1978 and 1981 added over $13 billion to the costs of servicing their debts in 1981 alone.

DOMESTIC ECONOMIC POLICIES Even in the face of these shocks, not all debtors suffered equally. South Korea and Indonesia were among the world's largest debtors, yet sound economic policies enabled them to service their debt while continuing to grow throughout the 1980s. In particular, these countries responded by reducing their budget deficits, restraining the expansion of domestic demand, and encouraging export production. But in many other countries, reactions by policy makers exacerbated the crisis.

In the face of rising imports and trade deficits, escalating prices, and growing budget deficits, it was much easier for governments to convince themselves that the oil crisis would be short-lived rather than try to limit demand. Many believed they could borrow to cover the deficits (from either their central banks or from foreign banks) and even increase spending, rather than restrain demand and close the deficits. But this strategy, while perhaps politically expedient in the short run, could not be sustained and typically led to even higher rates of inflation and larger debts.

At the same time, some governments tried to compensate for rising import prices by maintaining an overvalued exchange rate, which restrains price increases for imports and exportable goods. This step is popular with consumers because it makes imports cheaper, but it discourages export growth and encourages capital flight. More generally, countries that had been actively promoting manufactured exports, mainly in East Asia, were in a better position to further spur exports and close their deficits than countries that followed a less open trade strategy, such as in Latin America. Ultimately, countries with overvalued exchange rates and less-active export

promotion strategies faced larger current account deficits and, therefore, larger borrowing needs. Instead of improving, their debt crises just got worse.

IMPRUDENT BANK LENDING The foreign banks also bear some responsibility for making poor lending decisions. Banks were more than happy to keep lending to governments, even after the economic situation began to deteriorate and normal lending strategies suggested more prudence. Banks believed that, because the debts were **sovereign** (that is, either contracted or guaranteed by governments), they did not have to worry as much about the normal risks of default. And once the banks were heavily involved, further lending seemed a sensible way to keep debtor countries liquid enough to continue servicing earlier loans. Banks were encouraged by their own governments and the international agencies, all of which hoped that more lending would help countries grow out of the crisis, especially once the world economy recovered from its instability.

IMPACT ON THE BORROWERS

Once it became clear that many debtors would be unable to meet their obligations, the commercial banks stopped making new loans. From 1980 to 1982, private creditors provided more than $50 billion a year in new lending to developing countries; in 1987, the net resource flow was essentially zero.

The impact on the most indebted countries was severe. In effect, countries very quickly had to turn from net borrowers into net repayers of loans. We know from balance of payments accounting that net capital inflows always must equal the current account deficit, which in turn is equal to the balance of investment over domestic saving:

$$\text{Net capital inflows} = M - X = I - S_d \qquad [13\text{–}5]$$

Before the crisis, foreign borrowing (and other capital inflows) allowed the countries to finance imports in excess of exports and investment in excess of saving. But when countries began to default on the loans, banks demanded immediate repayment and reduced new lending. In the absence of other new capital inflows, countries had to reverse the signs and run a surplus of exports over imports and of saving over investment. Economic growth fell abruptly. Many governments found it difficult to raise taxes and reduce spending (to increase domestic saving), so they resorted to inflationary financing, which further destabilized the economy.

These kinds of dramatic changes were clearly evident in Mexico before and after the 1982 crisis, as shown in Table 13–3. In the five years up to and including 1981, imports exceeded exports in Mexico by the equivalent of about 2 percent of GDP and investment exceeded domestic saving by the same amount. There was strong growth in investment, GDP, and GDP per capita. But the crisis changed everything. Between

TABLE 13-3 Mexico before and after the 1982 Debt Crisis

	BEFORE THE CRISIS, 1977–81	AFTER THE CRISIS, 1982–86
Exports of goods and services (% of GDP)	9.6	16.9
Imports of goods and services (% of GDP)	11.5	10.6
Resource balance (% of GDP)	−2.0	6.3
Gross domestic investment (% of GDP)	24.7	20.4
Gross domestic savings (% of GDP)	22.7	26.7
Saving-investment gap (% of GDP)	−2.0	6.3
Gross domestic investment (real, average annual growth %)	13.0	−12.2
GDP (real, average annual growth %)	7.6	−0.6
GDP per capita (real, average annual growth %)	5.0	−2.9
Inflation, consumer prices (average annual %)	23.7	73.2

GDP, gross domestic product

Source: World Bank, *World Development Indicators*, http://databank.worldbank.org/ddp/home.do, accessed 2005.

1982 and 1986, imports fell and exports expanded sharply, providing the trade surplus necessary to pay foreign creditors. The **resource balance** (the excess of exports over imports, and the opposite of net capital inflows) shifted by a huge 8.3 percent of GDP (from −2 percent to +6.3 percent). Investment fell by over 4 percent of GDP, and domestic saving rose by a similar amount. GDP per capita fell, the budget deficit grew, and inflation skyrocketed to 73 percent per year.

These kinds of adjustments were not limited to Mexico. Between 1978 and 1981, a group of 15 heavily indebted countries imported about $9 billion more than they exported. But from 1983 to 1988, they were forced into a reverse transfer of almost $34 billion a year to service their debt. From 1970 to 1981, these countries had enjoyed per capita income growth of 2.7 percent a year, invested 25 percent of their GDP, and tolerated inflation of 39 percent a year on average. During the debt crisis, however, from 1982 to 1988, GDP growth per capita was −0.7 percent a year, investment fell to 18 percent of GDP, and annual inflation grew to 149 percent.[5]

ESCAPE FROM THE CRISIS, FOR SOME COUNTRIES

For the international commercial banks and many of the countries that had borrowed from them, especially the middle-income countries, the debt crisis of the 1980s essentially was over by the mid-1990s. For other countries (mainly low-income

[5]Rudiger Dornbusch, "Background Paper" (31), in Twentieth Century Fund Task Force on International Debt, ed., *The Road to Economic Recovery* (New York: Priority Publications, 1989).

countries in sub-Saharan Africa, South Asia, and Central America) that had borrowed predominantly from other governments and the international financial institutions, the debt crisis lingered unresolved into the new century.

Debt ratios improved substantially for Latin American countries from the 1980s to the 1990s and were much lower by 2005 compared to their previous high levels (Table 13–4). In most cases, improvements continued through 2009, albeit at a slower pace. Latin American debtors worked out of insolvency through debt relief agreements that involved a combination of several forms of debt restructuring and reorganization:

- *Refinancing,* involving making new loans to repay the old. In a refinancing, the amount of debt owed does not change, but the terms for repayments are eased, usually through longer repayment periods and perhaps lower interest rates.
- *Rescheduling,* closely related to refinancing, in which the original loans stay on the books, but the schedule of payments is altered to allow longer repayment periods and possibly lower interest rates.
- *Reduction,* in which the amount actually owed is reduced (that is, forgiven), either partly (a write-down) or completely (a write-off).
- *Buybacks,* whereby the debtor buys the loan from the creditor, usually for a percentage of the face value of the debt. Creditors might prefer getting an ensured payment today for part of the debt rather than taking the risk of less payment later.

TABLE 13–4 Debt Ratios, Developing Countries, 1980–2009 (percent)

	MAXIMUM OF 1980–97	2005	2009
All developing countries			
Long-term debt to GNP	43	26	22
Long-term debt to exports	203	75	75
Total debt service to exports	33	13	11
Latin American countries			
Long-term debt to GNP	60	29	24
Long-term debt to exports	343	111	111
Total debt service to exports	48	24	18
Selected countries: total debt service to exports			
Argentina	83	19	17
Bolivia	63	16	14
Brazil	82	46	23
Chile	71	15	23
Mexico	51	22	13

GNP, gross national product;.

Sources: World Bank, *Global Development Finance 2004* (Washington, DC: World Bank, 2004); and World Bank, *Global Development Finance 2011* (Washington, DC: World Bank, 2011).

- *Debt-equity swaps,* in which creditors are given equity in a company (such as a state-owned telecommunications company) in return for eliminating the debt outstanding.

Strategies to resolve the 1980s debt crisis evolved slowly in two broad stages. In 1985, nearly three years after Mexico first defaulted, U.S. Secretary of the Treasury James Baker announced what came to be known as the **Baker Plan**. This two-pronged strategy involved providing new finance to the debtor countries in return for them undertaking significant economic reforms aimed at stabilizing their economies and reducing their current account deficits. The plan did not reduce the amount of debt; it was based on the assumption that, with enough new financing and economic adjustment, the debtor countries could begin to grow and would fully repay the banks.

Financing was provided in part by the creditor banks refinancing and restructuring existing debts and partly through new funds provided by the International Monetary Fund (IMF), World Bank, and official aid agencies. The new financing provided some breathing room for the debtors because it reduced the amount of immediate adjustment necessary in the current account deficit. But adjustment was still necessary. The policy adjustment programs involved a range of steps aimed at reducing deficits in the government budget and the balance of payments, containing inflation, stimulating savings, and generating more resources for investment and debt service. These **stabilization and structural adjustment programs**, which were designed and overseen by the IMF and World Bank, typically involved devaluing the currency (or allowing it to depreciate through a float) to reduce imports and stimulate exports, cutting government expenditures, raising taxes, reducing growth in the money supply, and closing loss-making state companies.

These programs were controversial, and the austerity measures created huge burdens on the ordinary citizens of debtor countries in the form of reduced income, higher taxes and fees, reduced services from government, and other adjustments. Critics argued that, while some of these steps were necessary, the programs were not always well designed and required more austerity than was necessary while creating little burden for the creditors. Although it was hoped that this combination of new money and policy reforms eventually would allow countries to resume growth and fully repay the banks, in the end, the Baker Plan proved insufficient to resolve the debt crisis.

The key breakthrough came in 1989, when new U.S. Secretary of the Treasury Nicholas Brady unveiled a different strategy, this time incorporating the idea that the value of the debt would be written down and the commercial banks would share in some of the losses. The **Brady Plan** recognized that debtor countries needed a permanent reduction of their debt burdens to get back on sound economic footing. The strategy called for banks and debtor governments to renegotiate debts on a case-by-case basis from a menu of options, including reducing the face value of the debt,

reducing the interest rate, providing new loans, and other forms of restructuring. The debtor countries issued new bonds to the banks, called **Brady bonds**, to replace the old debt. The amount of debt relief encompassed in these deals varied from country to country. Mexico's commercial bank debt was reduced by about 35 percent, whereas Costa Rica's was reduced by about 65 percent. By May 1994, Brady deals were in place covering about $190 billion in debt for 18 countries, including Poland, Ecuador, Venezuela, the Philippines, Brazil, Peru, the Ivory Coast, and other countries. Over $60 billion in debt had been forgiven. As with the Baker Plan, the new scheme included stabilization and structural adjustment programs under the auspices of the IMF and World Bank, new lending from these organizations, and bilateral debt relief.

The key to the program was the recognition by the international community and the creditor banks that by continuing to insist on full repayment, they were undermining the strength of these economies and ultimately receiving less repayment than they would under a negotiated settlement. By the early 1990s, growth had resumed and prices had stabilized. Net capital inflows, after falling to $10 billion per year between 1983 and 1989, rose to $60 billion in 1992. Debt reduction alone could not have solved the crisis, but debt reduction combined with new financing, strong adjustment efforts by the countries themselves, and especially the revival of the world economy in the early 1990s helped bring an end to the crisis, at least for some countries. A similar approach would be adopted for other heavily indebted nations almost a decade later.

THE DEBT CRISIS IN LOW-INCOME COUNTRIES

Although the debt crisis effectively ended in the early 1990s for most middle-income countries that had borrowed heavily from commercial banks, it continued for many low-income countries, especially many countries in Africa. Most of the debts of these countries are owed to either official multilateral organizations, such as the IMF and World Bank, or to donor governments, rather than to commercial banks. Thus, although the terms of the debts were more generous (and usually subsidized), the Brady Plan was not a solution for these countries because it was designed for commercial banks, ultimately providing the creditor with new bonds that it could sell on private bond markets. Different approaches were needed for these countries.

Low-income countries borrowed in the 1970s and 1980s for many of the same reasons as other developing countries. They were hit by the same set of strong international economic shocks: high oil import prices, low commodity prices, the end of the gold standard, and weak market demand in the industrialized countries. These challenges were even more difficult to meet in the poorest countries, which had a limited number of well-trained financial and economic experts, many of which had

gained their independence only recently and had new and untested government institutions. The impact of these shocks was exacerbated in many countries by a prolonged period of economic mismanagement. Governments kept budget deficits high, erected significant barriers to trade, distorted market prices, and failed to provide basic infrastructure and health and education services.

Even where economic policies later improved, however, growth remained elusive, partly because many of these countries also face very difficult geographical challenges, such as being landlocked, located in the tropics (where disease is much more virulent), or located in or near the Sahara Desert. Several currently are ravaged by the HIV/AIDS pandemic, which further undermines their capacity for sustained economic growth and repayment of debt.

As with the debt crisis in Latin America, creditors sometimes added to the crisis by making loans for bad projects (many designed by the very same donors that provided the loans), although in this case the major creditors were government and multilateral agencies rather than commercial banks. In some cases industrialized country governments provided loans for poor investments and grandiose consumption projects in an attempt to win countries over to one side or the other of the cold war. Some Western loans supported corrupt governments that had little intention or capability of repaying the loans to finance presidential palaces, showcase steel mills, and other wasteful expenditures.

During the 1980s and early 1990s, the international community's strategy for the low-income countries had two components, similar to the original Baker plan: provide new finance in return for significant economic reforms in the hopes that the debtor countries could grow out of their difficulties. Beginning in 1988, the financing was augmented by partial write-downs from government creditors. Multilateral creditors, including the IMF, World Bank, and other multilateral organizations (such as the African Development Bank and Asian Development Bank) offered no write-downs until 1996.

DEBT REDUCTION IN LOW-INCOME COUNTRIES

Individual creditor governments provide debt rescheduling and debt reduction through an informal group called the **Paris Club**. The governments of the United States, the United Kingdom, Japan, Germany, France, and 14 other creditor countries coordinate to offer common terms for debt restructuring to each debtor. The first meeting was in 1956, when Argentina met with a group of its creditors in Paris. Since that time, the Paris Club has negotiated over 400 agreements with at least 80 different countries. Until 1988, all Paris Club agreements involved debt rescheduling rather than reduction. Since then, agreements typically provide debt rescheduling for middle-income countries and partial debt reduction coupled with rescheduling for low-income countries. The first debt reduction deals in 1988 provided 33 percent reduction for certain qualifying debts; by 1999, the Paris Club was

providing 90 percent reduction for at least some debts owed by certain countries. Many creditor governments have gone further and provided 100 percent debt relief in some circumstances.

The arguments for providing debt relief to the poorest countries were similar but not identical to those made for the middle-income countries. The first argument was a variant of debt overhang: The debts were so large that they were impeding growth and development, partially by undermining private sector incentives but also by impeding the government's ability to make critical investments in health, education, and infrastructure.

The second argument was that the creditor governments and institutions should not have been providing so much of their financing as loans to the world's poorest countries in the first place because history suggested that accelerating growth sufficiently to repay the loans would be a major challenge. The world's poorest countries have had the most difficulty in initiating and sustaining growth, for reasons that we discussed: adverse geography, disease burdens, weak institutions, frequent climate and trade shocks, and poor policies. Providing loans to these countries rather than outright grants was based on the presumption that growth could be turned around relatively easily, providing these countries with the basis to repay their loans. But for many of the world's poorest countries, growth rates remained low, debts piled up, and many countries were able to service their loans only by receiving new loans.

The third argument is that creditors knowingly lent money to dictatorships to garner their political support, even when the dictators wasted and stole the money. Once the dictators left, the citizens of the country were left with the responsibility of repaying these debts. The Democratic Republic of the Congo, formerly Zaire, piled up large debts under the ruthless dictatorship of Mobutu Sese Seko, as Western donors happily lent him more money, even when it was obvious that most of the money was wasted or stolen. Many argue that *odious debts* (obligations accumulated by an illegitimate government in the name of a country) should be forgiven, with the costs born by the creditor (Box 13–2). The United States, which often has resisted attempts by some countries to claim odious debts, pushed hard in 2004 for other countries to forgive debts owed by Iraq. Although the United States was careful to avoid the term *odious debt* because of the legal precedent, it nonetheless argued that Iraq's debts were amassed by an illegitimate dictator and were a large impediment to future growth. In November 2004, the Paris Club agreed to cancel over $29 billion of Iraq's debts and reschedule an additional $7 billion, by far the largest debt forgiveness operation ever undertaken for a single country.

THE HEAVILY INDEBTED POOR COUNTRY INITIATIVE

By the middle of the 1990s, it was becoming increasingly clear that some of the poorest and most heavily indebted countries would not be able to repay their debts. Most of these countries almost never actually defaulted on their debt payments: Instead,

BOX 13-2 ODIOUS DEBT

Under the law in many countries, *individuals* do not have to repay if others fraudulently borrow in their name and *corporations* are not liable for contracts that their chief executive officers enter into without proper authority. The legal doctrine of odious debt makes an analogous argument that sovereign debt that is incurred without the consent of the people and that does not benefit the people should not be transferable to a successor government, especially if creditors are aware of these facts in advance. But many developing countries carry debt incurred by rulers who borrowed without the people's consent and who used the funds either to repress the people or for personal gain.

The doctrine of odious debt originated in 1898 after the Spanish American War. During peace negotiations, the United States argued that neither it nor Cuba should be responsible for debt the colonial rulers had incurred without the consent of the Cuban people and had not used for their benefit. Although Spain never accepted the validity of this argument, the United States implicitly prevailed, and Spain took responsibility for the Cuban debt under the Paris peace treaty.

Though legal scholars elaborated the details of the doctrine of odious debt, it has gained little momentum within the international legal community—still, many countries could qualify. For example, through the 1980s, South Africa's apartheid regime borrowed from private banks, devoting a large percentage of its budget to financing the military and police and repressing the black majority. The South African people now bear the debts of their repressors. Despite appeals—from the archbishop of Cape Town and South Africa's Truth and Reconciliation Commission—to have the apartheid-era debt written off, the post-apartheid government has accepted responsibility for it, perhaps out of fear that defaulting would make the country seem not to be playing by the rules of capitalism and would hurt its chances of attracting foreign investment. South Africa is not poor enough to qualify for debt relief under the HIPC Initiative.

Among other dramatic instances of odious debt, Anastasio Somoza was reported to have looted $100 to $500 million from Nicaragua by the time he was overthrown in 1979; Ferdinand Marcos amassed a personal fortune of $10 billion in the Philippines; Mobutu Sese Seko expropriated a reported $4 billion from the Democratic Republic of the Congo (then Zaire); and Jean-Claude Duvalier reportedly absconded with $900 million from Haiti. Some of the money that built these fortunes was drawn from amounts that these dictators borrowed in the name of their people. Recently, the United States successfully argued that debts accumulated by Saddam Hussein in Iraq were illegitimate and should be

forgiven, although, to avoid setting a legal precedent for other countries to follow, the United States did not use the *odious debt* terminology. The international community forgave over $29 billion in Iraqi debt in 2004.

Economists Michael Kremer and Seema Jayachandran have argued for the creation of an independent institution to assess whether regimes are legitimate and declare sovereign debt incurred by illegitimate regimes odious and thus not the obligation of successor governments.

Adapted from Michael Kremer and Seema Jayachandran, "Odious Debts," *Finance and Development* 39, no. 2 (June 2002).

as the debt payments grew, the creditors lent them new money so they could repay the old, a strategy known as **defensive lending** (because the loans defend the creditor against the possibility of the debtor defaulting). But this strategy could not be sustained indefinitely, and debt levels continued to grow while growth rates remained stubbornly low.

In 1996, the international community recognized that deeper debt reduction would be necessarily. The World Bank and IMF, in conjunction with other multilateral agencies and creditor governments, launched the rather awkwardly named Heavily Indebted Poor Countries (HIPC) Initiative.[6] The HIPC include 38 countries with per capita incomes below $900 (converted using official exchange rates rather than purchasing power parity [PPP]) and with NPV debt/export ratios exceeding 150 percent.[7] Most had much higher debts. From 1982 to 1992, the NPV debt/export ratio deteriorated from an average of 266 percent to 620 percent for the HIPC. The new initiative had two key features, deeper debt relief from bilateral government creditors (initially up to 80 percent of qualifying debts, then in 1999 up to 90 percent), and for the first time ever, partial write-downs of debts owed to the multilateral agencies.

The first version of the HIPC initiative turned out to be too cumbersome and time-consuming to be effective. In the first three years, only five countries received debt relief, and the amounts involved were relatively small. In 1999 a new, enhanced HIPC Initiative was introduced to provide more relief and provide it more quickly.

To be eligible for debt reduction, a country must first establish a track record of good economic policies, as determined by the IMF and World Bank, usually

[6]For a thorough analysis of the HIPC program and related debt issues, see Nancy Birdsall and John Williamson (with Brian Deese), *Delivering on Debt Relief: From IMF Gold to New Aid Architecture* (Washington, DC: Center for Global Development and Institute for International Economics, April 2002).

[7]In certain circumstances, where countries have very open economies and the NPV/export ratio may not be appropriate (because large exports drive down the ratio), countries qualify if the NPV debt/revenue ratio is above 250 percent. For more information on HIPC, see www.worldbank.org/hipc.

including a stable macroeconomic environment, lower trade restrictions, policies to support private-sector growth, and strengthening of the financial and legal systems. Countries must develop **Poverty Reduction Strategy Papers**, detailing their plans for growth, development, and poverty reduction. Once countries establish this track record, they reach the **decision point** at which creditors provide interim relief—that is, they forgive debt payments as they fall due but do not yet forgive the stock of debt. Countries that maintain strong economic policies for an additional year or more and carry out specific reforms determined at the decision point reach the **completion point**, at which time they receive irrevocable debt reduction.

Under the program as it was conducted until 2009, countries that reach the completion point received debt reduction sufficient to reduce their NPV debt/export ratio to 150 percent, a level the IMF and World Bank deem to be the threshold for sustainable debt for low-income countries. Two features of this approach are noteworthy. First, a single threshold level is used for all countries. Basic economics would suggest that each country can sustain a different level of debt, but politically and institutionally, it was simpler for these organizations to treat all countries the same. Second, note that whereas the Paris Club forgives a percentage of debt, HIPC forgives the excess over the threshold. Thus a country with a NPV debt/export ratio of 200 percent would receive 25 percent debt forgiveness, whereas a country with a 300 percent ratio would receive a 50 percent write-off.

In 2005, the shareholders of the IMF, World Bank, and African Development Bank decided to go a step further and offer 100 percent forgiveness of all obligations owed to the three organizations for countries that reached the completion point. This step represented a major change and is the first time any of the organizations provided 100 percent forgiveness.

As of 2009, 35 HIPC had reached the post-completion point, and 12 countries were at the interim or predecision point (several mired in civil conflict were technically eligible but had made little progress). The 35 countries that reached their post-completion points had received debt relief amounting to $118 billion from the HIPC Initiative and a variety of other related bilateral and multilateral debt relief initiatives. The debt service/export ratio fell from an average of about 18 percent in 1999 to 6 percent in 2009, while debt service/revenues fell from 22 percent to 6 percent. In many countries, the benefits from the program were already evident. For example, Ghana used its savings to construct 509 new classroom blocks around the country, provide micro credit to about 43,000 farmers, and fund 560 sanitation and 141 water projects. Both Uganda (Box 13–3) and Tanzania used some of their saving to eliminate primary school fees, leading to sharp increases in school enrollment.

The HIPC Initiative demonstrates that, even after policy makers decided that some debt forgiveness is warranted, they face several key questions. These issues are more difficult in forgiving debts owed to government and multilateral institutions because market-based mechanisms are less relevant. Which countries should be eligible and which should not (that is, what are the right cutoffs for “poor” and “heavily indebted”)? How much debt should be forgiven and over what time frame? Should the debtors be expected to undertake policy conditions in return for debt relief, and

BOX 13-3 DEBT RELIEF IN UGANDA

When Uganda gained its independence from Britain in 1962, there was widespread optimism that its vibrant agricultural base and diverse, talented people could provide the basis for sustained economic development. However, political instability in the late 1960s and a 1971 coup led by Idi Amin ushered in a long period of political terror and economic destruction. By the time Yoweri Museveni assumed power in 1986, Uganda was one of the poorest countries in the world. Food production had fallen by a third in 15 years; average income had declined by over 40 percent; and life expectancy was just 48 years.

The new government embarked on a broad-based economic rebuilding program that focused on rehabilitating infrastructure, restoring macroeconomic balance, lowering barriers to trade, and investing in the social sectors. In the 15 years that followed, the economy staged a remarkable recovery. Average income increased by nearly 60 percent, and the share of the population living below the poverty line fell from 56 to 44 percent. By the mid-1990s, however, Uganda had accumulated a substantial debt burden. By 1993, total external debt was more than 12 times the value of annual exports, one of the highest such ratios in the world (total debt equal to twice the value of exports is generally considered a heavy burden). Two thirds of the debt was owed to multilateral agencies, so the traditional methods for debt relief (for private sector or bilateral debt) had limited potential.

In April 1998, Uganda was the first country to become eligible for debt reduction under the Heavily Indebted Poor Countries (HIPC) Initiative. In the first formulation of the program, Uganda received about $650 million in debt relief, reducing its $3.2 billion debt stock by about 20 percent. Two years later, Uganda became the first country to qualify for the enhanced HIPC program, which offered deeper debt relief. This second phase reduced Uganda's debts by about $1.3 billion, bringing the total debt reduction to approximately $2 billion. These steps reduced Uganda's debt service payments by about $80 million per year, or about two thirds of debt service payments due. In 2005, Uganda became eligible for 100 percent forgiveness of debts owed to the International Monetary Fund (IMF), World Bank, and African Development Bank. Uganda's early success with the program was encouraging: The country sharply increased the net primary school enrollment rate from 62 to 86 percent between 1992 and 2003, raised investments in farm-to-market roads by 75 percent, and brought the share of the urban population with access to safe water from 54 percent in 2000 to 65 percent in 2003 (however, the country also increased its military spending in response to regional conflicts). Gross domestic product (GDP) per capita grew at a rate just below 3.6 percent between 1995 and 2010, increasing average income by 70 percent over that period.

This information is drawn primarily from multiple World Bank websites on the HIPC program and the country page for Uganda (look for links at www.worldbank.org).

if so, what conditions (and determined by whom)? How should the cost burden be shared? There are no obvious right or wrong answers to these questions because they involve economic, financial, political, and institutional considerations, so these issues will continue to be major challenges for policy makers in the years ahead.

EMERGING MARKET FINANCIAL CRISES

International capital flows played a central role in a series of financial crises that struck several developing countries starting in the mid to late-1990s, including Argentina, Brazil, Ecuador, Indonesia, Korea, Mexico, Russia, Thailand, Turkey, Uruguay, and Venezuela. These crises had some similarities with the 1980s and HIPC debt crises, but some critical differences as well. In most cases they struck very suddenly and ferociously, with enormous economic and financial consequences in a matter of weeks and months. Moreover, several of the worst-hit countries previously had strong economic performance and had been favorites of the international capital markets. These crises led to dramatic falls in GDP and investment, major disruptions of trade and banking relationships, widespread unemployment, and increases in poverty. They also led to widespread rethinking of both the role of foreign capital flows (especially short-term flows) and the proper timing and sequencing of financial liberalization in the development process.

At the heart of these crises were huge, sudden reversals of international private capital flows. Economies that had been receiving relatively large amounts of private capital suddenly were faced with withdrawals of lines of credit, demands to repay debts, an exodus of portfolio capital, and offshore flight by domestic investors. Table 13–5 shows both the rapid buildup in lending and the depth and speed of the subsequent reversal in capital flows for five Asian crisis economies: Indonesia, Malaysia, the Phil-

TABLE 13–5 **Five Asian Economies:* Private External Financing before and after the Crises (billion US$)**

	1994	1995	1996	1997	1998
Net private flows	40.5	77.4	103.2	−1.1	−28.3
Equity investment	12.2	15.5	19.7	3.6	8.5
Private creditors	28.2	61.8	83.5	−4.7	−36.8
Commercial banks	24.0	49.5	65.3	−25.6	−35.0
Nonbank private creditors	4.2	12.4	18.2	21.0	−1.7

*South Korea, Indonesia, Malaysia, Thailand, and the Philippines.

Source: Institute of International Finance, "Capital Flows to Emerging Market Economies," (Washington, DC: Institute for International Finance, January 1999).

ippines, South Korea, and Thailand. Net private capital flows to these five countries more than doubled *in just two years,* from $40 billion in 1994 to $103 billion in 1996. Net commercial bank lending alone nearly tripled from $24 billion to $65 billion. But, in the last six months of 1997, the private capital inflow of $103 billion suddenly turned into an outflow of $1 billion. This net reversal of capital flows of $104 billion was equivalent to about 10 percent of the combined precrisis GDPs of these five countries. Mexico suffered a similar fate during its 1994 crisis. The reversal in capital flows amounted to $40 billion over two years, equivalent to about 9 percent of GDP. Other crisis countries followed a similar pattern. With withdrawals of those magnitudes, it is little wonder these countries were plunged into crisis.

As must be the case, the reversals in capital flow immediately led to dramatic changes in trade balances, saving-investment gaps, and overall economic activity. In the five Asian crisis countries, current account balances changed from deficits averaging 5 percent of GDP in 1996 to surpluses averaging 5 percent of GDP in 1998. Economic output fell sharply in each of the crisis countries in either the year of the crisis or the year after (depending mainly on how early or late in the year the crisis struck), as shown in Table 13–6. The plunges in GDP growth in Argentina, Indonesia, and Thailand were particularly large. The poor were especially hard hit. For example, urban day laborers trying to eke out a subsistence living by loading trucks or working on construction sites suddenly were thrown out of work. Poverty rates in Indonesia doubled by official estimates from 12 percent to around 22 percent, with some unofficial estimates suggesting even higher levels.

Almost as striking as the collapse in growth was the speed of the rebound, at least in some countries. Argentina, Korea, Malaysia, Mexico, and Turkey all recorded GDP

TABLE 13–6 Gross Domestic Product (GDP) Growth before and after the Crisis

	REAL ANNUAL GDP GROWTH (PERCENT)				
COUNTRY	YEAR OF CRISIS	YEAR PRECEDING CRISIS	YEAR OF CRISIS	YEAR AFTER CRISIS	2 YEARS AFTER CRISIS
Argentina	1995	5.8	−2.8	5.5	8.1
Argentina	2001	−0.8	−4.4	−10.9	7.0
Brazil	1998	3.3	0.1	0.8	4.4
Indonesia	1997	7.6	4.7	−13.1	0.8
Korea	1997	6.8	5.0	−6.7	10.9
Malaysia	1997	10.0	7.3	−7.4	6.1
Mexico	1995	4.4	−6.2	5.2	6.8
Philippines	1997	5.8	5.2	−0.6	3.4
Thailand	1997	5.9	−1.4	−10.5	4.4
Turkey	1994	8.0	−5.5	7.2	7.0
Turkey	2001	7.4	−7.5	7.8	4.8
Venezuela	1994	0.3	−2.3	4.0	−0.2

Source: World Bank, *World Development Indicators 2004* (Washington, DC: World Bank, 2004).

growth of 5 percent or more two years after their respective crises. As we shall see, the relatively rapid recovery is at least partly a reflection of the central role played by creditor panic in many of these crises.

How did the crises happen? The affected countries had several characteristics in common. First, they tended to be middle-income and upper-middle-income countries that had been growing quickly. Second, all had received large flows of private international capital, much of it with short-term maturity structures. Third, they recently had liberalized their financial systems and had recorded a very rapid—perhaps too rapid—expansion of bank lending and other financial services. Fourth, most of the crisis countries had exchange rates heavily controlled by their central banks, often strictly fixed (or pegged) to the U.S. dollar. Fifth, some countries, but not all, had large government budget deficits financed by a combination of borrowing from local and overseas banks and bondholders. These similarities suggest that government policies contributed to the crises, especially banking, financial, and exchange rate policies. However, the quick recovery in several countries, the fact that crises struck so many countries in such a short period of time, and the consistent pattern of rapid buildup and then withdrawal of private foreign capital all suggest that flaws in the operations of international capital markets played an important role as well.

DOMESTIC ECONOMIC WEAKNESSES

Each of the crisis countries had liberalized its financial systems in the late 1980s and early 1990s and had done it in ways that inadvertently left the financial systems fragile and overextended. Entry requirements were eased for banks and other financial institutions, allowing new private banks to open. Governments removed regulations that controlled interest rates and forced banks to allocate credit to particular firms and investment projects, so banks had more flexibility in their lending and interest rate decisions. At the same time, banks were given much greater freedom to raise funds through offshore borrowing, indeed in some countries government policies actively encouraged banks to borrow from foreign banks and relend to domestic companies. In Thailand, total foreign liabilities of banks and financial institutions rose from 5 percent of GDP in 1990 to 28 percent of GDP in 1995, mostly reflecting Thai banks borrowing from foreign commercial banks. The combination of these changes led to very rapid increases in domestic lending by banks, which grew by the equivalent of more than 50 percent of GDP in just seven years in South Korea, Malaysia, and Thailand.

Of course, financial liberalization can bring about many benefits to developing countries, including the mobilization of additional resources, reduced intermediation costs, and improved allocation of credit. The problem was not financial liberalization per se but how it was done, especially how rapidly it was done. The speed and magnitude of the expansion of financial activities outstripped the government's ability to establish strong legal and supervisory institutions to safeguard the system.

Central banks did not have supervisors with the skills and authority necessary to determine which banks were vulnerable and take steps to penalize or close poorly performing institutions. Bank regulations were weak, poorly enforced, or both. In some cases, supervisors were pressured (or bribed) to overlook violations by banks with politically influential owners. As a result, some banks were undercapitalized, nonperforming loans were at high levels, and many prudential regulations were broken with no penalty. Over time, loans tended to go to weaker investment projects, and the quality of banks' loan portfolios deteriorated. This left the banks (and the financial systems more broadly) in a vulnerable situation. In Thailand, for example, extensive lending was directed at real estate, construction, and property. When property prices began to fall in late 1996, banks exposed to these markets began to weaken considerably, making these banks' foreign creditors increasingly nervous.

Exchange rate policies added to the problems. Each of the crisis countries had either fixed or heavily managed exchange rate systems; none had fully flexible currencies. Although fixed exchange rates help keep import prices stable and provide a price anchor in a highly inflationary environment, they can create three kinds of problems.

First, they tend to encourage short-term capital inflows, which are especially vulnerable to rapid withdrawals. With fixed exchange rates, investors believe there is little chance they will lose money from a rapid change in the exchange rate. A foreign investor is more likely to buy a one-month bond denominated in Mexican pesos if the investor believes there is little risk that the exchange rate will change during the month. In countries with flexible exchange rates, foreign investors must take into account the risk that a relatively small exchange rate movement quickly could wipe out any gain they realize from higher interest rates.

Second, fixed exchange rates tend to become overvalued, which makes imports cheap and undercuts the profitability of exports. Partly because of their exchange rate policies, the crisis economies generally experienced growing imports, a slowdown in export growth, and a widening of the trade deficit in the years preceding the crisis.

Third, and more subtly, once capital withdrawals begin in the early stages of a crisis, fixed exchange rates tend to help accelerate the withdrawals. Once investors recognize that withdrawals are under way, they begin to speculate against the local currency, betting that the government will have to remove the fixed exchange rate.[8] This speculation adds to the loss of reserves and the pressure on the exchange rate.

[8]Such speculation takes place along the lines of the following, very simplified example. A foreign investor believes that the Philippine peso will have to be devalued. The investor takes out a short-term loan of 25 million pesos from a bank in the Philippines. The investor then converts the money into dollars using the current exchange rate of 25 pesos to the dollar, yielding $1 million. If the investor is right and the exchange rate moves, fewer dollars will be required to repay the loan. For example, if the exchange rate moves to 50 pesos to the dollar, the investor need convert only $500,000 dollars to get the 25 million pesos needed to repay the loan, allowing the investor to pocket a tidy profit of the remaining $500,000. This strategy is called *shorting the peso.*

When the central bank finally runs out of reserves, it has little choice but to allow the currency to float, which usually leads to a very large depreciation. The Thai baht jumped from 25 baht to the dollar in July 1997 to 54 baht to the dollar in January 1998, and the Korean won moved from about 900 won to over 1,900 won to the dollar in just a few months before appreciating back in early 1998. It is easy to see why someone holding assets denominated in either baht or won and expecting a devaluation would have wanted to get money out of these countries as quickly as possible.

SHORT-TERM CAPITAL FLOWS

While policy weaknesses undoubtedly created vulnerabilities in these economies, to fully explain the speed and ferocity of the crises we must turn to the operations of the international capital markets and the actions of the foreign creditors.

A key reason that so much capital was able to leave these countries so quickly was that so much of it had very short-term maturity structures. A large portion of the loans to the firms, banks, and governments in the crisis countries was scheduled to be repaid in just a few months or even weeks. These **short-term loans** (with full repayment due in one year or less) were attractive to both borrowers and lenders. For the borrowers, short-term loans generally carry lower interest rates; for the lender, short-term loans carry lower risk (and require less provisioning by the supervisory authorities)[9] because the lender is not exposed over long periods of time. As long as these economies continued to grow, creditors were happy to roll over the loans when they fell due (that is, make a new loan for the same amount to repay the old loan), allowing borrowers to continue their operations. So as long as things are going well, short-term loans *appear* to pose few problems. However, as soon as there is any trouble—or more precisely, as soon as creditors think there *may* be trouble—creditors quickly withdraw their lines of credit and demand immediate repayment of loans. This is precisely what happened in East Asia: When Thailand's economy began to noticeably weaken in late 1996 and early 1997, creditors began to close off their lines of credit and demand repayment, setting off a chain of events that led to financial panic and severe economic crisis.

Economies become vulnerable to a sudden withdrawal of international capital when the short-term foreign exchange *liabilities* of the economy grow in excess of short-term foreign exchange *assets.* In that situation, economies can become *illiquid*: Roughly speaking, there may not be enough dollars (or whatever relevant foreign currency) on hand to pay all the international debts falling due. Table 13–7 shows the size of one significant type of foreign liability for the crisis countries: short-term debts

[9]Central banks require commercial banks to set aside (or provision) a percentage of all new loans to ensure that the bank has some capital on hand in case loans fail. The amount of provisioning varies by the perceived risk of the loan and generally is smaller for short-term loans. In part because banks do not have to provision as much for these loans, short-term loans carry lower interest rates than long-term loans.

owed to foreign commercial banks by the government, state-owned companies, commercial banks, and private corporations in each economy. The table also shows data for the main liquid foreign exchange asset of an economy: the foreign exchange reserves held by the central bank. The last column shows the key point: In each of the crisis economies, short-term foreign debts exceeded or nearly exceeded the available foreign exchange reserves. In this situation, economies are *vulnerable* to a severe crisis because if all short-term loans are called in for repayment, not enough foreign exchange is available to repay every debt.

Bear in mind that the short-term bank loans shown in Table 13–7 are just one kind of short-term foreign exchange liability. Other kinds of foreign capital also can be withdrawn quickly, including portfolio equity (that is, stock purchases), foreign-exchange bank deposits, hedging instruments, and long-term loans with clauses that

TABLE 13–7 Short-Term Foreign Debt and Reserves (million US$)

COUNTRY	PERIOD	SHORT-TERM DEBT	RESERVES	SHORT-TERM DEBT/ RESERVES
Crisis countries				
Argentina	June 1995	21,509	10,844	1.98
Argentina	June 2001	40,916	21,077	1.94
Argentina	September 2001	37,792	20,555	1.84
Argentina	December 2001	32,320	14,553	2.22
Brazil	December 1998	41,038	42,580	0.96
Indonesia	June 1997	34,661	20,336	1.70
Korea	June 1997	70,612	34,070	2.07
Malaysia	June 1997	16,268	26,588	0.61
Mexico	December 1994	33,149	6,278	5.28
Philippines	June 1997	8,293	9,781	0.85
Russia	June 1998	34,650	11,161	3.10
Thailand	June 1997	45,567	31,361	1.45
Turkey	June 1994	8,821	4,279	2.06
Turkey	June 2000	26,825	24,742	1.08
Turkey	September 2000	27,845	24,255	1.15
Turkey	December 2000	28,360	22,488	1.26
Venezuela	June 1994	4,382	5,422	0.81
Noncrisis countries				
Chile	June 1997	7,615	17,017	0.45
Colombia	June 1997	6,698	9,940	0.67
Egypt	June 1997	4,166	18,779	0.22
India	June 1997	7,745	25,702	0.30
Jordan	June 1997	582	1,624	0.36
Peru	June 1997	5,368	10,665	0.50
Sri Lanka	June 1997	414	1,770	0.23
Taiwan	June 1997	21,966	90,025	0.24

Sources: Bank for International Settlements, *The Maturity, Sectoral, and Nationality Distribution of International Bank Lending* (Basel, Switzerland: various issues); International Monetary Fund, *International Financial Statistics* (Washington, DC: various years); Joint BIS-IMF-OECD-WB Statistics on External Debt (November 30, 2004) at www.jedh.org accessed in 2005.

allow accelerated repayment, but data on these forms of capital are not available. Moreover, the withdrawals generally are not limited to foreigners: Local citizens also begin to try to convert their assets out of domestic currency and into dollars (or yen), putting further pressure on the exchange rate. Note that longer-term loans and FDI generally cannot be reversed as quickly as short-term loans and therefore are less prone to rapid withdrawals.

CREDITOR PANIC

These financial crises were at least partially the result of what are known as rational panics by the creditors. Under certain circumstances, investors may have the incentive to quickly withdraw their money from an otherwise reasonably healthy economy, if they believe that other investors are about to do the same thing. The classic example from within one economy is a bank run, in which bank depositors suddenly withdraw their funds and deplete the capital of the bank. The particular conditions under which a rational panic can occur are described in Box 13–4.

BOX 13–4 MODEL OF SELF-FULFILLING CREDITOR PANICS

Self-fulfilling creditor panics are best understood by beginning with the critical distinction between illiquidity and insolvency. An *insolvent* borrower lacks the net worth to repay outstanding debts out of future earnings. An *illiquid* borrower lacks the ready cash to repay current debt servicing obligations, even though it has the net worth to repay the debts in the long term. A *liquidity crisis* occurs if a solvent but illiquid borrower is unable to borrow fresh funds from the capital markets to remain current on debt-servicing obligations. Because the borrower is solvent, capital markets could in principle provide new loans to repay existing debts with the expectation that both the old loans and the new loans will be fully serviced. The unwillingness or inability of the capital market to provide fresh loans to the illiquid borrower is the nub of the matter.

Why might markets fail this way? The primary reason is a problem of collective action. Suppose each individual creditor is too small to provide all of the loans needed by an illiquid debtor. A liquidity crisis results when creditors as a group would be willing to make a new loan, but no individual creditor is willing to make a loan *if the other creditors do not lend as well*. One possible market equilibrium is that no individual creditor is willing to make a loan to an illiquid borrower precisely because each creditor (rationally) expects that no other creditor is ready to make such a loan.

Consider a simple illustration. Suppose that a borrower owes debt D to a large number of existing creditors. The debt requires debt service of θD in period one, and debt service of $(1 + r)(1 - \theta)D$ in period two, where r is the rate of interest charged on the unpaid balance of the loan. The debtor owns an investment project that will pay off Q_2 in the second period. (Note that for the project to be profitable, $Q_2/(1 + r)$ must be greater than the present value of total debt service payments in both periods $\theta D + [(1 + r)(1 - \theta)D]/(1 + r)$, which must be equal to D.) The debtor lacks the cash flow to repay θD because the investment project pays off only in the second period. Moreover, if the debtor defaults, the loan repayment schedule is accelerated (that is, creditors demand immediate repayment). The investment project is then scrapped, with a salvage value of $Q_1 < D$. In that case, the partial repayment of the outstanding loan from the salvage value is shared among the existing creditors on a pro rata basis.

Typically, this solvent but illiquid borrower would borrow a fresh loan, L, in the first period, use it to repay θD, and then service $(1 - \theta)D + L$ in the second period. Thus with $L = \theta D$, the total repayment due in the second period is $(1 + r)\theta D + (1 + r)(1 - \theta)D = (1 + r)D$, which by assumption is less than Q_2. In this case, then, the project remains profitable.

Suppose, however, that each individual creditor can lend at most λ, where $\lambda < D$ (that is, λ is much smaller than D). This lending limit might result from prudential standards imposed on individual bank lenders, which limit their exposure to particular debtors. If only one lender is prepared to lend in the first period, the borrower will be forced into default because it will not be able to service its debts in the first period. The new creditor lending λ in the first period would then suffer an immediate loss on its loans (indeed, it might receive nothing if repayments are prioritized such that all of the preceding creditors have priority on repayment). Obviously, a first period loan will require at least n_1 new lenders, where $n_1 = \theta D/\lambda$.

There are clearly multiple rational equilibria in this situation. In the normal case, n_1 lenders routinely step forward, the existing debts are serviced, and the future debts are also serviced. The investment project is carried to fruition. In the case of a financial crisis, each individual creditor decides not to lend on the grounds that no other creditor is making loans. The debtor is pushed into default. The debt repayments are accelerated, and the investment project is scrapped with sharp economic losses because the salvage value Q_1 is less than $Q_2/(1 + r)$. Each individual creditor, of course, feels vindicated in its decision not to lend; after all, the debtor immediately goes into default.

Source: Excerpt from "The East Asian Financial Crisis: Diagnosis, Remedies, Prospects" by Steven Radelet, et al., *Brookings Papers on Economic Activity*, No. 1, 1998. Reprinted by permission of Brookings Institution Press.

In an international context, two conditions provide the foundation for such a panic. The first is a high level of short-term foreign liabilities relative to foreign assets. In this situation, each creditor begins to recognize that, if all creditors demanded repayment, not enough foreign exchange would be available to pay everyone. The second is that some event makes creditors believe that other creditors may begin to demand repayment. The event could be a military coup, a natural disaster, a sharp fall in export prices, a fall in property prices that weakens the domestic banking system, or an event in a neighboring country that makes creditors nervous about an entire region. Once creditors believe that others might pull out, the only rational action for each creditor is to immediately demand repayment ahead of everyone else, to avoid being the last in line and left unpaid if foreign exchange reserves are depleted. This is why the very rapid withdrawals are called *rational* panics. The irony, of course, is that the simultaneous demand for repayment by all the creditors ultimately depletes reserves and brings on the very crisis that all would rather avoid. In this sense, these crises are often referred to as *self-fulfilling*: The fact that more creditors believe that a crisis is possible in itself makes a crisis more likely to actually happen.

In Asia, once Thailand's economy began to slow in late 1996 and early 1997 and banks came under increasing pressure from falling property prices, foreign creditors that had lent to Thai banks began to get nervous and withdraw their loans. Other creditors came to believe that Thailand might run out of foreign exchange reserves, devalue the baht, or both, and began to withdraw their credits. These events ultimately depleted foreign exchange reserves and forced the large depreciation of the baht that began in July 1997. The weaknesses in Thailand made creditors more nervous about its neighbors, specifically Korea, Malaysia, Indonesia, and the Philippines. A similar chain of events (with differences in the specific details and triggering events) occurred in most of the other crisis countries.

Once a panic begins, it tends to perpetuate itself for a period of time, for several reasons. First, as foreign creditors demand repayment and the exchange rate begins to depreciate, local citizens try to convert their financial assets from local to foreign currency, putting additional pressure on reserves and the exchange rate. This was a particularly large problem in Argentina, Brazil, Indonesia, and Russia. Second, the deprecations wreak havoc on the balance sheets of banks and corporations that had borrowed in foreign currency. For example, as the Indonesian rupiah jumped from about 2,500 rupiah to the dollar in mid-1997 to over 10,000 rupiah to the dollar in early 1998, Indonesian corporations with dollar debts had to come up with *four times* more rupiah to make their payments. The crippling effect of the exchange rate movement was obvious to the creditors, so the more the exchange rate fell, the faster the foreign creditors tried to withdraw any remaining loans, putting even more pressure on the exchange rate. Third, as exchange rates fall in one country, creditors begin to believe something similar will happen in neighboring countries, so they start to withdraw their funds from other emerging markets. There was no hint of trouble in

Malaysia, Indonesia, and the Philippines until the Thai baht collapsed in July 1997. Within weeks contagion had struck, and creditors were withdrawing their money from almost every country in the region.

These emerging market financial crises were not new phenomena. Similar events have been recurring in slightly different forms since the development of banks and the emergence of international capital flows several centuries ago. Economists have long recognized that financial markets (both domestic and international) tend to be prone to instability and panic. Early in the twentieth century, industrialized country governments put into place mechanisms and institutions specifically designed to reduce the frequency and severity of financial crises, including lender-of-last-resort facilities at the central bank, deposit insurance, and bankruptcy procedures, making crises much less common within these countries.

However, these key institutions generally do not exist in international financial markets, leaving these markets vulnerable to panics. When central banks face a run on their foreign exchange reserves (akin to a commercial bank facing a depositor run), no international lender of last resort stands ready to supply it with the foreign exchange it needs to remain liquid (the IMF only partially fulfills this role, as discussed later). Similarly, no international insurance mechanism akin to deposit insurance assures creditors that they will be paid if a borrower defaults. And there is no international bankruptcy court that can call for a mandatory standstill on debt service payments and oversee the distribution of assets when a country cannot meet its short-term foreign currency obligations. On each count, there is extensive debate as to whether these institutions *should* exist and, if so, how they might realistically be designed to operate effectively in an international context, where there is no single legal authority. For example, an international agency that insures creditors could lead to excessive lending to weak companies because the creditor always could rely on an insurance settlement if the company goes bankrupt. In any event, in the absence of these institutions, international financial markets continue to be prone to rapid oscillations and financial panics.

STOPPING PANICS

Once a financial panic from international capital withdrawals is under way, it is very hard to stop. There are four basic options:

1. Governments can try to convince foreign creditors and citizens to stop withdrawing funds and even supply new funding. They can reduce the demand for foreign exchange by tightening fiscal and monetary policies and, perhaps, restricting imports. They can also implement policy reforms aimed at correcting perceived weaknesses in the economy. For example, if creditors are worried about the banking sector, reforms to strengthen banks might convince creditors to begin lending again.

2. Governments can try to increase the supply of foreign exchange by borrowing from official sources (such as the IMF and World Bank), much like a commercial bank borrows from its central bank in the event of a bank run. New funds can help assure creditors that enough foreign exchange is on hand to pay everyone, if need be, and allow the economy to operate normally.
3. Governments can try to stop the creditor rush for repayment by restructuring foreign debts so they will be repaid over longer periods of time. This option may include a formal debt rescheduling, a limited standstill period during which loan repayments are postponed, or in some cases outright default.
4. The government can do nothing (intentionally or unintentionally) and let the panic run its course until all creditors have fled, foreign exchange reserves are exhausted, debts are in default, and the economy is in deep recession. At some point, even in these dire circumstances, some investors will begin to return to take advantage of low asset prices, the foreign exchange market will begin to stabilize (at a greatly depreciated exchange rate), and the economy slowly will begin to recover.

The appropriate combination of policy actions and financing depends on the root causes of the crisis and the perceptions of international creditors, but there is little doubt that both policy adjustments and financing are necessary to avoid complete collapse. The problem, of course, is that correctly diagnosing a panic and prescribing the right response is very difficult when markets are changing by the hour, little accurate information is available, and the perceptions and reactions of creditors are impossible to measure. It therefore is not surprising that both the affected countries and the international community often make mistakes when dealing with incipient crises, like the ones faced by the emerging market economies.

In most of the affected economies, once the crisis was under way, the government turned to the IMF for advice and financial assistance. The IMF was established in 1945 in the aftermath of the Great Depression and World War II to help support stability in the international monetary system, including promoting the balanced expansion of world trade, the stability of exchange rates, and orderly corrections of balance-of-payments problems.[10] It aims to prevent crises or moderate their impact by encouraging countries to adopt what it sees as sound economic and financial polices and by providing temporary financing when necessary to address balance-of-payments problems. The IMF was undoubtedly in a difficult situation in working on these crises. The policies and programs it promoted during the crises were hotly debated, with analysts divided on whether the programs initially helped ease the

[10]See the description of the purposes of the IMF at www.imf.org/external/pubs/ft/exrp/what.htm.

crisis or unintentionally added to the panic, at least in its early stages, by shaking the confidence of investors and further weakening the economy.[11]

The IMF's initial set of policy prescriptions, consistent with option 1, centered on tightening both fiscal and monetary policies (to reduce aggregate demand and the demand for foreign exchange) and closing weak financial institutions. Each requires a delicate balance. In countries in which the fiscal deficits were large and part of the problem (because they required significant financing) such as Argentina and Turkey, some fiscal tightening was necessary. However, in other countries, particularly those in Asia, where the fiscal balance was either a small deficit or even a surplus, tightening fiscal policy was more debatable, and some analysts believe this step may have added to the economic contraction. Tightening monetary policy and raising interest rates is even more complex. On the one hand, high interest rates might attract some foreign currency and slow capital withdrawals. On the other hand, foreign creditors might believe that high interest rates would further weaken banks and corporations, leading them to accelerate their demands for repayment. The empirical evidence on this issue is far from conclusive. Perhaps the most controversial issue was reform of financial institutions. There is no doubt that banking systems were weak and overextended and needed significant reform. However, the very abrupt closure of some of these institutions may have added to the panic in the short run, leading to an acceleration of bank runs as depositors feared their bank would be the next closed.

Actions by governments in several of the crisis economies exacerbated rather than eased the crisis. Thailand's reluctance to float its currency earlier and take steps to address the problems facing banks that were heavily exposed to property markets made the crisis more severe than it otherwise would have been. The Indonesian central bank made huge loans to try to prop up weak banks, ultimately fueling capital flight and further weakening the currency.

Turning to the second option, these programs were supported by substantial financing from the IMF and, in some cases, bilateral funding from the governments of Japan, the United States, and European countries. This financing was designed to augment the foreign exchange reserves of the crisis economies, convince creditors that sufficient funds would be available to repay everyone, and ease pressure on the exchange rate. In several cases, the amount of financing pledged by the international community was extremely large (Table 13–8). International financial commitments to Mexico, Brazil, South Korea, and Turkey all exceeded $30 billion.

[11]For debates on these issues, see Steven Radelet and Jeffrey D. Sachs, "The East Asian Financial Crisis: Diagnosis, Remedies, Prospects," *Brookings Papers on Economic Activity* 1 (1998), 1–90; Jason Furman and Joseph Stiglitz, "Economic Crises: Evidence and Insights from East Asia," *Brookings Papers on Economic Activity* 2 (1998), 1–136; Martin Feldstein, "Refocussing the IMF," *Foreign Affairs* 77, no. 4 (March–April 1998), 20–33; Stanley Fischer, "In Defense of the IMF," *Foreign Affairs* 77, no. 4 (July–August 1998); Timothy Lane et al., "IMF-Supported Programs in Indonesia, Korea, and Thailand: A Preliminary Assessment," IMF Occasional Paper 178 (1999); World Bank, *East Asia: The Road to Recovery* (Washington, DC: World Bank, 1998); Martin Feldstein, ed., *Economic and Financial Crises in Emerging Market Economies* (Chicago: University of Chicago Press, 2003); Nouriel Roubini and Brad Setser, *Bailouts or Bail-Ins? Responding to Financial Crises in Emerging Economies* (Washington, DC: Institute for International Economics, 2004).

TABLE 13–8 International Financing for Selected Crisis Countries

	IMF COMMITMENTS		BILATERAL COMMITMENTS		TOTAL COMMITMENTS		ACTUAL DISBURSEMENTS	
COUNTRY	BILLION DOLLARS	PERCENT GDP	BILLION DOLLARS	PERCENT GDP	BILLION DOLLARS	PERCENT GDP	BILLION DOLLARS	PERCENT GDP
Mexico (1995)	18.9	4.6	20	5.0	38.9	9.6	27.6	6.8
Thailand (1997)	4.0	2.2	10.0	5.5	14.0	7.7	11.2	6.2
Indonesia (1997)	11.3	5.0	15.0	6.6	26.3	11.6	10.8	4.7
Korea (1997)	20.9	4.0	20.0	3.8	40.0	7.7	19.4	3.7
Russia (1998)	15.1	3.5	0	0.0	15.1	3.5	5.1	1.2
Brazil (1998–99)	18.4	2.3	14.5	1.8	32.9	4.1	17.5	2.2
Turkey (1999–2002)	33.8	17.0	0	0.0	33.8	17.0	23.1	11.6
Argentina (2000–01)	22.1	7.8	1.0	0.4	23.1	8.1	13.7	4.8
Uruguay (2002)	2.7	14.5	1.5	8.0	2.7	14.5	2.2	11.8
Brazil (2001–02)	35.1	6.9	0	0.0	35.1	6.9	30.1	5.9

GDP, gross domestic product.

Source: Nouriel Roubini and Brad Setser, *Bailouts or Bail-Ins? Responding to Financial Crises in Emerging Economies* (Washington, DC: Institute for International Economics, 2004).

The amounts actually disbursed were often smaller, either because the simple announcement of the amounts available helped ease creditor concerns, the crisis began to pass before the full amounts were necessary, or sometimes because the recipient countries did not fulfill specified conditions so funds were withheld. In Mexico, South Korea, Brazil, and Turkey, the funds clearly were sufficiently large to help ease the panic; in other cases, it is less clear the amounts were sufficient. However, some analysts believe that large financing packages are harmful in the long run: Some suggest that the bailout of Mexico in 1995 made private creditors more relaxed about lending aggressively to other emerging markets and so may have contributed to the buildup in capital flows that led to other crises in 1997 and 1998. This kind of situation is known as **moral hazard**, the risk that an agreement or contract will alter the behavior of interested parties in perverse ways.[12] In this case, the agreement to bail out Mexico might have led investors to believe they could take greater risks in other emerging markets, as they would get bailed out in the event of another crisis. However, many analysts dispute whether Mexico's bail out had this effect.

Option 3 is to restructure the debts to either stretch out the maturity dates so the debts can be paid later or to write off part or all of the debt. In the most extreme cases, such as Argentina, Ecuador, and Russia, governments adopted a unilateral **standstill**—that is, they halted all debt payments—or simply defaulted. In effect, these steps effectively forced the creditors to provide financing involuntarily, what economists Nouriel Roubini and Brad Setser refer to as "bailing in" the creditors.[13] The option for default is even more complicated when the creditors are not just foreign banks but local banks or citizens who hold government bonds. When governments have borrowed from local banks, defaulting on those debts further weakens the domestic banking system and can lead to a freeze on deposits, obviously a politically unpopular step.[14]

Steps less dramatic than outright default are possible and less disruptive. Brazil and Turkey persuaded at least some of the foreign banks to roll over existing lines of credit, but these were not mandatory and many banks did not do so. More-formal restructurings are also an option. Turkey exchanged about $8 billion of short-term local currency debt, held mostly by domestic banks and residents, into longer-term dollar and lira debt. Argentina extended the maturity of about $15 billion in government debt and capitalized interest on another $15 billion in long-term debt. These deals were expensive in terms of the interest rates charged, but they provided immediate financing and avoided the disruptions of default.

Perhaps the most interesting and important debt restructuring was in South Korea. In late December 1997, under pressure from the U.S. government and the IMF,

[12]The classic example of moral hazard is insurance markets, where coverage against a loss might increase risk-taking behavior by the insured.

[13]Roubini and Setser, *Bailouts or Bail-Ins?*

[14]Defaulting on debts owed to foreign banks similarly can weaken those banks, but they tend to be a much smaller share of the foreign banks' balance sheets, so the impact is more limited.

the major creditors to South Korea's banks agreed to reschedule about $22 billion in debt payments that were to fall due in the first quarter of 1998. This step was taken very early in the crisis, and its preventative effect in stopping the panic was immediate: The Korean won began to appreciate, and the Korean stock market rebounded the day after the rescheduling was announced. Within weeks, the most intense part of the Korean crisis effectively was over, and the Korean economy rebounded very quickly.

Debt restructurings have several advantages. First, they at least partially share the burden of adjustment between the creditors and debtors, rather than forcing the debtor to make all the adjustments. This is not only more fair, it may have a preventative aspect: With standstills or restructuring realistic possibilities, creditors should be less likely to engage in excessive lending. Second, when debts are restructured less new finance is required from the international community (and its taxpayers). Third, because pressure to immediately repay debts is at the heart of the creditor panic, restructuring can have an immediate salutatory effect, as was the case in Korea. Debt restructurings and standstills are the centerpiece of bankruptcy proceedings in most industrialized economies, but engineering them in cross-border situations is much more difficult. Several analysts have suggested different ways to establish an international bankruptcy regime, but putting these ideas into action has proven difficult.

Panics eventually end, even when little or no action is taken to stop them, simply because foreign exchange eventually is depleted, so creditors stop demanding repayment and debt falls into default. But the resulting economic contraction can be very deep, and the financial system can be left in a shambles. In Indonesia, for example, where the financial crisis quickly cascaded into a political crisis that led to widespread rioting and destruction of property and ultimately the resignation of President Suharto in May 1998. GDP contracted by about 15 percent in 1998 before the economy finally began to stabilize in 1999, and GDP per capita did not recover to the level of 1997 until 2004.

LESSONS FROM THE CRISES

There are several things we can learn from such economic crises. First, these crises are cautionary tales about rapid financial liberalization and the difficulties involved in building strong institutions in emerging markets. The crisis economies liberalized their financial systems very quickly, without fully establishing and strengthening the institutions necessary to oversee and regulate financial transactions. Building well-functioning financial systems remains a major challenge in the development process. These crises suggest that governments should proceed carefully in liberalizing domestic financial transactions and ensuring parallel development of the requisite regulatory institutions. The financial crisis in the United States and Europe in 2008-10 further strengthened awareness of the need for caution in financial liberalization when it became apparent that even the most developed financial systems were vulnerable partly as a result of insufficient regulatory oversight of the financial system.

Second, the crises reveal the vulnerabilities of relying on a fixed exchange rate, at least for countries with large private capital inflows. The conditions under which fixed or floating exchange rates are preferable for long-run growth and development is an open question that economists have debated for two centuries, and the debate is far from resolved. Fixed rates reduce volatility in thin foreign exchange markets and can provide some certainty to skittish investors. But the crises indicate that fixed rates create vulnerability to a panic and ultimately lead to huge economic adjustments when the exchange rate no longer can be defended. Increasingly, many economists suggest that the choice for developing countries is between a very rigidly fixed rate defended with abundant foreign exchange reserves or a freely floating exchange rate. Most economists now tend to believe that the rigid fixed rate options, including a currency board or outright dollarization,[15] are appropriate in only a very limited number of developing countries and flexible exchange rates generally are the preferred choice. However, the debate on this issue is far from over and likely to continue in the years to come.[16]

Third, in terms of foreign capital inflows, there are clear differences between FDI and other long-term capital and short-term capital. Long-term capital flows are much less prone to panic and more strongly associated with long-term investment and growth. Because short-term flows create vulnerability to a panic, governments should be much more careful about, and at times possibly even discourage, short-term flows, especially while they are in the process of strengthening weak financial institutions. Some analysts support restrictions on short-term capital inflows as a temporary step to protect nascent financial systems, so long as the restrictions are limited to short-term capital (there is little support among economists for restrictions on long-term capital or outflows of any kind). Chile has been at the forefront of countries trying to encourage long-term capital flows and discourage short-term flows. During the 1990s, Chile required that foreign investors deposit a share of their investment funds in a non-interest-bearing account for one year, a step that made short-term investment much less profitable. Such restrictions reduced short-term capital flows to Chile without reducing aggregate capital flows.[17]

[15]Recall that, with a currency board, the government issues domestic currency only when it is fully backed by available foreign exchange reserves at the given (fixed) rate, thereby keeping the ratio of foreign currency to domestic currency constant. With dollarization, a country adopts the dollar (or another widely traded currency) as its legal currency, thereby giving up an independent monetary policy.

[16]An excellent recent survey of exchange rate policies is Michael W. Klein and Jay C. Shambaugh, *Exchange Rate Regimes in the Modern Era* (Cambridge: MIT Press, 2010).

[17]For discussions of short-term capital flows, see Richard Cooper, "Should Capital Controls Be Banished?" *Brookings Papers on Economic Activity* 1 (1999), 89–141; Jaime Cardoso and Bernard Laurens, "The Effectiveness of Capital Controls on Inflows: Lessons from the Experience of Chile," IMF Monetary and Exchange Affairs Department, 1998; Sebastian Edwards, "Capital Flows, Real Exchange Rates, and Capital Controls: Some Latin American Experiences," Department of Economics, University of California at Los Angeles, 1998; Felipe B. Larraín, ed., *Capital Flows, Capital Controls, and Currency Crises: Latin America in the 1990s* (Ann Arbor: University of Michigan Press, 2000).

Fourth, vulnerability to crises can be reduced not only by reducing short-term capital inflows but also by building up foreign exchange reserves. China and Taiwan employed just this kind of self-insurance and were able to avoid serious difficulties while crises swirled around them in 1997 and 1998, even though China, in particular, displays some of the same financial sector weaknesses evident in the crisis countries. Although China had over $30 billion in short-term foreign debt in 1997, its foreign exchange reserves were almost $150 billion. After the crises, China continued to build up its reserves, which reached an astonishing $3 trillion in 2010. Taiwan, with about $20 billion in short-term debt in 1997, had over $90 billion in reserves, so also was safe from this kind of crisis (it subsequently built its reserves to over $300 billion by 2010). Korea, whose usable reserves had dwindled to about $3 billion when its crisis erupted (compared to $70 billion in short-term debts), subsequently built up its reserves to $300 billion by 2011, effectively ensuring that it will not soon face a similar international payments crisis and have to look to the IMF for emergency financing.

Fifth, the operations of international financial markets and the immediate reactions by the official international community probably added to the severity of the crisis. As a result, there has been widespread discussion about possible reforms to the international financial architecture. These debates have proceeded on several fronts, including reforming the operations of the IMF itself, strengthening international banking standards, and establishing new international mechanisms for debt standstills and rollovers, but few fundamental changes have been made.[18] As they have been for two centuries, financial crises are likely to continue to be recurring yet difficult-to-predict phenomena affecting emerging markets around the world.

SUMMARY

- Debt flows to developing countries have oscillated widely since the 1970s. They grew rapidly until the 1980s, then fell in the aftermath of the 1980s debt crises. They accelerated again in the early 1990s, only to drop sharply after the financial crises of the late 1990s then rebound again after 2001. Middle-income countries tend to borrow from both private lenders and official creditors (such as the World Bank), whereas low-income countries borrow mainly from official sources (including IDA loans from the World Bank).

[18]See Roubini and Setser, *Bailouts or Bail-Ins?*; Feldstein, *Economic and Financial Crises in Emerging Market Economies;* Barry Eichengreen, *Toward a New International Financial Architecture: A Practical Post-Asia Agenda* (Washington, DC: Institute for International Economics, February 1999); Morris Goldstein, *Safeguarding Prosperity in a Global Financial System, Report of an Independent Task Force for the Council on Foreign Relations* (Washington, DC: Institute for International Economics, September 1999).

- Debt can play an important role in financing investment and stabilizing an economy buffeted by shocks, but too much borrowing, especially of the wrong kind and for the wrong purposes, can create debt problems. In particular, borrowing for consumption or weak investment projects, borrowing too much when GDP growth or export growth rates remain low, or borrowing too much on short-term debt can leave countries vulnerable to crisis.
- The debt crises of the 1980s caused major economic dislocations in many countries around the world, especially in Latin America. For countries that had borrowed heavily from private commercial banks, the debt crises by and large were resolved through Brady bonds restructurings in the late 1980s and early 1990s. However, for lower-income countries that had borrowed from official sources, debt burdens continued to grow through the 1990s. The heavily indebted poor-country initiative has begun to reduce the debt burdens for some, but not all, the countries.
- The financial crises of the late 1990s struck emerging markets that had borrowed substantial funds with short maturities and had relatively low levels of foreign exchange reserves, fixed exchange rates, and weak financial systems. The rapid withdrawal of funds caused exchange rates to plummet and interest rates to soar and caused significant damage to domestic banks, private companies, government financial accounts, and overall economic well-being.
- International capital flows are more vulnerable to these kinds of crises than domestic capital because the institutions that can help reduce risks domestically do not operate well across borders, including lender-of-last-resort facilities, deposit insurance, and bankruptcy procedures. Countries have taken steps to minimize their risks by reducing (or restricting) short-term capital flows, building up reserves, introducing more flexible exchange rates, and strengthening financial systems.

14

Foreign Aid

> One of the greatest accomplishments of the past 50 years—the massive drop in the number of child deaths from 20 million children in 1960 to 8.1 million deaths last year—is an example of the tremendous progress we've made, in large part, thanks to foreign assistance. If the world comes together on a plan for financing development, the impact on health and development will be enormous. (philanthropist Bill Gates, April 2011).[1]

> I have long opposed foreign aid programs that have lined the pockets of corrupt dictators, while funding the salaries of a growing, bloated bureaucracy. (U.S. Senator Jesse Helms, January 11, 2001)[2]

The two viewpoints given in the quotations opening this chapter succinctly reveal the diversity of opinions and the contours of the debates about foreign aid. Foreign aid has always been controversial. As early as 1947 Congressman (later Senator) Everett Dirksen of Illinois labeled the **Marshall Plan**, a post–World War II aid program for European reconstruction now seen as one of the most highly regarded aid programs of all time, as "Operation Rat-Hole." Prominent

[1]Bill Gates, "Staying Committed to Development," *theGatesNotes,* April 11, 2011; www.thegatesnotes.com/Topics/Development/Staying-Committed-to-Development.

[2]Jesse Helms, "Towards a Compassionate Conservative Foreign Policy." Remarks delivered at the American Enterprise Institute, January 11, 2001, available at http://aei.org/speech/foreign-and-defense-policy/towards-a-compassionate-conservative-foreign-policy/.

economists such as Peter Bauer and Nobel laureate Milton Friedman have strongly criticized aid. Bauer believed that aid only enriches elites in recipient countries and famously quipped "aid is a process by which the poor in rich countries subsidize the rich in poor countries."[3] Friedman argued, beginning in the 1950s, that aid only strengthened and enlarged central governments, and as a result, aid did more harm than good. Critics, from both left and right, see aid as a political tool that distorts incentives, invites corruption, and entrenches corrupt dictators and elite business interests. Many believe that aid has little effect on growth and often has done more harm than good for the world's poor. They cite the widespread poverty in Africa and South Asia despite four decades of aid and point to countries that have received significant amounts of aid and have had disastrous growth records, including the Democratic Republic of the Congo (formerly Zaire), Haiti, Papua New Guinea, and Sudan. Critics call for aid programs to be dramatically reformed, substantially curtailed, or eliminated altogether.

By contrast, supporters see foreign aid as an important ingredient in fighting poverty, accelerating economic growth, and achieving other development objectives in low-income countries, especially in the very poorest countries, where people may not be able to generate the resources needed to finance investment or health and education programs. Equally prominent economists, such as Columbia University's Jeffrey Sachs and Nobel laureate Joseph Stiglitz, argue that, although aid has not always worked well, its overall record is positive, and it has been critical for poverty reduction and growth in many countries and helped prevent even worse performance in many others. Advocates argue that many of the weaknesses of aid have more to do with the donors than with the recipients, and because substantial amounts of aid are given for political purposes, it should not be surprising that it has not always been effective in fostering development. They point to a range of successful aid recipients such as Botswana, Korea, Indonesia, and Taiwan, and more recently, Ghana, Mozambique, and Uganda. They also note broader aid-financed initiatives such as the green revolution, the campaign against river blindness, and the introduction of oral rehydration therapy. They note that in the decades since aid became widespread in the 1960s, poverty indicators have fallen in many countries, and health and education indicators have risen faster than during any other 50-year period in human history.

Since its modern origins in the aftermath of World War II, foreign aid has become an important form of international capital flow from the richest to poorest countries

[3]Peter Bauer, *Dissent on Development* (Cambridge: Harvard University Press, 1972). For more recent critiques of aid, see William Easterly, *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics* (Cambridge: MIT Press, 2001), and *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good* (New York: Penguin Press, 2006); Dambisa Moyo, *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa* (New York: Farrar, Straus & Giroux, 2009).

as well as a way to provide technical expertise and donate commodities such as rice, wheat, and fuel. While it plays a much less significant role in middle-income countries, it remains as much a source of debate and controversy there as in other countries. Official development assistance from Organization for Economic Co-Operation and Development (OECD) nations to low- and middle-income countries totaled close to $130 billion in 2010. Aid given to richer countries and aid provided by China, the oil-rich nations, and others add to this total. More than 40 governments around the world provide aid, and 150 countries receive at least some aid inflow. For some countries, the amounts are trivial, amounting to half of a percent of gross national income (GNI) or less. In others aid flows are substantial, totaling 20 percent, 40 percent, or even more of GNI.

This chapter explores the motivations for and impacts of foreign aid. The empirical evidence on aid effectiveness is decidedly mixed, with some research showing little or no relationship between aid and development and others showing a positive impact. On balance, the evidence suggests that on average aid has had a modest positive impact on development outcomes, but with wide variation. Aid has supported growth and development in some countries and contributed to more broad-based improvements in certain areas, such as health and agricultural technology. But, in other countries, it has had little effect and did not spur growth; in some countries, it probably held back the process of development, particularly when donors have given it to political allies with corrupt or ineffective governments that showed little or no interest in economic development.

This mixed record has led to sharp debates. Where, when, and how should aid be provided? Which countries are most likely to use aid effectively? Who should have the major responsibility for designing and implementing aid programs? What kinds of conditions should donors impose on recipients? And how can donors ensure that aid is not wasted and gets to the people who need it most and can use it most effectively?

DONORS AND RECIPIENTS

WHAT IS FOREIGN AID?

Foreign aid consists of financial flows, technical assistance, and commodities given by the residents of one country to the residents of another country for development purposes, either as grants or as subsidized loans. Aid can be given or received by governments, charities, foundations, businesses, or individuals. Not all transfers from wealthy countries to poor countries are considered to be foreign aid (or the equivalent term, **foreign assistance**). It depends on who gives it, what it is given for, and the terms on which it is provided. A commercial loan from Citibank to build an electricity generator is not aid nor is a grant from the British government to purchase military

equipment. However, a grant from the British government to build an electricity generator counts as foreign aid.

An official source for definitions, data, and information on foreign aid is the Development Assistance Committee (DAC) of the OECD, an international organization with membership consisting of the governments of over 30 industrialized countries, including almost all the major donors. According to the DAC, to be counted as foreign aid the assistance must meet two criteria:

- It must be designed to promote economic development and welfare as its main objective (thus excluding aid for military or other nondevelopment purposes).
- It must be provided as either a grant or a subsidized loan.

The grants and subsidized loans that make up foreign aid are often referred to as **concessional assistance**, whereas loans that carry market or near-market terms (and therefore are not foreign aid) are categorized as **nonconcessional assistance**.[4] Distinguishing between subsidized and nonsubsidized loans requires a precise definition. According to the DAC, a loan counts as aid if it has a "grant element" of 25 percent or more, meaning that the present value of the loan (taking into account its interest rate and maturity structure) must be at least 25 percent below the present value of a comparable loan at market interest rates (usually assumed by the DAC, rather arbitrarily, to be 10 percent with no grace period). Thus the grant element is zero for a loan carrying a 10 percent interest rate, 100 percent for an outright grant, and something in-between for other loans.

Official development assistance (ODA) is the term used to describe aid provided by DAC donor governments (hence the term *official*) to low- and middle-income countries for development purposes. If a donor provides foreign aid for military purposes—as the United States does for many nations, including Egypt and Israel—this may be included in the U.S. foreign aid budget but is not included as ODA by the OECD. **Private voluntary assistance** is another and growing form of foreign assistance that also is not included in tallies of ODA because it is not official assistance. Private assistance includes grants from NGOs, religious groups, charities, foundations, and private companies. The Bill & Melinda Gates Foundation provides over $1 billion per year to support programs especially in health and education in low-income countries. World Vision, Save the Children, Care, and Oxfam International are NGOs that may be familiar to you. They collect billions of dollars each year from donations by individuals, mostly citizens of rich nations, and spend them on humanitarian and other forms of assistance in poor nations. Some estimates suggest

[4]Nonconcessional loans from donor agencies are counted as part of official development *finance* but not as official development *assistance*.

[5]Homi Kharas, "The New Reality of Aid," in L. Brainard, and D. Chollet, eds., *Global Development 2.0: Can Philanthropists, the Public, and the Poor Make Poverty History*? (Washington, DC: Brookings Institution, 2008).

that private voluntary assistance may amount to as much as one half to two thirds of ODA.[5] This is a dramatic increase from even a few decades ago.

WHO GIVES AID?

Although economic assistance from one country to another has occurred for centuries, today's foreign aid programs trace their origins to the 1940s and the establishment of the United Nations, the World Bank, the International Monetary Fund (IMF), and the Marshall Plan (Box 14–1). Historically, most aid has been given as **bilateral assistance** directly from one country to another. Some of the major bilateral aid agencies today include the U.S. Agency for International Development (USAID), the United Kingdom's Department for International Development (DFID), the Japan International Cooperation Agency (JICA), the Saudi Fund for Development, the Canadian International Development Agency (CIDA), and Swedish International Development Cooperation Agency (SIDA). Some governments have multiple bilateral aid agencies. The U.S. government has well over a dozen departments and agencies that provide bilateral aid, including USAID; the Millennium Challenge

BOX 14-1 THE MARSHALL PLAN

When World War II ended in 1945, world leaders hoped that Europe would not require much outside assistance and that the key economies (especially Britain and France) could quickly rebuild themselves. But by 1947, there had been little progress, and there was growing concern that discontent in Europe could encourage the spread of communism or fascism. U.S. Secretary of State George Marshall proposed a new approach to reconstruction at a commencement address at Harvard University on June 5, 1947, in which the United States would provide substantial amounts of funding, but only if the European nations, for the first time, could work together to draw up rational plans for reconstruction.

Despite some objections by isolationist members of the U.S. Congress, President Truman signed legislation funding the program, officially called the European Recovery Program, in April 1948. By June 1952 (when the program officially ended), the United States had provided $13.3 billion in assistance (equivalent to over $100 billion in today's dollars) to 16 countries in Europe. Almost 90 percent

of the funds were provided as grants. The largest recipients in dollar terms were the United Kingdom, France, Italy, Germany, and the Netherlands. For most of the major recipients, funding exceeded 1 percent of gross domestic product (GDP). From the perspective of the United States, these were large commitments representing 1.5 percent of its GDP. The Marshall Plan was seven times larger than current U.S. official development assistance, which in 2010 was 0.21 percent of U.S. GDP.

The Marshall Plan is generally regarded as having been a huge success in helping stimulate rapid growth and recovery in Europe (although some analysts believe that recovery was already under way and would have occurred without the funding). While the Marshall Plan is often held up as a model (Marshall won the Nobel Prize for Peace, in 1953), it differed in many ways from today's aid programs. Most important, it was aimed at countries that already had relatively high incomes, highly skilled workforces, and established financial and legal institutions, characteristics absent from most low-income countries. The Marshall Plan was designed to help relatively advanced countries rebuild infrastructure and rebound to their earlier levels of productive capacity, whereas today's aid programs are aimed at the much more difficult task of initiating growth and development in countries where it (by and large) has yet to occur. Nevertheless, the Marshall Plan and its perceived success provided the foundation for today's aid programs.

Corporation; the Departments of Agriculture, Defense, Health and Human Services, State, and Treasury; and the Peace Corps.

While most bilateral aid is provided to support recipient country governments, some is disbursed to churches, research organizations, universities, private schools and clinics, and nonprofit agencies. Even private companies sometimes receive foreign assistance, such as U.S. assistance through *enterprise funds* that make investments in firms (such as in Poland in the early 1990s) or aid to micro-finance agencies, like the Grameen Bank in Bangladesh, that provide loans to small-scale entrepreneurs. Sometimes aid is provided as cash, but often it is delivered as goods (such as food aid or medicine) or services (such as technical assistance), or in the form of debt relief. An example of the latter is the nearly $5 billion in debt relief Liberia received in 2010 from the World Bank, IMF, African Development Bank, and other creditors. This was a massive sum, equivalent to almost $1,200 for every Liberian. The debt that was taken off the books had been incurred by Liberia's previous governments in the 1970s and 1980s, but 15 years of civil war had destroyed the economy, making repayment by the current government impossible.

In terms of total dollars, the United States has consistently been the world's largest donor, except for a few years in the mid-1990s, when Japan provided the largest amount of aid. In 2009, the United States provided $29 billion in ODA, with France,

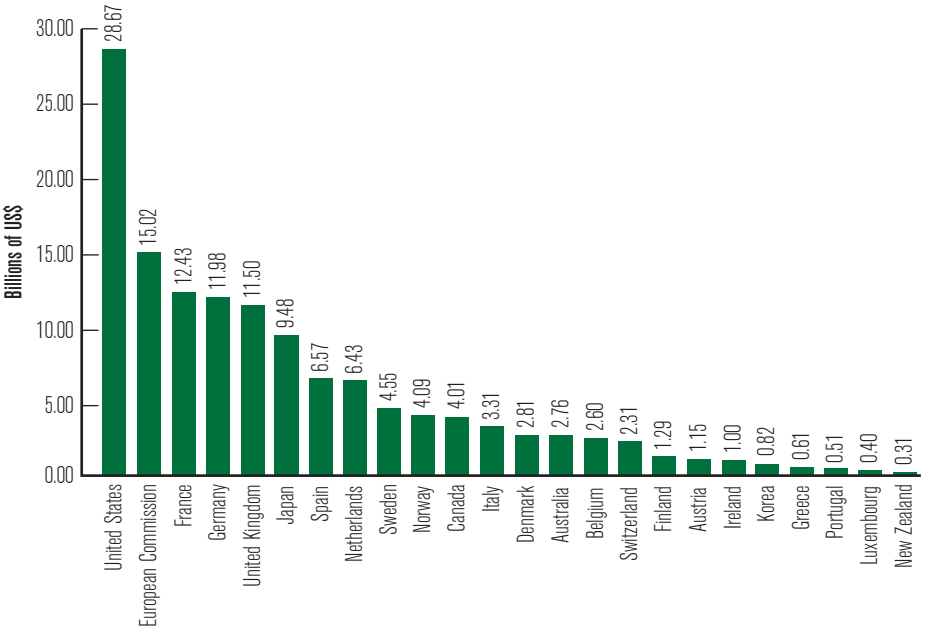


FIGURE 14–1 Net Official Development Assistance, 2009

Source: Organisation for Economic Co-Operation and Development Aid Statistics. Available at http://www.oecd.org/document/49/0,3746,en_2649_34447_46582641_1_1_1_1,00.html.

Germany, and the United Kingdom the next-largest donor countries (Figure 14–1). However, when aid is measured as a share of the donor's income, a different pattern emerges (Figure 14–2). The most generous donors from this perspective are Sweden, Norway, Luxembourg, Denmark, and the Netherlands, each of which provided between 0.82 and 1.12 percent of GNI in ODA in 2009.[6] The United States is one of the smallest donors by this measure, with ODA equivalent to about 0.2 percent of U.S. income. This figure is well below its 1970 level of 0.32 percent and U.S. contributions during the 1960s, when ODA averaged 0.51 percent of U.S. income. Foreign aid makes up less than 1 percent of the U.S. federal budget. This is much smaller than the public believes: Surveys show that Americans think that the United States spends upward of 20 percent of its budget on foreign aid.[7] Although U.S. ODA levels as a

[6]Over the past decade one of the largest donors in the world measured as a share of its income was Saudi Arabia, which provided more than 1 percent of its GNI in aid. Saudi Arabia had been a much smaller donor until it increased assistance to Afghanistan and several other countries in the aftermath of the September 11 terrorist attacks on the United States. It is uncertain whether this trend will continue.

[7]Since 2001, surveys of public opinion in the U.S. consistently find that Americans overestimate the share of government spending devoted to foreign aid. In 2010, World Public Opinion.org asked a random sample of over 800 American adults, "Just based on what you know, please tell me your hunch about what percentage of the federal budget goes to foreign aid." The median response was 25 percent. A follow-up question then asks, "What do you think would be an appropriate percentage of the federal budget to go to foreign aid, if any?" The median response was 10 percent. Survey available at www.worldpublicopinion.org/pipa/pdf/nov10/ForeignAid_Nov10_quaire.pdf; accessed February 2012. Foreign aid is not alone; the public generally overestimates the relative size of most government expenditures.

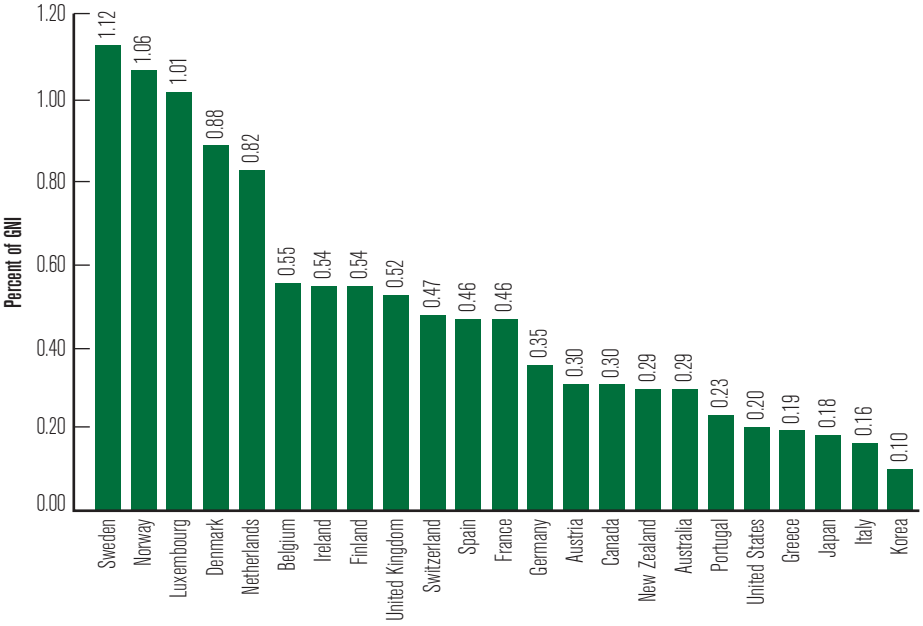


FIGURE 14–2 Net Official Development Assistance As a Percentage of Donor GNI, 2009

GNI, gross national income.

Source: Organisation for Economic Co-Operation and Development, Aid Statistics. Available at http://www.oecd.org/document/49/0,3746,en_2649_34447_46582641_1_1_1_1,00.html.

share of income are small compared to other countries, the United States and other rich countries affect poor countries in many more ways than just foreign aid: Policies on trade, technology, migration, security, and others areas are also important, as described in Box 14–2.

In nominal terms, global ODA increased steadily from the 1960s until it reached a peak of $60 billion in 1991, just after the end of the cold war (Figure 14–3). Aid flows then declined for a decade before rebounding in the early 2000s. They have risen steadily since due to an increased commitment by rich nations to assist poor ones and to the growth in development assistance provided to Iraq and Afghanistan. In real terms, increases in total ODA over the past decade indicate an even sharper rate of increase from earlier trends. Measured as a share of donor income, ODA fell throughout most of the 1980s and especially the 1990s. The 2000s reversed this trend, with ODA as a share of GNI now on a level last achieved during the 1970s. Given the economic problems currently facing many donor nations it is unclear if this trend will continue. Foreign assistance may start to decline in nominal terms and as a percentage of GNI as rich nations cut back on many areas of government spending.

The values in Figures 14–1 and 14–2 include both the amounts that donors give directly as bilateral aid and the amounts they provide indirectly as **multilateral aid**, which pools together resources from many donors. The major multilateral

BOX 14-2 THE COMMITMENT TO DEVELOPMENT INDEX

Rich countries affect poor countries in many ways, including through their policies on foreign aid, trade, migration, military spending, and the environment. Sometimes, these policies help poor countries; sometimes, they hurt. Open-trade policies can help provide markets for poor countries, but tariffs and quotas imposed by rich countries can keep poor countries out and slow their growth and development. To capture these broader channels beyond aid, the Center for Global Development produces an annual Commitment to Development Index that ranks 22 of the world's richest countries on the quality of their policies that affect poor countries in seven areas:

- Foreign aid, both in terms of its quantity (the amount of both government aid and private charitable contributions) and its quality, in terms of how it is given, to whom, and how much is required to be spent in the donor country.
- Trade policies, capturing tariffs, quotas, and subsidies for domestic farmers that impede trade or otherwise disadvantage poor countries trying to compete in rich-country markets.
- Foreign investment policies, including tax rates; programs that provide political risk insurance for investors against coups or other political instability; and other related policies.
- Migration, capturing the extent to which policies ease or hinder migration (particularly of unskilled rather than skilled workers) from poor to rich countries.
- Environmental policies, including greenhouse gas emissions, fishing subsidies, or other actions that lead to environmental degradation in poor countries.
- Security, including financial contributions to peacekeeping operations, naval operations that secure international shipping lanes, and penalties for certain arms exports.
- Technology, including policies that support research and development and protection of intellectual property rights, which enable creation and dissemination of innovations that help poor countries (like vaccines).

In 2010, Sweden came out on top, thanks to its large foreign aid program (relative to the size of its economy) and liberal immigration policies. Denmark ranked 2nd mostly because of its aid program. The United States ranked 11th, scoring relatively well on trade and security policies but poorly on aid programs.

South Korea, a newcomer to the Development Assistance Committee (DAC), ranked at the bottom, with weak scores on aid, environment, and migration and a strong score on technology. Japan, a more established donor, scored 2nd from the bottom with low scores in every category but technology. As with any composite ranking, the Commitment to Development Index is far from perfect, and different assumptions could change the rankings. Its intent, however, is not to be definitive in the rankings but to generate debate and discussion on the many different ways in which rich countries affect poor countries and on how the most important policies can be improved.

Source: David Roodman, a Senior Fellow at the Center for Global Development, is the architect of the index. The 2010 results are available at www.cgdev.org/section/initiatives/_active/cdi, accessed February 2012.

institutions include the World Bank; the IMF; the African, Asian, and Inter-American Development Banks; the United Nations; and the European Commission. The basic rationale for multilateral institutions is that they can provide larger amounts of aid with (presumably) lower bureaucratic costs (because donors do not have to duplicate efforts in each country) and fewer political ties (because funding decisions cannot be driven as easily by the political concerns of a single donor).

The largest multilateral agency is the **World Bank**, which began as the International Bank for Reconstruction and Development (IBRD) after its founding at a conference held in Bretton Woods, New Hampshire, in July 1944. The word *reconstruction* in the IBRD's title describes its first task, which was to help finance the reconstruction of Europe after World War II. The World Bank group today consists of five affiliated institutions operating in over 100 countries: the IBRD, the International Development Association (IDA), the International Finance Corporation (IFC), the Multilateral Investment Guarantee Agency (MIGA), and the International Centre for Settlement of Investment Disputes (ICSID). Most of the funding provided by the World Bank is *not* foreign aid. The IBRD lends to middle-income countries at market rates. It obtains its funds by borrowing on world capital markets then relends them to member countries at slightly higher rates. Because the IBRD has an excellent credit rating, it is able to borrow and relend to developing countries at cheaper rates than the recipients could obtain on their own in private markets. The IFC lends on commercial terms and takes minority equity positions in private companies. The MIGA provides guarantees to private foreign investors against loss caused by noncommercial (political) risks, and the ICSID helps settle investment disputes between foreign investors and host countries.

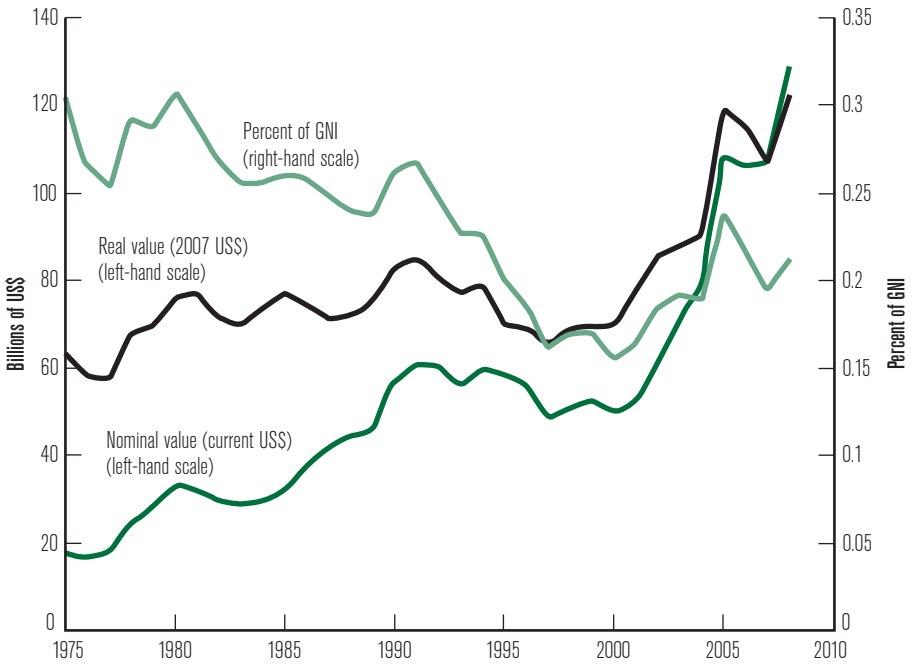


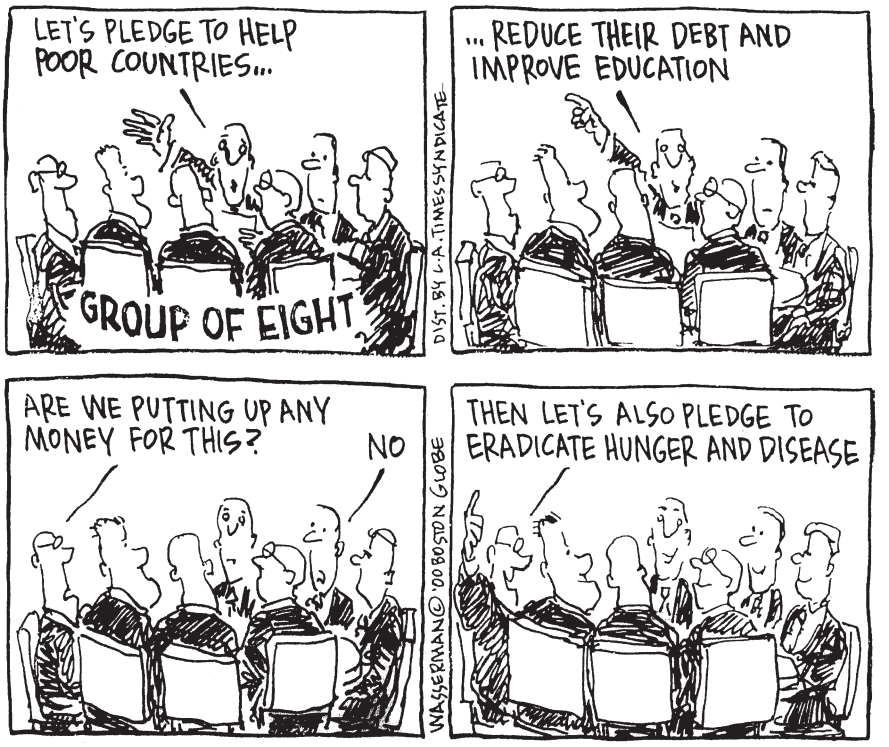
FIGURE 14–3 Global Official Development Assistance, 1975–2008

GNI, gross national income.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

The only part of the World Bank that actually provides foreign aid is the IDA. Donor governments contribute funds to the IDA, which uses it to provide highly concessional loans and grants to low-income countries. Funds from the IDA are used for a wide variety of development purposes, including building roads and ports, funding agricultural research and extension, purchasing medicines and schoolbooks, and training government officials. Its standard loan terms include a "service charge" of less than 1 percent per year and a repayment period of between 20 and 40 years, including a 10-year "grace" period during which no repayment is required. To be eligible for IDA funds today, countries must have (2011) GNI per capita of less than $1,165 (although the World Bank makes a few exceptions) and not have access to borrowing on private international capital markets. About 80 countries are eligible to borrow from the IDA. Among the largest borrowers, including countries that are no longer IDA eligible but that were in the past, are India, Vietnam, Tanzania, Ethiopia, and Nigeria. The IDA has disbursed about $222 billion since 1960; annual commitments today average close to $13 billion, 20 percent of which is in the form of outright grants.

The World Bank is owned and controlled by its 187 member governments, with each member having a voting share determined by the size of its shareholding, which



Source: "Group of Eight" cartoon by Dan Wasserman from *The Boston Globe*, January 1, 2000. Copyright © 2000 Boston Globe. All rights reserved. Used by permission and protected by the Copyright Laws of the United States. The printing, copying, redistribution, or retransmission of this Content is prohibited.

in turn is roughly commensurate with its portion of global gross domestic product. The U.S. is the largest shareholder in the IBRD (16 percent). The next-largest shareholders are Japan, Germany, the United Kingdom, and France. China and India each have about 2.7 percent of voting shares in the IBRD. There are on-going calls to reapportion voting rights away from high-income countries to low- and middle-income nations. Some progress has been made in this direction.

The **International Monetary Fund** was also founded at the 1944 Bretton Woods conference.[8] The IMF's original mission was to reestablish an international system of stable national currencies as part of an effort to rejuvenate world trade after World War II. All the IMF's early programs were in today's industrialized countries; frequent lending operations occurred with the United States, the United Kingdom, France, Germany, and other countries until the early 1970s. The IMF played a very

[8]The third organization that was meant to be launched at Bretton Woods was the International Trade Organization (ITO). However, the U.S. Congress failed to pass the legislation necessary to ensure U.S. participation, and the ITO never came into existence. Instead, the international community established the General Agreement on Tariffs and Trade, which was the forerunner of today's World Trade Organization (WTO).

limited role in developing countries until the late 1970s and early 1980s, when many low-income countries began to face severe balance-of-payments problems and debt crises.

Today most of the fund's operations are in low- or middle-income countries. IMF involvement in Greece and Ireland after their financial crises are important exceptions. The IMF's main purpose is to provide temporary financing to countries facing significant balance-of-payments problems stemming from a sharp fall in export prices, a rise in import prices (for example, from an increase in world oil prices), a financial crisis that leads to significant international capital flight, and other shocks. IMF funds are used to shore up the country's foreign exchange reserves and stabilize the currency, *not* to finance investment projects or consumption. The IMF is concerned primarily with helping countries achieve and maintain macroeconomic stability rather than directly supporting economic growth and development (although, of course, a stable environment is important for long-run growth). To receive IMF financing, countries must agree (sometimes controversially) to undertake policy reforms typically aimed at reducing the government budget deficit, tightening monetary policy, increasing the flexibility of the exchange rate, and quickly reducing the current account deficit to restore macroeconomic stability.

The vast majority of the IMF's financing is *not* foreign aid. For its emergency loans (called *stand-by credits*) it typically lends money for a one- to three-year period and charges market interest rates. Only about 10 percent of the IMF's outstanding loans are categorized as ODA, all of it provided to low-income countries. Although IMF programs are designed to be short term, in many instances they are renewed for many years. The IMF had continuous or nearly continuous programs beginning in the 1980s and continuing into the 2000s in Bolivia, Ghana, Malawi, the Philippines, Uganda, and several other countries.

Africa, Asia, and Latin America each have their own **regional development banks**: the African Development Bank, the Asian Development Bank, and the Inter-American Development Bank.[9] Like the World Bank, each of the major regional banks provides financing on both concessional and nonconcessional terms, depending on the income level of the borrower. Regional member countries and the major aid donors contribute to the capital of these banks, which also borrow on private capital markets to finance "hard" (nonconcessional) loans and receive contributions from aid donors for their "soft" (concessional) loan and grant operations. Other multilateral donors include the European Union (EU), the Organization of the Petroleum Exporting Countries (OPEC) Fund, and the Islamic Development Bank.

[9]A fourth, the European Bank for Reconstruction and Development, finances investment in eastern Europe and central Asia, but almost none of its finance is concessional.

The United Nations also has a range of foreign assistance programs that are of a considerably smaller scale than the lending programs of the World Bank or IMF. Much of this assistance comes as technical cooperation (consultants, advisers, and other expertise), primarily from the United Nations Development Programme (UNDP), the United Nations Population Fund (UNFPA), and the World Health Organization (WHO). Other United Nations agencies provide grants for development projects and humanitarian assistance, including the World Food Program (WFP), the United Nations Children's Fund (UNICEF, winner of the 1965 Nobel Prize for Peace), and the United Nations High Commissioner for Refugees (UNHCR).

WHO RECEIVES FOREIGN AID?

According to the DAC, over 150 countries and territories around the world are receiving aid. Table 14–1 shows the largest 10 recipients in 2009, each of which received more than $2 billion. Afghanistan topped the list (remember, ODA *does not* include military assistance). Total dollar amounts are important, but they do not tell the entire story. On a per capita basis, the aid flows to some of these countries are fairly small. Bangladesh received $1.2 billion in ODA, which sounds like a big number, but Bangladesh is a very populous country with over 160 million people. ODA inflows were equivalent to just a few percent of its GNI, or about $7.50 per Bangladeshi, the cost of a few kilograms of rice. But for some recipients with smaller economies, the amounts involved are quite large. Sierra Leone received $450 million in 2009 equal to 24 percent of its GNI that year. For its almost 6 million people, ODA worked out to be about $78 dollars per person, compared to a GNI per capita of $340 in the same year. For other (even smaller) countries a little bit of aid goes even farther. Tuvalu, a middle-income, island nation in the Pacific Ocean with a population of under 10,000 received $17 million in aid in 2009, which translated into 44 percent of GNI and over $1,700 per Tuvaluan. The difference in the apparent magnitudes of aid measured in nominal terms, as a share of GNI and aid per capita, reflects not only the behavior of the donors but the size of the population and income level of the recipient countries. A high aid to GNI ratio can indicate a large amount of aid, but it can also result from a low GNI or small population.

On a regional basis, sub-Saharan African countries received ODA flows averaging 4.9 percent of GNI in 2009, or $53 per person (Table 14–2). Despite the massive numbers of people living in absolute poverty in South Asia, exceeding the levels in sub-Saharan Africa, South Asia received less than 1 percent of its GNI in ODA, about $9 per person. For low-income countries around the world, donors in 2009 provided aid averaging about $47 per recipient, an increase from $14 in 2003. What happened over the those six years? Did the rich nations suddenly triple their commitment to the poorest nations? Not really. Since 2003, India graduated from low- to middle-income status, significantly raising the aid per capita amounts for low-income nations.

TABLE 14–1 Major Official Development Aid Recipients (2009)

RANK	COUNTRY	TOTAL ODA (MILLIONS OF CURRENT US$)
1	Afghanistan	6,235
2	Ethiopia	3,820
3	Vietnam	3,744
4	West Bank and Gaza	3,026
5	Tanzania	2,934
6	Iraq	2,791
7	Pakistan	2,781
8	India	2,502
9	Côte d'Ivoire	2,369
10	Congo, Dem. Rep.	2,354

RANK	COUNTRY	ODA (PERCENT OF GNI)
1	Liberia	70
2	Afghanistan	45
3	Solomon Islands	44
4	Tuvalu	43
5	Burundi	42
6	Micronesia, Fed. Sts.	41
7	Marshall Islands	32
8	Palau	28
9	Sierra Leone	24
10	Congo, Dem. Rep.	22

RANK	COUNTRY	ODA PER CAPITA (MILLIONS OF CURRENT US$)
1	Mayotte	2,751
2	Tuvalu	1,785
3	Palau	1,737
4	Marshall Islands	1,101
5	Micronesia, Fed. Sts.	1,093
6	West Bank and Gaza	748
7	Dominica	533
8	Grenada	463
9	Vanuatu	442
10	Kosovo	437

GNI, gross national income; ODA, official development assistance.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

Generally speaking, aid is an important type of capital flow for low-income countries, but not most middle-income countries. Aid flows averaged 9.3 percent of GNI in low-income countries in 2009 but just 0.1 percent of GNI in upper-middle-income countries. As discussed in Chapter 10, in the 1990s private capital flows grew sharply in middle-income countries, more than compensating for a fall in aid. In low-income countries, private capital rose much more slowly and remained significantly smaller than aid.

TABLE 14-2 Official Development Aid Receipts by Region and Income, 2009

	MILLIONS OF US$	PERCENT OF GNI	DOLLARS PER PERSON
Sub-Saharan Africa	44,553	4.9	53
South Asia	14,607	0.9	9
East Asia and Pacific	10,276	0.2	5
Europe and central Asia	8,101	0.3	20
Middle East and North Africa	13,589	1.1	42
Latin America and Caribbean	9,104	0.2	16
Low income	36,279	9.3	47
Lower-middle income	40,482	1.1	17
Upper-middle income	13,222	0.1	5
High income	433	0.0	0.4

GNI, gross national income.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

THE MOTIVATIONS FOR AID

What determines to whom donors provide aid and how much they give? Donors have a variety of motivations in providing aid, only some of which are directly related to economic development.

FOREIGN POLICY OBJECTIVES AND POLITICAL ALLIANCES There is little question that foreign policy and political relationships are the key determinants of aid flows. During the cold war, both the United States and the Soviet Union used aid to vie for the support of developing countries around the world. The United States provided significant amounts of aid to countries fighting communist insurgencies with little regard as to whether the aid was used for development, including Vietnam in the 1960s; Indonesia, the Philippines, and Zaire in the 1970s and 1980s; and several countries in Central America in the 1980s. The Soviet Union countered with aid to North Korea, Cuba, and countries across Eastern Europe. Both sides vied for support in newly independent countries across Africa and used aid to gain support for crucial votes in the United Nations or other world bodies. The United States and the Soviet Union are not the only countries to have used aid in this kind of competitive way: France, Germany, the United Kingdom, Taiwan, and China have used aid (among other policy tools) to try to gain support and recognition for their governments from countries around the world (Box 14–3).

Many donors provide significant aid to their former colonies as a means of retaining some political influence. Being a former colony substantially increases the probability of receiving aid flows. Between 1970 and 1994, 99.6 percent of Portugal's aid went to former colonies, whereas 78 percent of aid from the United Kingdom and

BOX 14-3 CHINA'S FOREIGN AID

China has had a small foreign aid program since the founding of the People's Republic of China in 1949. During the following three decades, China was a low-income country, with annual foreign exchange earnings of less than $10 billion a year, barely enough for China to pay for essential imports. China became a member of the World Bank in 1980 and became eligible for International Development Association (IDA) loans, with their heavily subsidized interest rates, which were available only to low-income countries. During these early decades, Chinese foreign aid involved small loans and grants mainly to friendly neighbors (North Korea for example) and to countries willing to switch formal diplomatic recognition from the Republic of China on Taiwan to the People's Republic with its capital in Beijing.

This situation began to change in the late 1980s and 1990s and changed dramatically in the first decade of the twenty-first century. China does not report the specific dollar amounts of its total aid to developing countries nor does it give a breakdown of that aid by recipient country. It is also not easy to distinguish between what is development aid in accordance with the official definition of development assistance as used by international organizations, by which the grant proportion of any assistance must be at least 25 percent of the total. Still there is no doubt that China's foreign aid had risen to billions of dollars a year by 2010. Some of this aid was in the form of outright grants, but much of it took the form of heavily subsidized loans from the Export-Import Bank of China. The official position of the Chinese government is that developing countries know best how foreign aid should be used and therefore the government never attaches conditions to its assistance, unlike the World Bank and many bilateral aid agencies that typically require the recipient country to institute particular reforms in order to receive aid.

While officially the purpose of Chinese aid is to provide unconditional funds to low-income countries to help them raise their incomes, the reality is that much of the aid is directed to countries that are important to China's own development. China's rapid industrialization over the past three decades has turned China into one of the world's largest importers of natural resource products, ranging from oil to iron and copper ore and much else. Some of the largest recipients of Chinese foreign aid have been countries in Africa that are exporters of oil or have given China the right in their country to explore for oil (Angola, Equatorial Guinea, Ethiopia, and Sudan among others). Many of these countries have repressive governments, and China often has been criticized for aiding governments that carry out repressive measures against their own people, including in the Darfur region of Sudan.

China's goal is to obtain the rights to develop and import natural resource products from whoever is willing to sell, and ownership and drilling rights are seen as a way of securing reliable sources of these products. Aid in turn helps acquire those rights. The aid finances large infrastructure construction projects, ranging from roads to government buildings and hospitals, often built by Chinese construction companies and with Chinese labor battalions. Many countries use their aid programs to support their own economic and geopolitical interests abroad so China is not unique in this respect.

Sources: State Council of China, "White Paper: China's Foreign Aid" *China Daily*, April 22, 2011; Thomas Lum, Hannah Fischer, Julissa Gomez-Granger, and Anne Leland, "China's Foreign Aid Activities Africa, Latin America, and Southeast Asia," Congressional Research Service, Washington, DC, February 25, 2009; Robert Rotberg, ed., *China into Africa: Trade, Aid and Influence* (Washington, DC: Brookings Institution Press, 2008).

between 50 and 60 percent of aid from Australia, Belgium, and France went to their former colonies.[10]

For the United States, the most important geopolitical concern outside of the cold war has been the Middle East. Since 1980, two of the largest recipients of U.S. foreign aid (which goes beyond ODA) have been Israel and Egypt because the United States provided financial support to back the Camp David peace agreement signed by those two countries in November 1979. More recently, the largest U.S. aid recipients are countries important to key national security interests, such as Afghanistan, Pakistan, and Jordan. Beginning in 2002, Iraq became the largest recipient of U.S. aid (and the largest aid recipient in the world), and its reconstruction is likely to become the largest single foreign aid program ever recorded. Aid commitments to Iraq began to decline in 2009 while increasing amounts of development assistance shifted to Afghanistan.

INCOME LEVELS AND POVERTY Income and poverty are important considerations in aid allocation, at least for some donors, although not as much as is sometimes assumed.[11] For many people in rich countries, the main rationale for aid is to help those in most need in the poorest countries. Income levels and poverty influence both the amount of aid donors provide and the extent of its concessionality. Donors generally provide their most concessional aid to the poorest countries and give fewer

[10]Alberto Alesina and David Dollar, "Who Gives Foreign Aid to Whom and Why?" *Journal of Economic Growth* (March 2000), 33–63.

[11]Alesina and Dollar, "Who Gives Foreign Aid to Whom and Why?," found that income levels were the primary motivation for aid from Denmark, Finland, Norway, and Sweden.

grants and subsidized loans for higher-income countries. Some aid programs are designed explicitly with this objective in mind. The World Bank's IDA program has an income ceiling (as do the concessional windows of the regional development banks). Once countries reach that ceiling, in most cases, they "graduate" from IDA to nonconcessional IBRD loans. Other programs have less formal graduation rules but still tend to provide less aid as incomes grow.

As incomes of the poorest countries rise, the composition of their capital inflows tends to change, with aid flows declining and private capital inflows increasing. Botswana is an example. It received aid flows equivalent to 15 percent of GNI in the early 1970s, when its income averaged less than $800 per capita, but with its average income now over $6,000, ODA amounts to only a few percent of GNI. But Botswana grew rapidly. The transition from high to low aid receipts generally takes many years because most nations do not grow as quickly as Botswana did. One study calculated that, for the most successful countries, the half-life of aid is about 12 years, meaning that aid falls to about 50 percent of its peak level after 12 years and about 25 percent of its peak after 24 years.[12] So, for many of the poorest countries, a rapid transition to private capital flows is unlikely.

COUNTRY SIZE Donors provide much more aid (either as a share of GDP or per person) to smaller countries than to large countries. If donors were concerned strictly with allocating aid to where the largest numbers of poor people live, much more aid would go to China, India, Indonesia, Bangladesh, and Pakistan. However, the very size of some of these countries sometimes daunts donors. They prefer to provide aid to smaller countries, where the difference it makes is more noticeable. Aid of $50 million would be huge in The Gambia, but would not be noticed in India. For political reasons, donors generally want to influence as many countries as possible, which tends to lead to a disproportionate amount of aid going to small countries. A vote in the UN General Assembly counts the same if it is from a small country or a large one, so donors may try to influence as many small countries as possible.

COMMERCIAL TIES Bilateral aid is often designed at least partially to help support the economic interests of certain firms or sectors in the donor country. Multilateral aid is less prone to these pressures, although by no means immune. Many analysts have concluded that commercial ties are an important determinant of Japan's aid. Food donations help support farmers that produce the crops in the United States and European Union. Many donors tie portions of their aid to purchases within the donor

[12]Michael Clemens and Steven Radelet, "The Millennium Challenge Account: How Much Is Too Much, How Long Is Long Enough?" Working Paper No. 23, Center for Global Development, Washington, DC, February 2003.

country. Automobiles, airline tickets, and consulting services financed by U.S. foreign aid, in most cases, must be purchased from U.S. firms. Tying aid has a long history: Much of the machinery and equipment for the Marshall Plan was purchased from U.S. companies, and all of it had to be shipped across the Atlantic on U.S. merchant vessels. Tying aid probably helps strengthen political support for aid programs within donor countries, but it adds to the costs of aid programs and makes them less effective. One study found that tying aid added 15 to 20 percent to its cost, meaning that recipients received less of a benefit from the aid notionally allocated to their country.

DEMOCRACY Historically, donors have provided foreign assistance with little regard to whether recipient country governments were authoritarian or democratic. This was particularly the case during the cold war, but since the breakup of the Soviet Union, donors have tended to increase their aid to countries that have become democracies. More aid has been aimed at strengthening fragile democracies, supporting the transition of nondemocracies to democracies, or building democratic institutions, including financing for parliaments, election monitoring systems, and groups supporting civil rights and free speech. Alberto Alesina and David Dollar, in a study covering the years 1970 through 1994, found that the typical country received a 50 percent increase in aid after switching to become a democracy.

AID, GROWTH, AND DEVELOPMENT

Many people think of foreign aid primarily as either building infrastructure (such as roads or dams) or providing emergency humanitarian relief after natural disasters, but its purposes often are more diverse. Most foreign aid is designed to meet one or more of four broad economic and development objectives:

- To stimulate economic growth through building infrastructure, supporting productive sectors such as agriculture, or bringing new ideas and technologies.
- To promote other development objectives, such as strengthening education, health, environmental, or political systems.
- To support subsistence consumption of food and other commodities, especially in emergency situations after natural disasters or humanitarian crises.
- To help stabilize an economy after economic shocks.

The extent to which aid has been successful in helping achieve these objectives is a matter of continued debate and controversy.

Despite these broader objectives for aid, economic growth has always been the main yardstick used to judge aid's effectiveness, with more aid expected to lead to

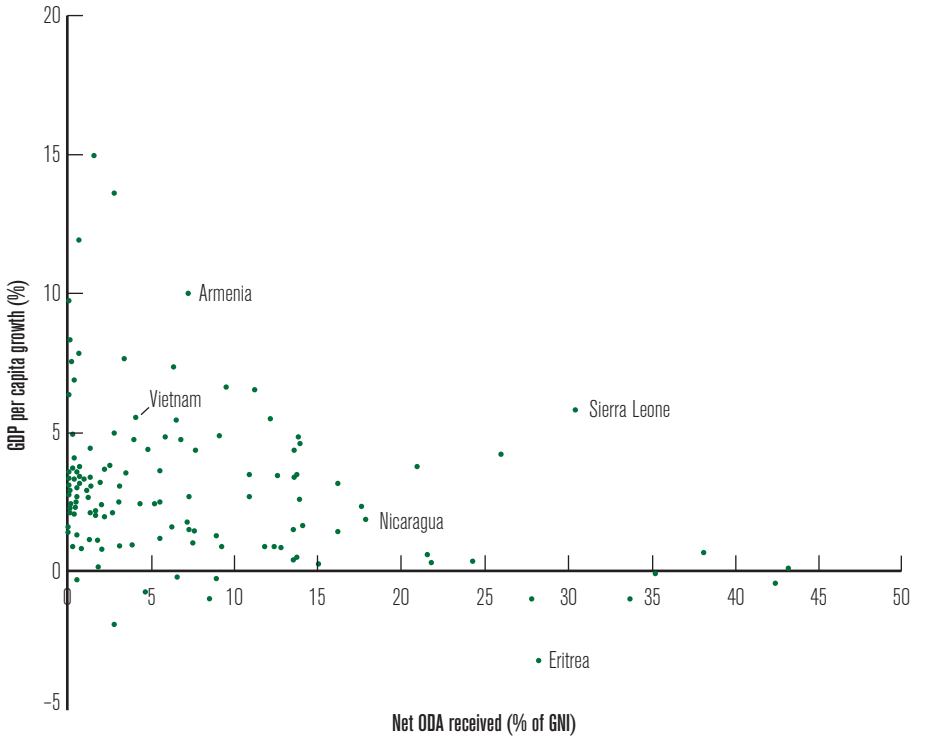


FIGURE 14–4 Foreign Aid and Growth, 1999–2008

GNI, gross national income; ODA, official development assistance.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

faster growth. But, at a very broad level, there is no apparent simple relationship between aid and growth, as shown in Figure 14–4. Some countries that received large amounts of aid recorded rapid growth (Sierra Leone), whereas others recorded slow or even negative growth (Eritrea). At the same time, some countries that received very little aid did very well, whereas others did not (note the scatter of points in Figure 14–4 close to the *y*-axis.)

What does the absence of a simple relationship mean? For some analysts, it is evidence of a failure of aid to achieve its basic objectives. But for others, this simple correlation is misleading because other factors affect both aid and growth. Some countries that received large amounts of aid may face endemic disease or poor geography or may be emerging from long-standing civil conflict, in which case aid might have a positive impact on growth even if the overall growth performance remains weak. Or the causality could run in the opposite direction: Donors give more aid to countries with slow growth rates and much less to rapid growers like China. These analysts suggest that, once these other factors are taken into consideration, a positive relationship emerges. Still others conclude that aid works well under certain circumstances but fails in others. Aid might help spur growth in countries with

reasonably good economic policies but fail to do so where corruption is rife and the economy is badly mismanaged. In this view, while the overall trend line is important, the variance around the trend and the reasons for that variance are also critical in understanding the true underlying relationships.

Debate on these issues has been ongoing for many years, and continues today. There is general agreement on some broad issues. Even aid pessimists (at least most of them) agree that aid has been successful in some countries (such as in Botswana or Indonesia or more recently in Mozambique and Tanzania), that aid has helped improve health by supplying essential medicines, and that aid is an important vehicle in providing emergency relief after natural disasters. Similarly, aid optimists concede that much aid has been wasted or stolen, such as by the Marcos regime in the Philippines and the Duvalier regime in Haiti, and that, even under the best circumstances, aid can have adverse incentives on economic activity. Debate continues on the overall general trends, the conditions under which aid works or does not work, and on what steps can be taken to make aid more effective. Empirical evidence is mixed, with different studies reaching different conclusions depending on the time frame, countries involved, and assumptions underlying the research. We summarize these by looking at three views on aid and growth.

VIEW 1. ALTHOUGH NOT ALWAYS SUCCESSFUL, ON AVERAGE, AID HAS A POSITIVE IMPACT ON ECONOMIC GROWTH AND DEVELOPMENT

The clearest economic rationale for aid is to promote growth by financing new investment, especially investment in public goods. Aid used to build roads, ports, electricity generators, schools, and other infrastructure augments the process of capital accumulation, which (if the investments are productive) should accelerate the rate of growth. This motivation for aid has been invoked since the Marshall Plan. Indeed, early analyses of foreign aid, such as Walt Rostow's popular 1960 book *The Stages of Economic Growth,* saw aid wholly in terms of financing investment and adding to the capital stock.[13] This idea is fully consistent with the Harrod-Domar model, which views capital formation as the main driver of growth, as well as with the Solow model, which sees capital accumulation as important for growth, albeit with diminishing returns and with a strong role for new technology. In the context of these models, aid adds to the total amount of saving, which increases investment and the capital stock, which in turn accelerates the rate of economic growth.[14]

[13]W. W. Rostow, *The Stages of Economic Growth: A Non-Communist Manifesto* (Cambridge: Cambridge University Press, 1960).

[14]As discussed in Chapter 4, in the Harrod-Domar model the effect is a permanent increase in the growth rate; in the Solow model it is a transitory increase in the growth rate until the economy achieves a new steady level of output.

In this view, poor countries are unable to generate sufficient amounts of saving on their own to finance the investment necessary to initiate growth, or if they do they can finance only very slow growth. As discussed in Chapter 10, saving rates tend to be low in the very poorest countries, and even where they are moderately high, the actual amount of saving is low. A country with per capita income of $200 and a saving rate of 10 percent generates saving of $20 per person per year, which cannot purchase much in terms of capital goods. In the strongest version of this view, the poorest countries may be stuck in a **poverty trap**, in which their income is too low to generate the saving necessary to initiate the process of sustained growth.[15] Total saving may be too small to compensate for depreciation, let alone add to the capital stock. In a more moderate version, the poorest countries may be able to save enough to begin to grow, but only at very slow rates. Thus aid flows provide a way to augment domestic saving and accelerate the growth process.

Aid can also support growth by building knowledge and transferring new ideas, technology, and best practices from one country to another. Some aid helps support research in low-income countries that can accelerate the pace of growth. One of the best examples is aid-funded research in the 1960s and 1970s on new varieties of seeds, fertilizers, and pesticides that helped transform agriculture in Asia through what became known as the *green revolution*. Similarly, a significant portion of aid finances advisers and technical assistance that ultimately is aimed at strengthening institutions and increasing productivity. This kind of aid, when it is effective, can be thought of as shifting the production function upward, thereby helping increase the rate of growth.[16]

The majority of studies, although far from all, find a positive relationship between aid and growth, on average, after controlling for the impact of other factors on growth.[17] In particular, beginning in the mid-1990s researchers began to investigate whether aid might spur growth, albeit with diminishing returns—that is, small amounts of aid might have a relatively large impact on growth, but each additional dollar of aid might have less effect. This may seem like an obvious point because it is the standard assumption about capital accumulation in most growth models (such as the Solow model), but it is surprising that earlier research tested only a linear relationship. Many (but not all) of the studies that allow for diminishing returns and control for other variables find a positive relationship, on average, between total aid

[15]See Jeffrey Sachs et al., "Ending Africa's Poverty Trap," *Brookings Papers on Economic Activity 1* (2004), 117–240.

[16]In the Solow model, the increase in the rate of growth lasts only during the transition to a new steady-state level of income.

[17]For a recent review of the research, see Michael Clemens, Steven Radelet, Rikhil Bhavnani, and Samuel Bazzi, "Counting Chickens When They Hatch: Timing and the Effects of Aid on Growth," *Economic Journal*. Published electronically December 1, 2011. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0297.2011.02482.x/full.

and growth.[18] In terms of Figure 14–4, these studies find that other variables, such as geography, political conflict, policies, and institutions explain much of the variance in growth rates among aid recipients. They conclude that, after controlling for these variables and allowing for diminishing returns, a positive relationship between aid and growth emerges, albeit with important variance around the trend line. This view is captured by Figure 14–5a. But as we shall see, other research reached different conclusions, so the debate about the relationship between aid and growth remains open.

In addition, for defenders of View 1, aid can have a positive impact on other important development objectives that may affect growth only indirectly or affect it only after a long period of time, such as health, education, or the environment. Similarly, it can provide emergency assistance or humanitarian relief or be used to help achieve macroeconomic stability. We briefly review each of these in turn.

PROMOTING HEALTH, EDUCATION, AND THE ENVIRONMENT Nearly half of all aid flows are aimed at improving health, education, the environment, and other objectives. Aid that finances immunizations, medical supplies, medicines, bed nets (to prevent the spread of malaria), clean water, and school supplies is meant to help provide essential goods and services to people too poor to afford them. In economic terms, most of these kinds of expenditures count as additions to consumption rather than investment (although some aid for education and health, such as building schools and clinics, count as investment).

Economic growth is often a secondary, long-term objective of this kind of aid. Healthier, better-educated workers should be more productive, shifting the production function and expanding output and income. A similar argument can be made for at least some aid aimed at protecting the environment. Improving forest or fisheries management, cleaning water supplies, or reducing air pollution can improve welfare and increase productivity over time. The impact on growth, however, can take a long time to materialize. Programs to immunize children or improve the quality of primary schools do not affect labor productivity until these children join the workforce, which is many years after the aid was received.

Research on the relationship between aid and other development outcomes tends to concentrate on evaluating specific projects and interventions rather than broad cross-country trends. As with growth, in many cases, aid appears to have had

[18]Cross-country econometric studies that allow for diminishing returns typically do so by including both aid and aid squared as determinants of growth. This specification yields a parabola in which aid has a positive effect on growth with diminishing returns, but in theory, with enough aid, the marginal impact could be negative. However, in these studies, the data points are concentrated on the upward portion of the curve, so they do not conclude that large amounts of aid have a negative impact. See, for example, Carl-Johan Dalgaard, Henrik Hansen, and Finn Tarp, "On the Empirics of Foreign Aid and Growth," *Economic Journal* 114, no. 496 (2004), 191–216; Robert Lensink and Howard White, "Are There Negative Returns to Aid?" *Journal of Development Studies* 37, no. 6 (2001), 42–65; Henrik Hansen, and Finn Tarp, "Aid and Growth Regressions," *Journal of Development Economics* 64 (2001), 547–70.

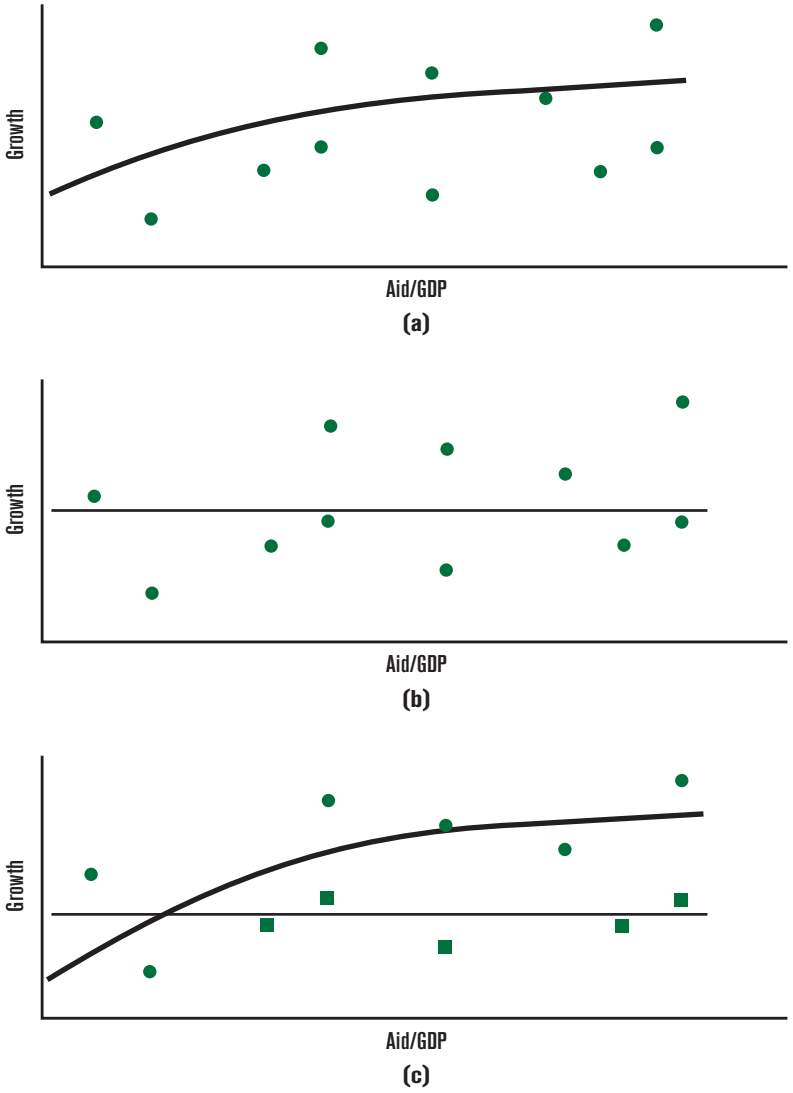


FIGURE 14–5 Three Views on Aid and Growth

(a) View 1: Aid has a positive impact on growth with diminishing returns after controlling for the impact of other variables. (b) View 2: Aid has little or no impact on growth and may have a negative impact. (c) View 3: Aid has a positive impact on growth in some circumstances (circles) but no impact in others (squares).

GDP, gross domestic product.

a strong impact on health and education, along with many interventions that failed. Perhaps the best documented are aid-financed interventions in health. A group of experts convened by the Center for Global Development documented 17 cases of large health interventions in developing countries that have been very successful in improving health outcomes and saving lives, several of which are discussed in

Chapter 9 and in Box 14–4.[19] A massive global effort spearheaded by the World Health Organization succeeded in completely eradicating smallpox from the world in 1977, perhaps the greatest public health achievement ever. A worldwide effort to fight polio, supported by both government aid agencies and private donors such as Rotary International, succeeded in nearly eliminating the disease. The introduction of oral rehyrdration therapy has saved millions of lives after being developed first in East Pakistan (now Bangladesh) in the 1970s. In Egypt, a donor-funded oral rehydration program helped reduce infant diarrheal deaths by an astonishing 82 percent between 1982 and 1987.

BOX 14-4 CONTROLLING RIVER BLINDNESS IN SUB-SAHARAN AFRICA

Onchocerciasis, or river blindness, is a pernicious, parasitic disease that afflicts approximately 18 million people worldwide, with well over 99 percent of its victims in sub-Saharan Africa. Spread through the bite of an infected blackfly living near fast-moving waters, river blindness causes symptoms such as disabling itching, rashes, muscle pain, skin lesions, and in its most severe cases, blindness.

In 1974, governments and donors in 11 West African countries jointly launched the Onchocerciasis Control Programme. At the time, 2.5 million of the area's 30 million inhabitants were infected with the disease, and approximately 100,000 were blind. The primary intervention was weekly aerial spraying of larvicide during rainy seasons along the region's waterways to control the disease-spreading blackflies. The aerial spraying, augmented by hand spraying of breeding grounds, persisted even through civil and regional conflicts and coups. In 1987, the program began distributing the drug ivermectin, which treats the disease's symptoms with a single annual dose.

The program was implemented through unusual cooperation between several governments and a wide range of donors. It was launched under the leadership of the World Health Organization, the World Bank, the Food and Agriculture Organization, and the United Nations Development Programme. A total of 22 donor countries contributed $560 million to support the 28-year project, with much of the financing provided through long-term commitments. The pharmaceutical company Merck & Co. donated ivermectin to the program beginning in 1987, pledging to provide it to anyone that needed it as long as they needed it, making the donation program one of the largest public—private partnerships ever

[19]Ruth Levine with Molly Kinder, *Millions Saved: Proven Success in Global Health* (Washington, DC: Center for Global Development, 2004).

created. Governments and nongovernmental organizations in the recipient countries were critical to implementing the program effectively.

The program achieved impressive success between 1974 and its conclusion in 2002: Transmission was halted completely in all 11 West African countries involved, 600,000 cases of blindness were prevented, and 18 million children born in the program area are now free from the risk of the disease. In addition to the striking health benefits, the economic impact has been impressive. An estimated 25 million hectares of arable land, enough to feed 17 million people, are now safe for resettlement. In Burkina Faso, 15 percent of the country's land that had been deserted because of onchocerciasis has been completely reclaimed, and its new residents now enjoy a thriving agricultural economy.

The program was extremely cost-effective, with a yearly cost of less than $1 per protected person. The World Bank calculated that the annual return on investment (attributable mainly to increased agricultural output) to be 20 percent, and it is estimated that $3.7 billion will be generated from improved labor and agricultural productivity.

Source: Ruth Levine with Molly Kinder, *Millions Saved: Proven Success in Global Health* (Washington, DC: Center for Global Development, 2004), chap. 6.

There have also been several successful education initiatives.[20] A girls' primary school education initiative in Balochistanm, Pakistan, led to a tripling of the number of schools, a more than doubling of girls' enrollment, and an increase in the girls' primary school completion rate from 7 to 30 percent between 1990 and 1998. In Ethiopia, a systemwide set of education reforms implemented in the 1990s with the support of USAID, the World Bank, and UNICEF led to the construction of many new schools, increasing the number of teachers and improving their training, and expanding the availability of textbooks. Between 1991 and 2001, the overall primary school enrollment ratio increased from 20 to 57 percent, and the girls' enrollment ratio jumped from 12 to 47 percent. In Indonesia, donors partially supported the construction of over 61,000 primary schools in the mid-1970s, which Massachusetts Institute of Technology (MIT) economist Esther Duflo showed led to significant increases in educational attainment and increased wages for school graduates.[21]

[20]For a summary of successful education interventions, see Maria Beatriz Orlando, "Success Stories in Policy Interventions towards High Quality Universal Primary Education" (Washington, DC: Center for Global Development, 2004).

[21]Esther Duflo, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment" *American Economic Review* 91 (September 2001).

These individual success stories only tell part of the story. There also have been many failures, where donors provided funding for health and education reforms that did not happen, for technical assistance that did not help, or for interventions that led to little change in health or education. But beyond case studies, little systematic evidence has been gathered on the circumstances under which aid-financed interventions in health and education have been successful and where they have not.

PROVIDING EMERGENCY ASSISTANCE AND HUMANITARIAN RELIEF About one tenth of all foreign aid provides emergency relief after natural disasters or humanitarian crises. Relief programs after earthquakes, floods, or other disasters typically provide food, clothing, basic medicines and drugs, and other items to help meet subsistence needs. Similarly, humanitarian aid assists people living in refugee camps and those displaced by war or other conflict. From an economic point of view, most of this aid is meant to support basic consumption rather than investment and growth, although some of it is used to rebuild ruined infrastructure and help output from collapsing further than it otherwise might following disasters.

SUPPORTING ECONOMIC AND POLITICAL STABILITY Some aid is provided to help countries facing macroeconomic instability. A fall in export prices, a rise in import prices, or an economic shock, such as a drought, flood, or earthquake, can lead to a sharp depreciation of the exchange rate, which in turn can ripple through the economy and increase the prices of traded goods and services. Aid inflows add to the supply of foreign exchange and can help moderate the pressure on the exchange rate. The IMF plays the leading role in this area, with other donors such as the World Bank providing supporting funds. Aid flows that are a response to a macroeconomic crisis or a natural disaster, as described in the previous paragraph, are inversely correlated with economic growth. Growth falls because of the crisis, and aid increases as a result. These situations account for some of the "below-the-line" cases in Figure 14–4, where aid is associated with very low or even negative GDP growth. But because aid is responding to the low growth, these cases should not be seen as examples of a failure of aid to promote growth and development.

VIEW 2. AID HAS LITTLE OR NO EFFECT ON GROWTH AND ACTUALLY MAY UNDERMINE GROWTH

How could aid have no impact or a negative impact on growth, as suggested by Figure 14–5b? One simple way is if most of it simply is wasted. If donors build large bureaucracies or spend the money on expensive technical experts from their home country that write reports no one reads, aid will not help growth. Aid that winds up in the personal off-shore bank accounts of government officials or finances a fleet of expensive cars for members of parliament creates little stimulus to growth. If aid

builds a road but provides no funds for maintenance, there may be an initial burst of output, but this could be followed by a decline back to previous levels as the road deteriorates. Because donors typically like to finance capital costs of new projects but not maintenance, this is a common problem.

More insidious, if aid breeds corruption, government officials and their cronies spend their time plotting about how to siphon aid money to their own bank accounts rather than increasing output. Aid funds used to pay private contractors for construction projects or delivery of services can create opportunities for kickbacks and other illegal activities, just like any other public funds. One of the strongest critiques of aid and how it might undermine development is that it can prop up malevolent dictators and support political regimes that further impoverish rather than help the poor. U.S. aid to the Marcos regime in the Philippines during the 1970s and 1980s, for example, may have helped support an anticommunist ally but probably also lengthened the time in which his corrupt regime remained in power. The same argument can be made about aid in the 1970s and 1980s to the Central African Republic, Haiti, or Zaire.

But even if aid is not simply wasted or misused, it could have a smaller than expected impact on growth because of diminishing returns to investment as the recipient begins to reach its **absorptive capacity**. Government bureaucracies with relatively few skilled workers might have difficulty overseeing and managing larger and larger aid flows. Purchases of commodities and supplies might strain available warehouses or delivery systems. Large donations of medicines might sit unused because they cannot be delivered on time to clinics. In the long run, an increase in aid flows may help expand absorptive capacity because it can be used to train and hire more workers, expand warehouse facilities, or build infrastructure to ease bottlenecks, but in the meantime, the impact of aid may be diminished by these constraints.

Perhaps the most important way in which aid could slow growth is by undermining incentives for private sector activity. Large aid flows can spur inflation and cause a real appreciation of the exchange rate, which reduces the profitability of production of all tradable goods. Aid flows can enlarge the size of the government and related services supporting aid projects, drawing workers and investment away from other productive activities such as agro-processing, garments, or footwear exports, which have been important engines for growth in many countries. These "Dutch disease" effects are usually associated with revenues from natural resource export booms (Chapter 18), but can result from aid flows as well. Some analysts believe that large aid flows to Ghana in the late 1980s and early 1990s undermined export incentives.[22] Similarly, food aid can sometimes undermine local food production, as discussed in Box 14–5.

[22]See Stephen Younger, "Aid and the Dutch Disease: Macroeconomic Management When Everybody Loves You," *World Development* 20, no. 11 (November 1992), 1587–97.

BOX 14-5 FOOD AID AND FOOD PRODUCTION

For many people, it seems obvious that donating food is a good way for rich countries to help poor countries. Although sometimes that is true, under certain circumstances, donations of food can hurt local farmers by undermining the incentives for them to produce food. If all food is produced locally (that is, there are no imports), then an increase in food from aid donations can shift out the supply curve for food and drive down food prices, benefiting consumers but hurting farmers, as shown in Figure 14–6a. In an economy with no commercial imports, food aid shifts the supply curve for food from S_1 to S_2, causing the price to fall from P_0 to P_1. Producers react by reducing output from Q_0 to Q_1, whereas consumers increase consumption from Q_0 to Q_2, with food aid of $Q_2 - Q_1$ filling the gap. Consumers are clearly better off (important if there is a true food shortage) while producers are worse off.

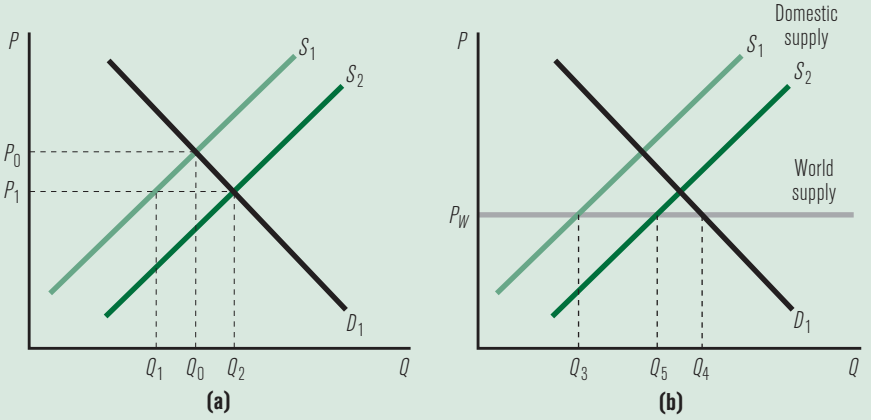


FIGURE 14-6 Food Aid, Prices, and Production

(a) Food aid and production in the absence of imports: Food aid lowers prices, benefiting consumers but displacing local production. (b) Food aid and production with imports: Food aid adds to the total supply without affecting prices or displacing production.

In an economy that produces some food but imports on the margin, the outcome is different, as shown in Figure 14-6b. With a world price of P_w, initial production takes place at Q_3 with consumption at Q_4 and imports of $Q_4 - Q_3$. In this situation, food aid displaces imports, with imports falling to $Q_4 - Q_5$ and food aid making up the difference $Q_5 - Q_3$. Domestic prices, food production, and consumption do not change, with food aid, in effect, saving the economy the cost of the displaced imported food.

AID, SAVING, AND TAX REVENUE Aid could also have a smaller-than-expected effect on growth if it has an adverse impact on saving and investment. Early analysts assumed that, because aid is a form of foreign saving, each dollar of aid would add one full dollar to investment. But this is unlikely to happen, for at least two reasons. First, as mentioned previously, not all aid is provided as investment goods or even aimed at increasing investment and growth. Second, even where aid is aimed directly at investment, the impact could be partially offset by a reduction in either private saving (through a decline in the rate of return on private investment) or government saving (through a fall in tax revenues). If a country that currently invests $100 million per year receives $10 million in new aid for investment (for example, building roads), it is highly unlikely that total investment will reach $110 million. It is far more likely that, with $10 million in new investment, some of the original $100 million will be shifted into consumption.

The impact of aid on saving is illustrated in Figure 14–7. Before the recipient country obtains aid, it can produce consumption goods and capital goods along the production possibilities frontier *P*. To simplify, the diagram ignores international trade and other capital inflows, so that consumption initially also must take place somewhere along the production frontier. Community tastes are defined by a set of indifference curves, of which two are shown (labeled *I* and *II*). The country's welfare is maximized if it produces and consumes at point *A*, where the indifference curve *I* is tangent to the frontier *P*, with consumption at C_1 and investment (and saving) at I_1.

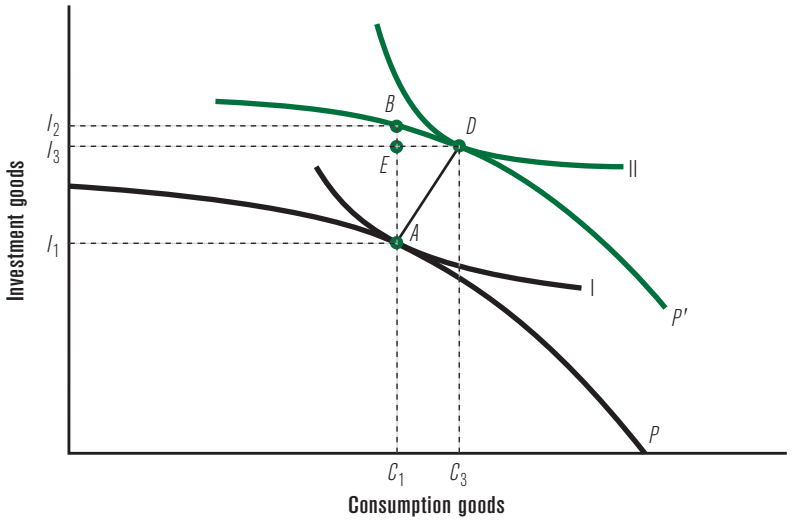


FIGURE 14–7 Impact of Aid on Investment and Consumption

Foreign aid totaling *AB* turns into actual new investment of only *AE* because the country maximizes welfare on the new frontier, *P′*, at point *D*, not *B*. *P′* is not really a production frontier but a supply frontier.

Source: Adapted from Paul Mosley, "Aid, Savings and Growth Revisited," *Oxford Bulletin of Economics and Statistics* 42 (May 1980), 79–91.

Now donor countries contribute an amount *AB* of aid with the intention that it be used to raise investment to I_2. However, the aid moves the *consumption frontier* outward from *P* to *P′*. Now, the country maximizes its welfare at point *D*, the tangency between *P′* and indifference curve I_2, where it consumes C_3 and invests I_3. Of the aid amount *AB*, *AE* (equal to $I_3 - I_1$) is invested and "BE" (equal to $I_2 - I_3$) is consumed. Even if all the new aid money is invested as intended, some of what the country had invested before is shifted to consumption, so in effect, the aid finances an increase in both investment and consumption, with the precise impact depending on production possibilities, community tastes, and other variables left out of the figure, such as trade. Aid may finance mostly investment if $I_3 - I_1$ is a large percentage of total aid (*AB*), or it may mostly finance consumption if $I_3 - I_1$ is a small percentage of *AB*.

Why would the country react in this way and not invest all the aid it receives? Saving always requires a sacrifice of current consumption, based on the anticipation that investment will lead to increased output and greater consumption in the future. When the country receives a new inflow of aid, investment increases and the country is less willing to sacrifice that last dollar of today's consumption to finance the last dollar of investment. Thus domestic saving falls somewhat, although with the addition of foreign saving from aid total saving (and investment) rises.

When substitutions of these kinds are possible and aid effectively finances activities for which it was not originally intended, aid is said to be **fungible**. This kind of substitution could be good or bad, depending how the money is spent. If donor funds to build a school allow members of the community to save less and purchase more food, clothing, and medicine for their families, few donors would take issue. But if instead the freed-up funds are used to buy several Mercedes Benzes for the mayor and his or her staff, donors are likely to be unhappy and economic development is not served.

Aid given as a cash transfer to the budget or as program aid rather than for a specific project deliberately gives the recipient the kinds of choices demonstrated in Figure 14–7. But even if all aid is used for very specific investment projects and the donor audits projects to confirm that the money was spent as planned, substitution is possible. Project aid might be used, for example, for investment projects that the government would have made without aid. In that case, the government's resources are freed up for other purposes. If the government is intending to build an electricity generator with its own funds, then a donor decides to provide the funding, the government's money is freed up for any purpose the government wishes. Even if aid finances projects the government was not planning to implement, the government might simply cut back on other projects because it wants to raise the share of consumption for economic or political reasons.

The tendency for aid flows to "stick" as increased government expenditures, rather than result in tax relief or offsetting expenditure reductions in other areas, is sometimes called the *flypaper effect*. Several studies find that $1 in aid translates to less than $1 in government spending, typically in the range of 33 to 67 cents on the dollar. But some studies for specific countries find a strong flypaper effect, with $1 of

aid translating into close to $1 of new spending. One study on Indonesia found that $1 in aid led to $1.50 in government expenditures, suggesting that aid "crowded in" government spending, perhaps by relieving constraints that were limiting government spending.[23]

Aid flows can similarly influence government revenues. A government that receives aid flows is likely to be less willing to raise marginal tax revenue from the general public, especially because taxes are not politically popular. A government without aid inflows might be willing to raise taxes to finance a new hospital, but if a donor finances the hospital, the government has much less incentive to raise the revenue. However, aid could help increase revenues if donors require governments to provide matching funds for a particular activity; if aid helped governments introduce institutional changes to strengthen the tax authority, or in the case of loans, governments knew they would need revenues in the future to repay the loans. The empirical evidence on this effect is mixed: In some countries, aid inflows corresponded with increased revenues, whereas in other cases, revenues declined.[24] Where revenues declined, the reduction may benefit ordinary citizens, who can save and consume more for their own purposes, which could help accelerate economic growth. Or it may be directed to individuals or businesses with special connections to the government.

Private saving might also be affected by aid inflows. A host of subtle influences from aid on relative prices may contribute to substitution that reduces private saving and crowds out private sector investment. Aid inflows could reduce interest rates, which in turn could reduce private saving. If investment is subject to diminishing returns, more capital in general could mean lower returns on investments and hence a greater tendency to consume rather than save in the recipient countries.

It is difficult, if not impossible, for donors to stop this substitution, although they may be able to limit it to some extent. Research suggests that aid funds are fungible, at least partially, meaning that recipients, to some extent, can substitute aid funds from their intended purpose to something else. However, the evidence suggests that aid is not perfectly fungible: Much of it goes to its intended purpose, and recipients cannot or do not transform all aid flows into other purposes.

AID DEPENDENCY Some analysts argue that aid loses some of its effectiveness when aid flows are relatively large for long periods of time, because prices, institutions, and expectations adjust to that level and, to some extent, recipients become

[23]See World Bank, *Assessing Aid: What Works, What Doesn't, and Why* (New York: Oxford University Press, 1998); Tarhan Feyzioglu, Vinaya Swaroop, and Min Zhu, "A Panel Data Analysis of the Fungibility of Foreign Aid," *World Bank Economic Review* 12, no. 1 (1998). For the study on Indonesia, see Howard Pack and Janet Rothenberg Pack, "Is Foreign Aid Fungible? The Case of Indonesia," *Economic Journal* 100 (March 1990).

[24]Mark McGillvary and Oliver Morrisey, "A Review of the Evidence on the Fiscal Effects of Aid," Research Paper no. 01/13 Center for Research in Economic Development and International Trade, Nottingham, UK, 2001. For a classic model on these effects, see Peter Heller, "A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment, and Taxation," *American Economic Review* 65 (1975), 429–45.

dependent on aid. This situation can raise several challenges. First, the larger the amount of aid, the greater is the risk of a sudden reduction in aid. Aid flows tend to be more erratic than domestic revenues, and the degree of volatility increases with greater dependence on aid.[25] If donors suddenly reduce aid, whether because they face budget cutbacks, in response to a corruption scandal, a change in government, or for some other reason, the macroeconomic consequences can be difficult to manage. Second, large aid flows give donors substantial leverage over the recipient, so that recipients become more willing to introduce whatever conditions the donors require, good or bad, to ensure that the aid continues to be disbursed. Aid flows thus can weaken public accountability and impede the development of civil society, if the recipient government becomes more responsive to the donors than to its own citizens.

As recipients become more dependent on aid, the process of gradually reducing aid over time as development proceeds becomes more difficult. If aid finances public expenditure, a reduction of aid must be matched with either an increase in revenue or a reduction in expenditure, neither of which is particularly easy or politically popular (although raising revenue is made easier by the process of economic growth). The incentives are strong for recipients to continue to request aid for as long as possible and resist these adjustments. Egypt has been one of the world's largest aid recipients since 1980, and some observers believe that these inflows have taken the pressure off of the government to reduce wasteful expenditures and introduce other needed reforms. The downfall in 2011 of the Egyptian government of Hosni Mubarak, in part, may have been an indirect result of these consequences of aid.

Aid programs may actually weaken the institutions that society needs to enhance the process of development and reduce the need for aid. Budget institutions might be weakened by aid, especially if most aid is disbursed as projects with the money not flowing through the budget. Aid agencies need qualified accounts, economists, and financial specialists to design, monitor, and evaluate their projects. By increasing the demand for these skills, aid agencies bid up their wages, making them more expensive for everyone else, including the government (not to mention domestic firms). Aid agencies sometimes recruit the most able and promising local staff, leaving government agencies and the rest of the economy with less-qualified (and usually lower-paid) staff. In Mozambique, after the end of its civil war in the early 1990s, donors came flooding into the country and recruited many of the most-capable government workers, paying upward of five times the government salary and sometimes more than private sector salaries. But the departure of some of its most talented staff deprived the government of needed skills, and weakened government effectiveness as a result.[26]

[25]Aleš Bluír and Javier Hamann, "How Volatile and Unpredictable Are Aid Flows, and What Are the Policy Implications?" IMF Working Paper No. 01/167, International Monetary Fund, Washington, DC, 2001.

[26]Peter Fallon and Luiz Pereira da Silva, "Recognizing Labor Market Constraints: Government-Donor Competition for Manpower in Mozambique," in D. Lindauer and B. Nunberg, eds., *Rehabilitating Government* (Washington, DC: World Bank, 1994).

This process is not limited to the budget office: It can affect the central bank, the ministry of health (which might lose physicians or epidemiologists), the ministry of education, and other agencies. Aid agencies might be able partially to countervail these tendencies by supplying technical assistance to train government workers (although some then take other jobs) or otherwise strengthen institutional systems. Over time, there may be some benefit, as the increased wages entice more school graduates to become accountants and financial experts. In any case, aid agencies often do not take into account these possible deleterious impacts.

To some extent, the dependency can work both ways, as aid agencies become accustomed to operating large-scale programs. Donor agencies, in effect, are asked to work themselves out of their jobs by making aid less necessary and less important over time. But, like any agency, most aid agencies like to be bigger and have more aid flows to disburse. Individuals working on specific projects often like to see them continued over time; and the larger the programs, the more influence they may have over government policy choices. Promotions and other rewards often are geared to the size of the aid program they manage. Thus the incentives facing the donors sometimes encourage dependency between donor and recipient.

Proponents of View 2 point to a body of research that finds little relationship between aid and growth. Keith Griffin and John Enos were among the first to report empirical research questioning the effectiveness of aid, finding negative simple correlations between aid and growth in 27 countries. Since then, other studies concluded that there is no relationship or even a negative relationship between aid and growth,[27] in contrast to the studies cited earlier that support View 1.

VIEW 3. AID HAS A CONDITIONAL RELATIONSHIP WITH GROWTH, STIMULATING GROWTH ONLY UNDER CERTAIN CIRCUMSTANCES, SUCH AS IN COUNTRIES WITH GOOD POLICIES OR INSTITUTIONS

This perspective begins by accepting the idea that aid has had mixed results: Even where research has found a generally positive relationship, no one claims that aid has worked across all countries all the time. It recognizes that aid seems to have stimulated growth in some countries under certain circumstances but not in others and focuses on trying to decipher the key characteristics that might explain these

[27]See, for example, K. B. Griffin and J. L. Enos, "Foreign Assistance: Objectives and Consequences," *Economic Development and Cultural Change* 18, no. 2 (July 1970), 313–27; Paul Mosley, "Aid, Savings, and Growth Revisited," *Oxford Bulletin of Economics and Statistics* 42, no. 2 (1980), 79–96; Peter Boone, "The Impact of Foreign Aid on Savings and Growth," Working Paper no. 677, Centre for Economic Performance, London School of Economics, London, 1994; William Easterly, Ross Levine, and David Roodman, "New Data, New Doubts: A Comment on Burnside and Dollar's 'Aid, Politics, and Growth,'" *American Economic Review* 94, no. 3 (June 2004); David Roodman, "The Anarchy of Numbers: Aid, Development, and Cross-Country Empirics," Working Paper no. 32, Center for Global Development, Washington, DC, July 2003. For a review, see Clemens et al., "Counting Chickens When They Hatch."

differences. This view is represented by Figure 14–5c. Three "conditional" explanations have emerged, suggesting that the relationship between aid and growth might depend on the characteristics of the recipient country, the type of aid being provided, or the way in which donors provide it.

CHARACTERISTICS OF THE RECIPIENT COUNTRY The most influential of the conditional perspectives is that the impact of aid depends on the quality of institutions and policies in the recipient country. According to this view, in countries with poor macroeconomic and trade policies, high levels of corruption, and low accountability of government officials, aid is likely to have little or no impact. By contrast, in countries with reasonably good policies and institutions, aid programs can help accelerate growth. Jonathan Isham, Daniel Kaufmann, and Lant Pritchett initiated this line of research by finding that rates of return on aid-financed investments were higher in countries with strong civil liberties and other measures of governance. Craig Burnside and David Dollar took the next step by finding a significant positive relationship between aid and growth in countries with good policies and institutions and no relationship otherwise.[28] Other researchers explored the possibility that the impact of aid may differ depending on the type of government, quality of human capital, location in the tropics (presumably a proxy for health), the magnitude and frequency of export shocks, and other factors.

Although subsequent analysis cast doubt on some of these results, the idea has had an enormous influence on donor policies, probably because the conclusions of a differential effect are consistent with the experience of many development professionals. These findings have led to a shift among donors to be more "selective" in allocating foreign aid, to provide more of it to countries with relatively stronger institutions and policies, as we discuss later in the chapter.

TYPE OF AID Different kinds of aid might affect growth in different ways. One study disaggregated aid flows into those most likely and least likely to affect growth within a few years, if at all.[29] It separated aid into three categories:

- Emergency and humanitarian aid, likely to be negatively associated with growth because it increases sharply at the same time growth falls because of an economic shock.
- Aid that might affect growth only after a long period of time, if at all, and so a growth impact may be difficult to detect, such as aid for health, education, the environment, and to support democracy.

[28]Jonathan Isham, Daniel Kaufmann, and Lant Pritchett, "Governance and Returns on Investment: An Empirical Investigation," Policy Research working paper no. 1550, World Bank, 1995; Craig Burnside and David Dollar, "Aid, Policies, and Growth," *American Economic Review* 90, no. 4 (September 2000), 847–68; Paul Collier and David Dollar, "Aid Allocation and Poverty Reduction," *European Economic Review* 45, no. 1 (2002), 1–26; World Bank, *Assessing Aid: What Works, What Doesn't, and Why* (New York: Oxford University Press, 1998).

[29]Clemens et al., "Counting Chickens When They Hatch."

- Aid directly aimed at affecting growth (building roads, ports, and electricity generators, or supporting agriculture).

It found a strong relationship between the third type of aid and growth, but the relationship with the other types was less detectable. The study found some support for the proposition that the third subcategory of aid had an even stronger impact in countries with stronger institutions.

DONOR PRACTICES Many analysts argue that differences in donor practices are likely to influence aid effectiveness. Multilateral aid might be more effective in stimulating growth than bilateral aid because less of it is determined by political factors (although some argue just the opposite, that multilateral agencies tend to be too large, unfocused, and less directly accountable to taxpayers). Aid "tied" to purchases of good and services in the donor countries or used to finance conferences and international meetings might be less effective than other aid. Large aid bureaucracies add to costs and slow down disbursements, making aid less effective. Donors that coordinate with other funders and recipient governments may be more effective than those that operate completely on their own. Donors with more effective monitoring and evaluation systems might be better able to channel aid to its most productive uses, enhancing overall effectiveness. An influential view that emerged in the late 1990s is that aid might be more effective where there has been a participatory approach among government and community groups in recipient countries in setting priorities and designing aid-supported programs, a topic we return to later in the chapter. Substantial debate centers around these issues, and there is little doubt that donor practices are critical, but to date there has been very little systematic research connecting specific donor practices to aid effectiveness.

In summarizing the aid and growth research, it seems that aid has been successful in some countries but not others, with the overall trend a subject of debate. Further research is looking beneath the surface to reveal what types of aid are most effective, which are ineffective, and the precise conditions under which aid has the largest impact on growth. Just as with the evidence on the causes of economic growth more generally discussed in Chapter 3, there is still much that we do not know about the relationship between aid and growth.

DONOR RELATIONSHIPS WITH RECIPIENT COUNTRIES

The criticisms about foreign aid and the evidence that it appears to have been effective in some countries but not in others led to debates about how aid programs can be improved to more effectively support growth and development. But the challenge is not easy. Aid programs face some inherent difficulties in trying to achieve a wide

range of objectives (which may differ between donor and recipient), provide financial oversight, and ensure results. In this section, we first explore some of the challenges in more depth, then turn to some specific changes that have been suggested and in some cases implemented by donors and recipients in an attempt to make aid more effective.

THE PRINCIPAL-AGENT PROBLEM

A key issue facing aid agencies is that there is only an indirect and distant relationship between the people actually providing the financing (taxpayers in donor countries) and the intended ultimate beneficiaries of aid projects (poor people living in low-income countries). Many institutions, organizations, and transactions are in between, and the decisions made along the way may not be in accord with the wishes or best interests of either the original taxpayers or the ultimate beneficiaries. All public-sector agencies and many private companies are faced with the **principal-agent problem**, but the international dimension makes it an even greater challenge for aid. Economist Bertin Martens analyzed the principal agency problem for aid agencies, and our account closely follows his.[30]

Principals in a private company (the owners), club (the members), or public administration (taxpayers) cannot make all decisions and carry out all tasks themselves, so they must delegate these responsibilities to *agents* to work on their behalf: managers, employees, elected officials, and civil servants. But agents have their own goals and motivations, which may not be the same as the principals'. Agents also have more information than the principals and can use that information in ways that run counter to the principals' interests. Therefore, principals are faced with the problem of writing contracts and establishing rules that more closely align agents' interests with their own. These issues are particularly important for public-sector agencies because they tend to have multiple objectives (not just profit) and their many principals may not agree on which objective holds the higher priority. These problems make it very difficult to match incentives, such as salary, bonuses, and promotions, with performance indicators.

Most aid programs have a long, complex chain of principal-agent relationships, starting with the taxpayers who delegate authority to elected officials, who in turn become principals who delegate authority to a new set of agents, the heads of aid agencies. These relationships continue to aid agency employees, contractors, and consultants. In the recipient country, similar relationships lie between the ultimate recipients, their government, and those who actually implement programs. In most public agencies, there is considerable slippage throughout these relationships, but

[30]Bertin Martens, "Introduction" in Berten Martens, ed., *The Institutional Economics of Foreign Aid* (Cambridge: Cambridge University Press, 2004).

the problem is compounded in aid agencies by the physical separation of the original taxpayers and ultimate beneficiaries. In domestic public programs (such as trash collection or local schools), the taxpayers and ultimate beneficiaries are the same people, so they may have clearer information about success or failure and can reward or penalize their agents accordingly by reelecting them or voting them out of office. But this feedback loop is broken for aid agencies, as Martens points out:

> A unique and striking characteristic of foreign aid is that the people for whose benefit aid agencies work are not the same as those from whom the revenues are obtained; they actually live in different countries and different political constituencies. This [separation] blocks the normal performance feedback process: beneficiaries may be able to observe performance but cannot modulate payments (rewards to agents) as a function of performance. Although donors are typically interested in ensuring that their funds are well spent, it is extremely difficult for them to do so, since there is frequently no obvious mechanism for transmitting the beneficiaries' point of view to the sponsors.[31]

The principal-agent problem affects nearly all aspects of aid delivery, including program design, implementation, compensation, incentives, evaluation, and allocation of funding. The problem can never be fully avoided: Private companies face similar issues between owners, managers, and employees, as do private aid foundations and charities. The challenge is to design institutions and incentives to try to mitigate these problems as much as possible to clarify goals, objectives, incentives, and rewards. A key challenge for donors is how best to apply conditions to their loans to encourage recipients to act more in accord with the donors' (and possibly the ultimate beneficiaries') interests and wishes.

CONDITIONALITY

Donors often require that recipient countries adopt specific policies, or packages of policies, as prerequisites for funding; and this **conditionality** is one of the most controversial aspects of aid. Broad policy conditions are most often associated with financing from the IMF and World Bank, but all donors use them to some extent. Many bilateral donors take their lead from these two multilateral organizations and do not provide the full amount of their aid until the recipient country has met IMF and World Bank conditions. The IMF typically requires that countries adopt a more flexible exchange rate, smaller budget deficits, slower growth of the money supply, a buildup of foreign exchange reserves, privatization of certain state-owned enterprises, a reduction in import restrictions, and the implementation of broad anticorruption measures. World Bank conditions sometimes reinforce these broad policy

[31]Martens, "Introduction," 14.

reforms and may add some sector-specific requirements, such as liberalizing fertilizer markets before providing an agriculture loan.

In many cases, the specific policy conditions required by the IMF and World Bank are justifiable and often supported by at least some people within the recipient governments. But, other times, many people believe that IMF and World Bank conditions go too far and their conditions are harder to justify, especially when the rationale for specific reforms are less clear-cut and seen by critics as based more on ideology than hard facts and sensible economics. For example, while privatization of state-owned marketing boards, retail outlets, or trading establishments is generally not controversial, privatization of public utilities is much more open to debate. Many countries may need to impose fiscal discipline as required by the IMF, but the Fund is often accused of imposing much stricter discipline than is really necessary to achieve stability, at a cost of dampening aggregate consumption and expenditures on important social programs. Sometimes, the donors simply impose too many conditions, seemingly wanting the recipient country to fix everything at once without showing any sense of priority or feasibility. Turkey's "Letter of Intent" with the IMF in April 2003 included a table listing 131 very specific policy actions (with dozens more subcategories of actions) that the government promised to undertake as part of its program.[32] The IMF admitted that, at times, it has gone too far, as with its program in Indonesia after the financial crisis of 1997.[33]

While donors are often criticized for imposing too many conditions, they are almost as often criticized for not imposing *enough* conditions. Some advocates that criticize the IMF for imposing too much fiscal austerity also insist that it require governments to spend a minimum amount on health and education. The World Bank is often asked to add conditions to force governments to take specific actions, for example, on projects that have potential environmental consequences. This can lead to difficult dilemmas. Sometimes governments ask the World Bank to be one of many co-funders for a project, hoping that World Bank participation will provide credibility and security. In return, the World Bank typically wishes to require environmental safeguards or anticorruption measures. But because the World Bank is only one lender, its leverage may be limited. If the World Bank requires too many conditions, the government can just go ahead without them (and without the environmental safeguards), but if they ask for too few, they may find themselves funding a project with inadequate conditions.

The rationale for these conditions is straightforward: Donors believe these broad-based conditions are important for growth and development, and without them, providing aid is futile. It is easiest to see this rationale in extreme cases: If gov-

[32]See International Monetary Fund, "Turkey: Letter of Intent," April 5, 2003, available at www.imf.org/external/np/loi/2003/tur/01/index.htm; accessed February 2012.

[33]IMF Independent Evaluation Office, "The IMF and Recent Capital Account Crises: Indonesia, Korea, and Brazil," (Washington, DC: IMF, 2003). Available at www.imf.org/external/np/ieo/2003/cac/pdf/all.pdf.

ernment policies have led to high rates of inflation, an overvalued exchange rate, massive inefficiencies and waste of public spending, and extensive corruption, then providing aid, whatever the specific purpose, without requiring fundamental change provides no benefits and perhaps perpetuates the damage. Some even argue that the primary purpose of aid is not the money but for aid to act as a lever for the policy reforms. The conditionality debate is a microcosm of the principal-agent problem: A donor requests a recipient to take specific actions in return for receiving aid, but the recipient may have different objectives and controls sufficient information to make compliance difficult.

Cornell economist Ravi Kanbur points out two further problems with conditionality. First, it is not always clear what conditions are the most appropriate to ensure sustained growth and development. Mahbub ul Haq, an influential development adviser long associated with the UNDP, once commented that sub-Saharan Africa, "often receives more bad policy advice per capita from foreign consultants than any other continent in the world."[34] Development doctrine, as discussed in Chapter 5, has swung from a state-led approach in the 1950s and 1960s, to basic human needs in the 1970s, to a macroeconomic approach focused on open markets in the 1980s and 1990s, to a focus on institutions beginning in the mid-1990s, to a greater focus on private sector led growth in the 2000s. As a result, the list of conditions is constantly evolving.

Second, conditionality does not seem to work. Most analysts agree that governments implement reforms only when it is in their interests to do so, and donor conditions have little, if any, impact on that decision. They perhaps can spur governments to implement changes faster than they otherwise would, or provide support to reformist elements in policy debates, but cannot persuade governments to do what they really do not want to do. Many donors continue to disburse aid even when recipients fail to meet conditions, sometimes repeatedly so. Over time, recipients learn that aid flows do not necessarily depend on meeting stated conditions, a process that gradually undermines the aid institutions and the conditions they attach. Donors are faced with their own internal incentives to continue to disburse aid to support the contractors and recipients that depend on it. They also face a "Samaritan's dilemma," that withdrawing aid would create short-term pain for the very people it is aimed to help.[35]

In the end, there are no clear-cut rules for conditionality. Striking the right balance between responsible oversight and accountability, on the one hand, and ensuring against

[34]Mahbub ul Haq, "Does Africa Have a Future?" *Earth Times New Service,* January 1998.

[35]Ravi Kanbur, "The Economics of International Aid," in Serge Christophe-Kolm and Jean Mercier Ythier, eds., *Handbook of the Economics of Giving, Altruism, and Reciprocity: Applications* (Amsterdam: North Holland, 2006). Also see Jakob Svensson, "Why Conditional Aid Does Not Work and What Can Be Done About It," *Journal of Development Economics* 70, no. 2 (2003), 381–402; William Easterly, *The Elusive Quest for Growth* (Cambridge: MIT, 2002).

high bureaucratic obstacles and the imposition of unnecessary controls or unwarranted policy changes, on the other, requires flexibility, judgment, and the ability to balance multiple objectives.

IMPROVING AID EFFECTIVENESS

Since the late 1990s, as global aid flows began to rise, there has been increased discussion and debate about how aid programs could be strengthened and become more effective in supporting growth and development. These debates recognized some of the weaknesses in aid programs and resulted in some specific ideas for change. The "Paris Declaration on Aid Effectiveness and the Accra Agenda for Action" is perhaps the clearest articulation of the consensus on best practices in aid delivery.[36] In recent years, donor agencies began to put some of these ideas in practice, and some donor practices changed noticeably as a result.

COUNTRY SELECTIVITY One influential idea is that donors should be more selective about which countries they provide aid to, based on the view that aid works best in countries with good policies and institutions. In the strongest version of this view, aid should be provided *only* to countries that meet these criteria and not otherwise, with more going to countries with the highest levels of poverty. A more moderate view is that *more* aid should be allocated to countries with stronger policies and institutions but continue to provide targeted aid to some countries with relatively poor institutions, especially in postconflict situations. This proposal takes a turn on the conditionality debate: Instead of providing aid to encourage reforms, provide it to countries that already decided to implement key reforms. Economists Paul Collier and David Dollar suggested a "poverty-efficient" allocation of aid in which funding would be provided to the poorest countries with relatively stronger policies and institutions to maximize the impact of aid on global poverty reduction.[37]

Donors began to move in this direction. The World Bank uses its Country Policy and Institutional Assessment index to determine partly the allocation of its concessional IDA funds. Several European donors moved toward providing broad budget support or financing for sector-wide approaches, but only for a relatively small number of countries considered to be the most responsible. The U.S. Millennium Challenge Corporation (MCC) provides aid to only a small number of recipient countries, based largely (although not completely) on their performance on 17 indicators of policies and governance.[38] As of late 2011, the MCC had signed compacts with 23 low-income and lower-middle-income countries around the world.

[36]See the Organization for Economic Cooperation and Development (2008), "The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action," available at www.oecd.org/dataoecd/11/41/34428351.pdf; accessed February 2012.

[37]Collier and Dollar, "Aid Allocation and Poverty Reduction."

[38]Steven Radelet, *Challenging Foreign Aid: A Policymaker's Guide to the Millennium Challenge Account* (Washington, DC: Center for Global Development, 2003).

Because so much aid is allocated for political, security, and other foreign policy reasons, there are limits to how far donors are likely to go in reallocating their aid based on strict economic development criteria. This is especially true for the major bilateral donors but is true for multilateral donors as well. One frequent suggestion is for donors to more clearly separate funding primarily aimed at foreign policy goals from funding aimed at development. This would allow programs to be designed, implemented, and evaluated in different ways so donor country governments, their taxpayers, and recipients could better understand and appraise aid effectiveness.

RECIPIENT PARTICIPATION A second influential idea is that aid has been weakened by donor domination in setting priorities, designing programs and projects, choosing implementers (often consulting firms from the donor country), monitoring and evaluating results, and that aid recipients should play a much larger role in these areas. In this view, bureaucrats and activists from donor countries design too many programs, leading to a poor choice of priorities, flawed design, or weak commitment among recipients for programs they do not feel are their own. Advocates have pushed for a more participatory approach in which various groups in recipient countries (government, NGOs, charities, the private sector) play a more active role. The idea is to eliminate some of the problems in the long chain of principal-agent relationships and more tightly integrate the ultimate beneficiaries in key aspects of the aid-delivery process. This might have two benefits. First, projects might be better designed, with a more accurate view of the highest priorities and the most appropriate implementation methods to meet local needs. Second, increasing recipient participation in the design and implementation process may provide them with more ownership of the activity and a higher stake in ensuring its success. One of the first movements in this direction was the introduction of Poverty Reduction Strategy papers as the basis for World Bank and IMF debt relief and other financing. Similarly, both the Global Fund to Fight AIDS, Tuberculosis, and Malaria and the Millennium Challenge Corporation rely on a significant degree of local participation in designing and implementing the programs they finance. USAID's Feed the Future program, which aims to improve food security by supporting increased agricultural production and improved nutrition, relies on country strategies developed by local leaders and national experts.

While the participatory approach holds out promise and has been increasingly used since the early 2000s, there is no clear empirical evidence yet on the extent to which (or the circumstances under which) it improves aid effectiveness. There is a clear and inescapable tension between country ownership, on the one hand, and donor priorities and conditionality, on the other. Donors are more likely to facilitate a participatory approach in countries in which governments show a strong commitment to sound development policies and less so in countries with corrupt and dictatorial governments.

HARMONIZATION AND COORDINATION Managing aid flows from many different donors can be a constant challenge for recipient countries because different donors

tend to implement their own initiatives and insist on using their own unique processes for initiating, implementing, and monitoring projects. Recipients can be overwhelmed by donor requirements to provide multiple project audits, environmental assessments, procurement reports, financial statements, and project updates. According to the World Bank, developing countries typically work with 30 or more aid agencies across a wide variety of sectors, with each sending an average of five missions a year to oversee their projects. Governments can find themselves hosting three or more aid missions a week.[39] Many recipient countries have only a limited number of skilled and highly trained technocrats; and all the donors want to meet with these top people, leaving them with much less time to deal with other pressing concerns. The government of Tanzania, which hosts several hundred aid missions each year, has introduced a "quiet time" from April to August of each year during which it asks donors to minimize meetings and missions so that the government has time to adequately prepare its annual budget.

These concerns have led to numerous suggestions and pledges for donors to more closely coordinate their activities; harmonize their accounting, monitoring, and evaluation systems; or pool their funds.[40] Some progress has been made in the form of donors providing more aid as budget support in certain countries, acting together through joint missions, and agreeing to use similar monitoring and evaluation procedures. At the same time, some newer donor initiatives appear to want to use their own new methodologies, which could compound this problem going forward.

RESULTS-BASED MANAGEMENT The emphasis on demonstrating the effectiveness of aid has led to calls for improved monitoring and evaluation and results-based management. In this view, aid programs should aim to achieve very specific quantitative targets, and decisions about renewing or reallocating aid going forward should be based on those results. There are three basic objectives. First, results-based management can help donors allocate funds toward programs that are working. Second, ongoing reviews can detect problems at an early stage and help modify and strengthen existing programs. Third, donors and recipients can better learn what approaches have worked and what have not. Stronger monitoring and evaluation helps strengthen the principal-agent relationship so that aid agencies have clearer incentives and taxpayers have better information about the impact of aid on its intended beneficiaries. USAID, for example, introduced in 2011 a new evaluation approach aimed at strengthening the impact of the agency's program on achieving specified results. There has been an explosion in empirical investigations on the impact of development interventions through randomized controlled trials, led by

[39]World Bank Development News Media, "Cutting the Red Tape," February 21, 2003.

[40]Ravi Kanbur and Todd Sandler, "The Future of Development Assistance: Common Pools and International Public Goods," Policy Essay No. 5, Overseas Development Council, Washington, DC, 1999.

the path-breaking work of the Jameel Poverty Action Lab at MIT. These and other similar efforts are leading to a much deeper understanding of what works and what does not work in aid-supported programs around the world.

SUMMARY

- Most aid is given by bilateral donors, but a significant portion is channeled through multilateral agencies. Global aid flows peaked in 1991, then declined in real and nominal terms until 1997, before rebounding again. Official development assistance is higher today, almost $130 billion in 2010, than ever before but not as a share of the GNI of today's developed nations. The United States currently provides the largest dollar amount of aid but is among the smallest in terms of aid as a share of its income.
- For some countries, aid flows are large and significant, while in others they are small. Sub-Saharan Africa received aid flows equivalent to about 5 percent of its income in 2009, or about $53 per person. Because of the war in Iraq, the Middle East received $42 per person in 2009, an amount much higher than the $9 per person received by South Asia, which is a much poorer region.
- For most donors, the primary motivation for providing aid is to support foreign policy objectives and political alliances. Income and poverty tend to be secondary objectives. Country size, commercial and historical ties, and the extent of democracy also play a role.
- There are three broad viewpoints on the relationship among aid, growth, and development. First, some analysts believe that aid supports growth and development by adding to investment and the capital stock, helping in the transfer of technology, supporting key health and education programs, and enhancing economic stability. Second, others believe that aid has no impact on growth and might undermine growth by distorting incentives for private production, encouraging corruption, enlarging the government, or creating aid dependency. Third, some argue that aid works under certain conditions but not others, depending on the policy and institutional environment in the recipient country, the purpose of the aid, or the donor's practices and procedures. The empirical evidence on these relationships is mixed, with different studies reaching different conclusions.
- Aid can influence both saving rates and tax revenues through a host of subtle impacts on prices, preferences, and incentives. Aid flows add to total saving and investment but less than one for one. The fungibility of aid can cause private and government saving to decline as aid increases. The impact on government revenues varies across countries, with some

experiencing a decline and others an increase corresponding to higher aid flows.

- Aid agencies face a classic principal-agent problem, compounded by the fact that the original funders (taxpayers in donor countries) have only indirect and distant connections to the intended ultimate beneficiaries (poor people in low-income countries). A particular challenge for donors is how to place appropriate conditions on their aid. Donors often are criticized for putting too many, too few, or the wrong kinds of conditions on aid, and conditionality in general appears to have been largely ineffective.
- In recent years, donors began to try to make aid more effective by becoming more selective in choosing recipients, encouraging more participation by recipients in program design and implementation, coordinating more closely with each other, and managing programs based on results. It is too early to judge the effectiveness of these changes, but they are likely to affect aid programs for some time to come.

15 Managing Short-Run Crises in an Open Economy

Economic development takes place in the long term. Most of the processes discussed in the previous chapters, whether improving human welfare, increasing saving, or shifting toward manufactured exports, take years and even decades to bear significant results. If policy makers in developing countries gaze only at the far horizon, however, they are unlikely ever to reach it. Much happens in the short term, within a few months or a couple of years, to throw an economy off balance and make pursuing long-term strategies difficult and sometimes impossible. Policy makers need to emulate a ship's captain, who, always steering toward the port of destination, nevertheless must deal decisively with any storms at sea.

Among the most dangerous and likely of these storms are changes in world prices that throw the balance of payments into deficit, excessive spending that fuels inflation or unsustainable debt, and droughts or other natural disasters that disrupt production. Unless a government counteracts these economic shocks, they create greater uncertainty and higher risk for private producers and investors, who take evasive actions that reduce future investment, worsen the crisis, and cause development efforts to flounder.

During the 1970s and 1980s, the late 1990s, and in Europe in 2010–11, as pointed out in Chapters 12 and 13, many economies became unbalanced because of unstable world market conditions and their own macroeconomic mismanagement. In Chapters 5 and 11 through 13, we discussed the consequences of such **macroeconomic instability**. Countries with overvalued exchange rates and rapid inflation were unable to grow rapidly. *Stabilization programs*, many funded by the International Monetary Fund (IMF), were intended to correct these macroeconomic imbalances.

In this chapter, we develop a mechanism for analyzing the macroeconomic policies that a developing country should pursue to stabilize its economy and create a climate for faster economic growth. The model developed here incorporates the two main policy approaches for correcting macroeconomic imbalances: changing the level of domestic expenditures and adjusting relative prices. In many cases, managing economic crises specifically requires expenditure reduction (by lower government budget deficits and slower creation of money) and exchange-rate devaluation.

EQUILIBRIUM IN A SMALL, OPEN ECONOMY

Developing economies[1] have two features central to understanding how macroeconomic imbalances occur and can be corrected. First, they are **open economies**, in that trade and capital flow across their borders in sufficient quantities to influence the domestic economy, particularly prices and the money supply. Most economies are open in this sense, especially because of economic reforms in China beginning in the late 1970s and in Eastern Europe and the states of the former Soviet Union beginning in the early 1990s. Today only a few economies, such as Cuba, North Korea, and Burma, are so heavily protected and regulated (and subject to foreign embargoes) that they might not qualify as open to trade and finance.

Second, these are **small economies**, meaning that neither their supply of exports nor their demand for imports has a noticeable impact on the world prices of these commodities and services. Economists call these countries *price takers* in world markets. A number of developing countries can exert some influence over the price of one or two primary exports in world markets: Brazil in coffee, Saudi Arabia in oil, Zambia in copper, South Africa in diamonds, for example. But they almost never affect the price of goods they import, and for macroeconomic purposes, it usually is adequate to model even these countries as price takers.[2]

These two qualities, smallness and openness, are the basis for the **Australian model** of a developing economy.[3] Countries typically trade both importable and

[1]In developing this and the next two sections, we acknowledge an intellectual debt to Shantayanan Devarajan and Dani Rodrik, who wrote an excellent set of notes for their class on macroeconomics for developing countries at Harvard's John F. Kennedy School of Government in the late 1980s and to Richard E. Caves, Jeffrey A. Frankel, and Ronald W. Jones, who develop the open economy model in Chapter 19 of *World Trade and Payments: An Introduction* (Glenview, IL: Scott, Foresman, Little, Brown, 1990).

[2]Among developing countries, China and India are large enough that they could become exceptions to the small country rule, given continued growth in China and both greater growth and openness in India.

[3]So called because it was developed by Australian economists, including W. E. G. Salter, "Internal Balance and External Balance: The Role of Price and Expenditure Effects," *Economic Record* 35 (1959), 226–38; Trevor W. Swan, "Economic Control in a Dependent Economy," *Economic Record* 36 (March 1960), 51–66; W. Max Corden, *Inflation, Exchange Rates and the World Economy* (Chicago: University of Chicago Press, 1977). Australia also is a small, open economy.

exportable goods and services. The Australian model lumps importables and exportables together as *tradables* and distinguishes these from all other goods and services, called *nontradables.* (We use this specification again in Chapter 18's discussion of Dutch disease.)

Tradable goods and services are those whose prices within the country are determined by supply and demand on world markets. Under the small economy assumption, these world market prices cannot be influenced by anything that happens within the country and so are *exogenous* to the model (determined outside the model). The domestic (local currency) price of a tradable good is given by $P_t = eP_t^*$ where e is the nominal exchange rate in local currency per dollar (pesos per dollar for Mexico or rupees per dollar for Pakistan) and P_t^* is the world price of the tradable in dollars. Even if the supply of and demand for tradables change within an economy, the local price will not change because domestic supply and demand have a negligible influence on the world price. Yet, changes in the nominal exchange rate change the domestic price of tradables commodities. If a country **devalues** its nominal exchange rate, it increases the amount of local currency required to purchase a dollar (e increases); if a country **revalues** its nominal exchange rate it decreases the amount of local currency required to purchase a dollar (e decreases). Devaluations thus increase the local currency price of tradables (all else equal), while revaluations tend to decrease the local currency price of tradables. Because this model simplifies all tradables into one composite good, the price of tradables P_t is best thought of as an index, a weighted average of the prices of all tradables, much like a consumer price index.

Tradables include exportables, such as coffee in Kenya and Colombia, rice in Thailand, beef in Argentina, cattle in West Africa, palm oil in Malaysia and Indonesia, copper in Peru and Zambia, oil in the Middle East, and textiles and electronics in East Asia, and importables, such as rice in West Africa, oil in Brazil or Korea, and intermediate chemicals and machinery in many developing countries.

Nontradables are goods and services, such as transportation, construction, retail trade, and household services that are not easily or conventionally bought or sold outside the country, usually because the costs of transporting them from one country to another are prohibitive or local custom inhibits trade. Prices of nontradables, designated P_n, therefore, are determined by market forces within the economy; any shift in supply or demand changes the price of nontradables. Nontradable prices thus are *endogenous* to the model (determined within the model). The term P_n, like P_t, is a composite or weighted average price incorporating all prices of nontradable goods and services.

INTERNAL AND EXTERNAL BALANCE

Figure 15–1 depicts equilibrium under the Australian model. The vertical axis represents nontradables (N); the horizontal axis takes both exportables and importables and treats them together as tradables (T). The production possibility frontier shows

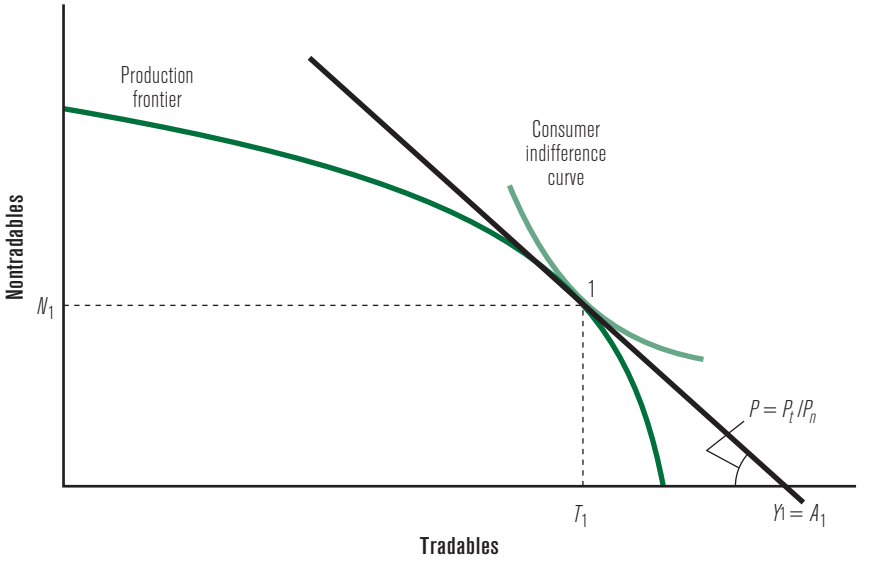


FIGURE 15–1 Equilibrium in the Australian Model

With equilibrium at point 1, the tangency of the production frontier and a community indifference curve, the country produces and consumes T_1 of tradables and N_1 of nontradables. The relative price, P, is a measure of the real exchange rate (see text). Y_1, national income measured in tradable prices.

the menu of possible combinations of outputs of the two kinds of goods, N and T. The community indifference curves show consumer preferences between consumption of tradables and nontradables.

Equilibrium is at point 1, the tangency of a consumer indifference curve and the production possibilities frontier. At this point, the production of tradables, determined by the production frontier at point 1, is T_1, equal to the demand for tradables, determined by the indifference curve at 1, and similarly, for nontradables, supply equals demand at N_1. This is a defining characteristic of equilibrium in the Australian model: At point 1, the markets for both goods are in balance. Put another way, there is **external balance (EB)**, because the supply of tradables equals demand, and **internal balance (IB)**, because the supply of nontradables equals demand.

Point 1 simultaneously indicates the optimal (profit maximizing) combination of tradables and nontradables for producers, the optimal (utility maximizing) combination of tradable and nontradables for consumers. For both producers and consumers, this optimum occurs with respect to relative prices—in this case the relative price of tradables to nontradables. This relative price is indicated by the slope of the price line in Figure 15–1. This joint equilibrium for producers and consumers is indicated by the tangency of the indifference curve and the production possibility frontier. The tangency of the indifference curve and production frontier is jointly determined with the relative price of tradables in terms of nontradables, $P = P_t/P_n$. This relative

price, P, is one way to define the **real exchange rate (RER)**, one of the important innovations of the Australian model.[4] (Box 15–1 provides a more detailed explanation of the real exchange rate.) This formulation separates out prices that are under the influence of monetary and fiscal policy and domestic market forces, P_n, from prices that can be changed only by adjustments of the nominal exchange rate, $P_t = eP_t^*$. Note that the slope of the price line that is tangent to the production possibility curve and the consumer indifference curve is the only real exchange rate consistent with equilibrium in the model.

If P rises (the price line becomes steeper in the diagram), tradables become more expensive relative to nontradables. Producers then attempt to switch along the production frontier away from N goods, toward T goods. Consumers attempt to switch in the opposite direction, up along the indifference curve to consume fewer T goods and more N goods. Therefore, a rise in P should increase the surplus of T-good production over consumption.

If the production of T goods exceeds consumption of T goods, there is an external surplus, which is identical to a surplus in the balance of trade. To see this, start with the definition of the trade balance as

$$B_t = X - M \qquad [15\text{–}1]$$

where X and M are exports and imports. Because exports are the surplus of supply over demand for exportable goods, while imports are the opposite, a surplus of demand over supply, we can write the balance of trade as

$B_t =$ value of X-goods supply $-$ value of X-goods demand $-$ (value of M-goods demand $-$ value of M-goods supply),
$=$ value of X-goods supply $+$ value of M-goods supply $-$ (value of X-goods demand $+$ value of M-goods demand),
$=$ value of tradables supply $-$ value of tradables demand;

or if we let the supply of tradables be S_t and demand be D_t,

$$B_t = P_tS_t - P_tD_t = P_t(S_t - D_t) \qquad [15\text{–}2]$$

In Figure 15–1, with the economy in equilibrium, consumption of tradables is equal to production, so the balance of trade is 0.

The value of income (GDP) also can be found in Figure 15–1. It is the sum of the value of output of N goods (N_1) and T goods (T_1). This value is given by Y_1, the

[4]Chapters 18 and 19 will define the real exchange rate index as RER $= R_oP_w/P_d$. The term R_o is an index of the nominal exchange rate; in this chapter we use e, the nominal exchange rate itself. The term P_w is an index of world prices, often the U.S. consumer or wholesale price index and is similar or identical to P^* as measured in practice. But P_d is a domestic consumer or wholesale price index that includes both tradable and nontradable prices, whereas P_n is an index of nontradable prices only. Thus, the Australian formulation of the real exchange rate is a more-precise definition than those given in the later chapters.

BOX 15-1 REAL VERSUS NOMINAL EXCHANGE RATES

Most anyone who has traveled outside their home country has experience with nominal exchange rates. The nominal exchange rate is simply the number of units of local currency you can buy from the country you visit with one unit of your own currency. For instance, in August 2010, you could buy just over 46 Indian rupees, or 12.6 Mexican pesos, for US$1. (Note, this implies that you could also have bought 3.65 Indian rupees with one Mexican peso.) But when you bring a dollar into India or Mexico what you really care about is not how many rupees or pesos you can buy with that dollar but rather the quantity of actual goods and services that you can buy with that dollar. This depends on the prices of goods and services in the host country in addition to the price of its currency. This distinction is the key idea underlying the concept of the real exchange rate (RER).

In the most general sense, RER is the relative price of foreign goods in terms of domestic goods. In practice, economists have developed a range of approaches to quantifying this idea. Economists working on developed economies typically measure the RER as the relative price of domestic and foreign goods. In contrast, economists working on developing countries typically measure the RER as the relative price of tradables and nontradables. Tradable commodities are goods that are *or could be* traded internationally, in contrast to nontradable goods (such as housing and many services), which are not traded internationally. A key practical distinction between these categories of goods is that there are world market prices for tradable goods, but the prices of nontradables are determined purely by local supply and demand conditions in each country.

The nominal and real exchange rates are linked together by the requirement that the relative prices of tradable and nontradable goods in the RER be expressed in the same currency units. The prices of tradables are typically expressed in U.S. dollars, whereas the prices of nontradables are expressed in units of the local currency. Thus we need to use the nominal exchange rate to convert the dollar-denominated price of tradables into units of the local currency to calculate the RER.

We can construct this RER as follows. Expressing the nominal exchange rate, e, in terms of the number of local currency units per dollar, and expressing the prices of tradables and nontradables as P_t and P_n, respectively, we can construct the RER as

$$\text{RER} = \frac{eP_t^*}{P_n} = \frac{P_t}{P_n}$$

In this equation, the asterisk on P_t in the second term indicates that the price of tradables is expressed in dollars. Multiplying this dollar-denominated price by e, the nominal exchange rate as defined above, converts the price of tradables into the same local currency units as the price of nontradables.

It is important to consider several issues related to this concept of the RER. First, note that the RER is based on price indices rather than actual nominal price levels. The RER is thus expressed relative to a base year, and changes in the relative price of tradables to nontradables, say from 1 to 1.2, would indicate percentage changes (in this case, 20 percent) relative to the base year. We refer to an increase in the RER as a *depreciation* of the local currency, and a decrease in the RER as an *appreciation* of the local currency. While this may sound counterintuitive, the rationale for these terms is that when a currency depreciates in real terms, a given quantity of foreign goods can be exchanged for a greater quantity of that country's domestic goods (and vice versa in the case of an appreciation). It is for this reason that the RER is often thought of an as an indicator of a country's international competitiveness: When a country's currency depreciates in real terms relative to its trading partners' currencies, that country's goods become less expensive to foreigners.

An important practical challenge in constructing an RER lies in the need to choose price indices for tradables and nontradables. P_t and P_n are indices of the prices of entire categories of goods. Thus constructing these price indices first requires deciding which goods (and services) belong in which category. Specific price data may also be lacking. One short cut for addressing these challenges may be to use the U.S. consumer price index in place of P_t and a similar indicator from the home country (with the same base year) in place of P_n.

Although the availability of such price indicators as the consumer price index makes them convenient, their use in constructing RERs is problematic. Theory calls for an index of nontradables prices, but the consumer price index (CPI) is typically constructed to reflect the price of a basket of consumption goods that includes both tradable and nontradable goods. The larger the share of tradable goods in that basket, the greater the divergence between what the RER tells us in theory and what we actually measure if we construct an RER using those broad price indices. In practice, aggregate price indices purely for nontradables rarely exist. Similar problems exist in choosing a price index to represent tradables prices (for use in the numerator of the RER for a given country). In this case, in which the goal is to choose a price index based to the greatest extent possible on tradables, many authors use the wholesale price index (WPI) from the United States or from a given country's trading partners. Yet this approach too is problematic because (as Lawrence Hinkle and Peter Montiel note[a]) for-

[a]Lawrence E. Hinkle and Peter Montiel, *Exchange rate misalignment: concepts and measurement for developing countries*. (Washington, DC: World Bank, 1999).

eign WPIs may not provide a very close indication of the tradables prices actually faced by consumers in the home country. There is no perfect match between the theoretical requirements and practical data availability in constructing empirical RERs. A common compromise is to use the foreign WPI to represent P_t and the domestic CPI to represent P_n.

An additional question is whether the relevant RER is purely between two specific countries (the home country and a single trading partner—that is, the bilateral RER) or between the home country and multiple trading partner countries. In general, policy makers in a given country will be more concerned with how their currency relates in real terms with all of their trading partners. In that case, it is necessary to take a (trade-weighted) average of all the bilateral RERs between the home country and its trading partners. This average is called the **real effective exchange rate (REER)**.

The central challenge for policy makers concerned with their country's international competitiveness is whether the level of the REER at any given time reflects its equilibrium value or whether it is overvalued or undervalued relative to that equilibrium. Equlibrium in this setting generally refers to the level of the REER at which a country's internal market (that is, its supply and demand for nontrable goods and labor) and its external market (that is, its supply and demand for tradable goods) are in balance.

intersection of price line P from point 1 to the T axis.[5] In national income accounting, we distinguish two concepts. Gross domestic *product*, a measure of the value of output, is given by

$$\text{GDP} = C + I + X - M \qquad [15\text{–}3]$$

where C and I are consumption and investment by both the government and the private sector. Gross domestic *expenditure*, often called **absorption**, is

$$A = C + I = \text{GDP} + M - X \qquad [15\text{–}4]$$

When, as in Figure 15–1, the economy is in equilibrium, $X = M$ and income equals absorption. Indeed, this is the other condition for equilibrium in the Australian model. From equation 15–4, we can also see that $A - \text{GDP} = M - X$, indicating

[5]Along the T axis, Y_1 is measured in prices of the T good, so $P_tY_1 = P_tT_1 + P_nN_1$ or $Y_1 = T_1 + (P_n/P_t)N_1$. But $P_n/P_t = \Delta T/\Delta N$, with $\Delta N = N_1$ and $\Delta T = Y_1 - T_1$, the distance along the T axis from T_1 to Y_1. Thus the value of both goods in T prices is $T_1 + Y_1 - T_1 = Y_1$.

that any excess of expenditures over income implies a negative trade balance of equal amount. (The appendix to this chapter provides a more detailed review of balance of payments accounting.)

This exploration of the Australian model yields three results. First, macroeconomic equilibrium is defined as a balance between supply and demand in two markets: nontradable goods (internal balance) and tradable goods (external balance). Second, to achieve equilibrium in both markets, two conditions must be satisfied: Expenditure (absorption) must equal income, and the relative price of tradables (the real exchange rate) must be at a level that equates demand and supply in both markets (the slope of P in Figure 15–1). Third, this also suggests two remedies for an economy that is out of balance: A government can achieve equilibrium (stabilize the economy) by adjusting absorption, the nominal exchange rate, or both. Generally, both instruments must be used to achieve internal and external balance.

THE PHASE DIAGRAM

Using the perspective of trade theory, we tie the small, open economy model of macroeconomic management to the tools of analysis already used in this text. But the principles of stabilization can be explored from a more-useful perspective, the **phase diagram**. To develop this approach, consider the markets for tradables and nontradables from the perspective of conventional supply and demand diagrams, as in Figure 15–2.

In these diagrams, we use the real exchange rate, which is the relative price of T goods in terms of N goods (P_t/P_n), as the price in both markets. For tradable goods, that gives a conventional supply and demand diagram: As the price rises, supply increases and demand decreases. But in the nontradables market, a rise in P means a fall in the relative price of N goods, so supply decreases and demand increases. Note that, in both markets, any increase in expenditure, or absorption, A, causes an outward shift of the demand curve: At any price, consumers buy more of both goods.

To use these diagrams as a basis for macroeconomic analysis, we need to change the interpretation of the supply curve for tradables. Until now, we have assumed that all tradables are produced within the home country. But foreign investment and foreign aid can add to the supply of tradables by financing additional imports. Therefore, the supply curve should not be S_t, but $S_t + F$, where F is the inflow of long-term foreign capital in the form of aid, commercial loans, and investment.

Figure 15–2 constitutes a simple model of the small, open economy that is based on two variables: The real exchange rate, P, on the vertical axis and absorption, A, which determines the position of the demand curves. These, of course, are the conventional variables of microeconomics, price and income. But in this model, they also are the two main macroeconomic policy tools of government: The exchange rate and the level of expenditure. Because these two variables are central to macroeconomic management, it would be helpful to develop a diagram that uses them explicitly on the axes.

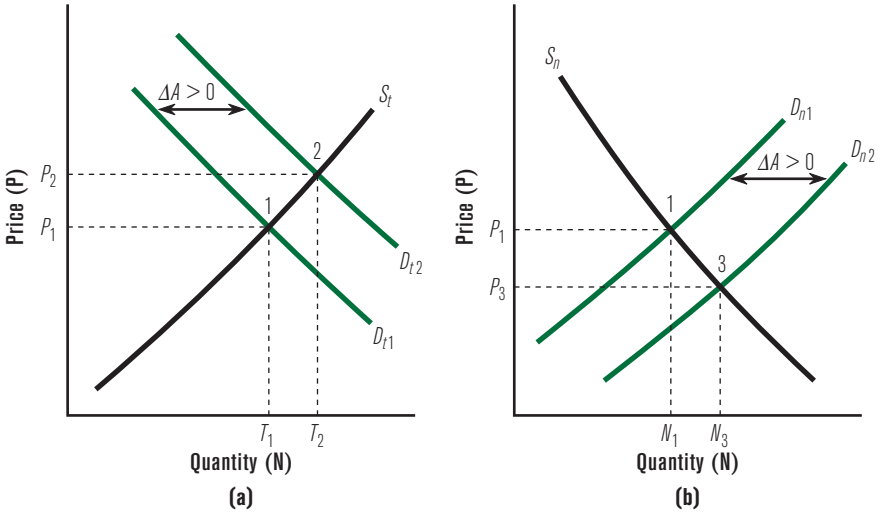


FIGURE 15–2 Tradables and Nontradables Markets

(a) Tradables market: The demand and supply curves for tradables S_t and D_t have the conventional slopes. (b) Nontradables market: The slopes are reversed. S_n falls as P rises (because the relative price of N is falling) and D_n rises as P rises. In both markets, demand increases when absorption (expenditure) increases, shown by and outward shift of D_t and D_n.

S, supply; D, demand; P (price) $= P_t/P_n$.

Figure 15–3 does this. It puts the real exchange rate, $P_t = eP_t^*/P_n$, on the vertical axis and real absorption, A, on the horizontal axis. The diagram also contains two curves, each representing equilibrium in one of the markets. Along the EB, or external balance, curve, the T-goods market is in balance ($S_t = D_t$). Along the IB, or internal balance, curve, the N-goods market is in balance ($S_n = D_n$).

The slopes of the two curves, EB and IB, can be derived from Figure 15–2. In the tradables market, when absorption is A_1, equilibrium is at P_1, where T_1 is produced and consumed. This equilibrium point 1 also is shown in Figure 15–3a. If absorption increases to A_2 in Figure 15–2a, the demand curve moves outward and shifts equilibrium to point 2. Note that with higher absorption, A_2, the real exchange rate, P_2, must be higher to restore equilibrium in the T-goods market. Increased absorption raises the demand for T goods. To meet this demand, it is necessary to raise output, which can be achieved only through a higher relative price of T goods, P_2. This higher price also helps regain balance by reducing the demand for T goods along the new demand curve. Point 2 is transferred to Figure 15–3a at (P_2, A_2).

In the nontradables market, when absorption is A_1, equilibrium is at P_1, where N_1 is produced and consumed. This equilibrium point 1 also is shown in Figure 15–3b. If absorption increases to A_3 in Figure 15–2b, the demand curve moves outward and shifts equilibrium to point 3. In the N-goods market, higher absorption, A_3,

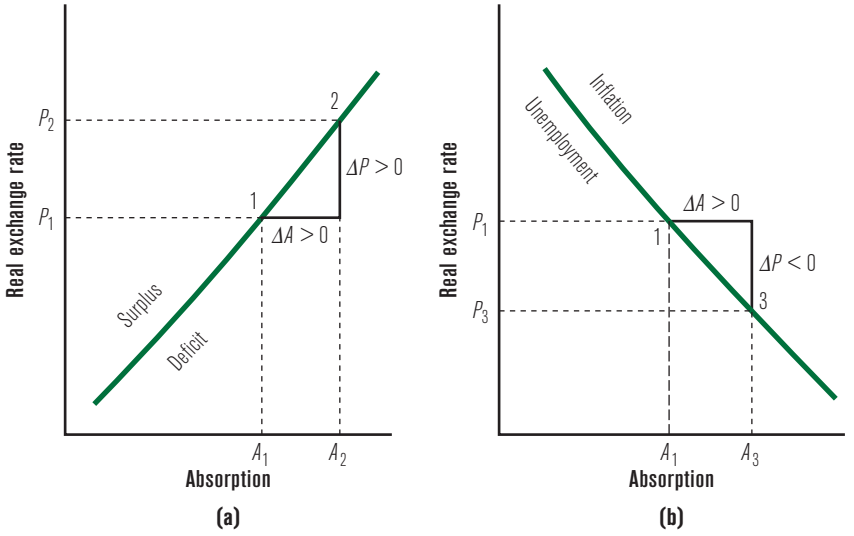


FIGURE 15–3 The Phase Diagram

(a) External balance (EB). (b) Internal balance (IB). The axes are the main policy variables: the real exchange rate, P, and the real absorption, A. The curves show equilibrium in the T-goods market (EB) and N-goods market (IB).

requires a lower, or appreciated, real exchange rate to restore equilibrium. Increased absorption raises demand for N goods, met by raising output, which can be achieved only through a lower relative price of T goods, P_3. This lower real exchange rate, or higher price of N goods, also helps regain balance by reducing the demand for N goods along the new demand curve. Point 3 is transferred to Figure 15–3b at (P_3, A_3).

Figure 15–3 also shows the **zones of imbalance**. In the T-goods market (panel a) for any given level of absorption, say, A_1, any real exchange rate greater than P_1 causes external surplus: The production of tradables exceeds the demand for tradables because the relative price, P, is at a more depreciated level than required for equilibrium. Any real exchange rate below (more appreciated than) P_1 causes an external deficit and the demand exceeds the supply of tradables. Therefore, the zone of surplus is northwest of EB and the zone of deficit is southeast.

In the N-goods market (Figure 15–3b), inflation is to the right of the IB curve, where the demand for N goods exceeds the supply. In that region, for any given real exchange rate, such as P_1, absorption is too high, say, A_3. To the left is the zone of unemployment, where there is an excess supply of N goods. In that region, for any given real exchange rate, say, P_3, absorption is too low, say, A_1.

The meanings of *inflation* and *unemployment* are precise in our model but not in the real world. It is best to think of inflation as being an increase in prices faster than is customary in the country in question. That rate would be quite low in Germany, Japan, or China, probably less than 5 percent a year, but quite high in

Brazil or Argentina. Unemployment implies not only jobless workers but also idle capital and other factors of production. In other words, there is unemployment when an economy is inside the production frontier in Figure 15–1. A country may have high levels of labor unemployment but be unable to increase output because it is fully utilizing its capital or land.

EQUILIBRIUM AND DISEQUILIBRIUM

The two balance curves are put together in Figure 15–4. All along the external balance curve, the demand for T goods equals the supply produced at home plus any net foreign capital inflow. All along the internal balance curve, the demand for N goods equals the supply of N goods. The only point at which there is both internal and external balance (equilibrium in both the T- and N-goods markets) is the intersection of the two curves. This is sometimes called the *bliss point*. It is the same as the tangency of the indifference curve to the production frontier in Figure 15–1 at point 1. The objective of macroeconomic policy is to adjust the exchange rate and absorption to keep an economy stable, in both external and internal balance.

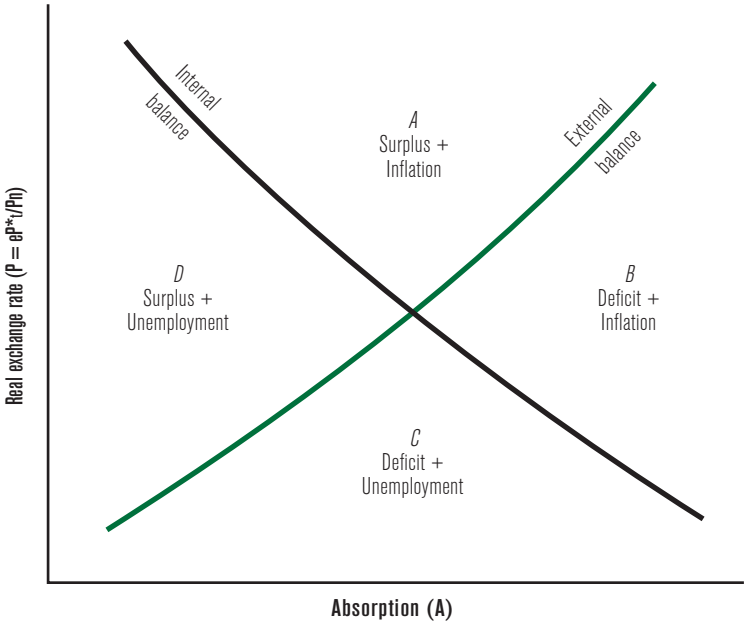


FIGURE 15–4 Zones of Imbalance

The economy is in equilibrium only at the intersection of the external balance (EB) and internal balance (IB) curves. Zones of imbalance are labeled. For example in zone A, the supply of T goods exceeds demand, so there is a surplus, and the demand for N goods exceeds supply, so there is inflation.

Economies spend considerable time in one of the four zones of imbalance shown in Figure 15–4. Zone *A* to the north is a region of external surplus and inflation, where the exchange rate is *undervalued*. In zone *B* to the east of equilibrium, the economy faces inflation and a foreign deficit, due principally to excessive expenditure (absorption is greater than income). To the south is zone *C*, where the exchange rate is *overvalued* (too appreciated) and there is both unemployment and an external deficit. And west of the bliss point the economy is in zone *D*, where, because of insufficient absorption, there is unemployment of all resources but a foreign surplus.

Once in disequilibrium, economies have built-in tendencies to escape back into balance. Figure 15–5 describes them separately for external balance (panel a) and internal balance (panel b). Start with an external surplus, point 1 (Figure 15–5a). The excess supply of tradables generates two self-correcting tendencies. First, the net inflow of foreign exchange adds to international reserves. If the central bank takes no countermeasures, the money supply increases and interest rates fall and induce both consumers and investors to spend more. The increase in absorption moves the economy rightward, back toward external balance. Second, the inflow of foreign exchange creates more demand for the local currency and, if the exchange rate is free to float, forces an appreciation. This is a move downward in the diagram, also toward the EB line. The net result of these two tendencies is the resultant, shown as a solid line in the diagram, heading toward external balance. If, instead, the economy starts in

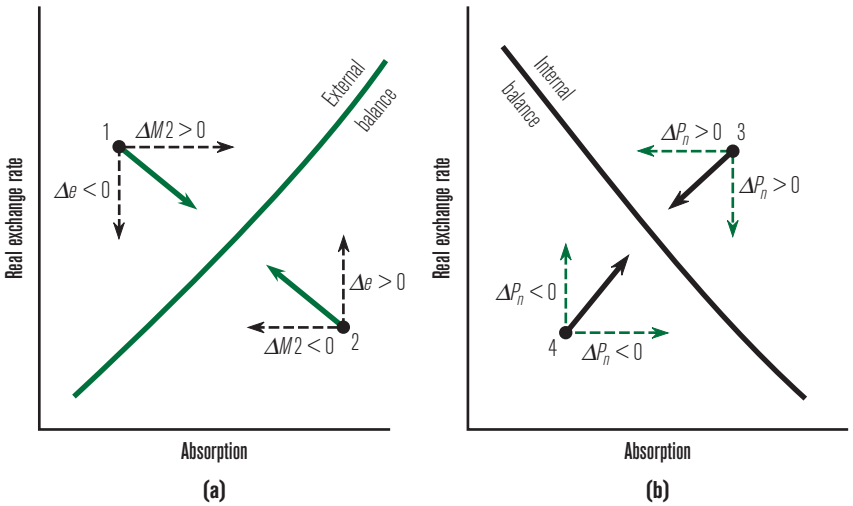


FIGURE 15–5 Tendencies toward Equilibrium

(a) External balance (EB). If the economy faces an external surplus (point 1), reserves and the money supply tend to rise while the exchange rate tends to appreciate; this drives the economy toward EB. (b) Internal balance (IB). For a deficit, if the economy faces inflation (point 3), the rise in prices leads to real appreciation of the exchange rate and a reduction in the real value of absorption; this moves conditions toward IB. Conversely for unemployment at point 4, but only if prices can fall flexibly.

external deficit at point 2, the tendencies are the opposite but the result is the same, a tendency to regain external balance.

The tendency to regain internal balance is shown in Figure 15–5b. When there is inflation (point 3), it affects both the real exchange rate and real absorption. If the nominal exchange rate remains fixed (or is not allowed to depreciate as fast as inflation), the rise in P_n causes a real appreciation. At the same time, the rise in prices can cause a fall in the real value of absorption, assuming that the central bank does not take steps to increase the money supply to compensate for inflation. Under these assumptions, the economy would move from inflation at point 3 back toward internal balance. Unemployment (point 4) would be self-correcting also if prices are able to fall as easily as they rise, but this seldom is the case.

Despite these self-correcting tendencies, in practice, they often fail to work smoothly or quickly enough because of *structural rigidities* in the economy. For instance, exchange-rate changes may take time to affect actual imports and exports, perhaps as long as two years to have a full impact. In economies like Ghana and Zambia, dominated by one or two export products such as cocoa, oil, and copper, with long gestation periods for new investment, supply elasticities for tradables may be especially low, and foreign deficits can persist for a time despite real devaluations.

Nontradables prices probably rise very quickly when demand exceeds supply, as in Figure 15–5b. But in many developing economies, inflation, once started, may resist corrective policies, and prices do not fall so easily when there is unemployment: Unions strike wage bargains that try to maintain real wages by continually raising nominal wages; banks use their market power to keep interest rates high; producers depend on imports, the prices of which are responsive only to exchange rate adjustments; and large firms with monopoly or oligopoly power keep prices up to cover costs that resist downward pressures. Such rigidities have frequently been cited to explain chronic trade deficits and inflation in Latin America, especially in Argentina and Brazil.

However, arguments about structural rigidities can be overstated. There is some flexibility in production for most export industries, even in the short term. And many producer prices are quite flexible, including those of most farm products, those in the large informal sector, and even those of some modern manufacturing firms. Nevertheless, the automatic tendencies toward external and internal balance depicted in Figure 15–5 are likely to be too slow and politically painful to satisfy most governments.

Not all the barriers to adjustment are structural. Sometimes, policies work against adjustment. When foreign reserves fall, for example, the money supply also falls automatically unless the central bank's policy is to *sterilize* these shifts by expanding domestic credit to compensate for the fall in reserves and keep the money supply from falling. Sterilization prevents the move from points 1 or 2 of Figure 15–5a toward external balance. And nominal exchange rates respond to changing market conditions only if the exchange rate is allowed to float or the government makes frequent adjustments in the nominal exchange rate to match changing economic conditions.

However, the opposite policy, a fixed nominal exchange rate, is needed if inflation in nontradables prices is to cause a real exchange-rate appreciation, as depicted at point 3 of Figure 15–5b. This fixed nominal rate is called an exchange rate *anchor* because the fixed rate alone can halt the upward drift of prices as the economy moves due south from point 3. Chile used such an anchor to slow inflation during the late 1970s (Box 15–2). If government devalues the rate to keep up with inflation,

BOX 15–2 PIONEERING STABILIZATION: CHILE, 1973–84

In the last year of the Salvadore Allende regime in Chile, when the public sector deficit soared to 30 percent of the gross domestic product (GDP) and was financed mostly by printing money, inflation exceeded 500 percent a year. In 1973, General Augusto Pinochet overthrew Allende and established an autocratic regime. An early goal of his government was to stabilize the economy. It proved to be a difficult task of many years, with important lessons for later stabilizations in Latin America.

Faced by rapid inflation and unsustainable external deficits, the government imposed a fiscal and monetary shock on the economy. The budget deficit was cut to 10.6 percent of GDP in 1974 and again to 2.7 percent in 1975. Monetary policy was tight: From the second quarter of 1975 through the middle of 1976, it has since been estimated, households and firms were willing to hold more money than was in circulation. But inflation persisted; consumer prices nearly doubled in 1977.

Despite draconian measures, prices continued to rise for two reasons. First, the peso was aggressively devalued to improve the foreign balance, the more so because of the 40 percent fall in copper prices in 1975. In 1977, the peso was worth about one-80th its 1973 value against the dollar. Second, wages in the formal sector were determined by rules that permitted adjustments based on the previous year's rate of inflation, a rule that helped perpetuate the higher rates of earlier years. It also was argued by some that the monetary policy was not stringent enough.

In 1978, the government switched gears and began using the exchange rate as its main anti-inflation weapon. At first a crawling peg was adopted with preannounced rates, the *tablita*, that did not fully adjust to domestic inflation. In 1979, the rate was fixed at 39 pesos to the dollar for three years. The appreciating real exchange rate, or *anchor*, helped control inflation, which was down to 10 percent by 1982. But it also discouraged export growth and contributed to a growing

current-account deficit. At the same time, Chile liberalized its controls over foreign capital flows and attracted large inflows of loans: Net long-term capital rose from negligible amounts before 1978 to average over $2 billion a year in the following five years, equivalent to 8 percent of GDP in 1980. This inflow not only financed the growing current deficit but contributed to the real appreciation of the exchange rate.

Not until after 1984 did Chile finally achieve a semblance of both internal and external balance. It did so through a large real devaluation, approaching 50 percent, supported by tighter fiscal and monetary policies. After a decade and a half of falling income per capita, Chilean incomes grew by 5.8 percent a year from 1985 to 1991.

Source: Based on the account by Vittorio Corbo and Andrés Solimano, "Chile's Experience with Stabilization Revisited," in Michael Bruno et al., eds., *Lessons of Economic Stabilization and Its Aftermath* (Cambridge, MA: MIT Press, 1991).

Brazil's practice for many years, then real appreciation is thwarted and there is no anchor. Similarly, real absorption falls with inflation only if the government fixes its expenditure and its deficit in nominal terms and allows inflation to erode the real value of the expenditure and if the central bank restrains the money supply to grow more slowly than inflation. More typically, the fiscal authorities adjust the expenditure, while the monetary authorities adjust both the money supply and the nominal exchange rate, to fully compensate for inflation. In that case, rising prices have no impact on the real exchange rate or real absorption and an inflationary economy remains at point 3 in Figure 15–5b.

STABILIZATION POLICIES

Whether the barriers to rapid automatic adjustment are inherent in the economic structure or created by policy contradictions, in most cases, governments need to take an active role to stabilize their economies. They have three basic instruments for doing so: exchange-rate management, fiscal policy, and monetary policy.

Alternative **exchange-rate regimes** were introduced in Chapter 12. Governments can vary the exchange rate by having the central bank offer to buy and sell foreign currency at a predetermined or *fixed* official exchange rate (e in our nomenclature) that nevertheless can be changed from time to time or by allowing the rate to *float* in the currency market, although the central bank sometimes may intervene to influence the price. An intermediate case is the *crawling peg*, under which the central

bank determines the rate but changes it frequently, as often as daily, to ensure that the official rate stays in line with domestic and world inflation; this results in a constant or slowly adjusting real exchange rate (P).

Governments have two policies that can influence the level of absorption. Fiscal policy, adjusting levels of government expenditure and taxation, directly affects the government's components of consumption and investment. It also influences private expenditure, especially consumption, which depends on *disposable income*, or income net of taxes. Monetary policy also affects private expenditure. If the central bank acts to increase the money supply, as described in Chapter 12, it increases the liquidity of households and firms, lowers interest rates, and stimulates private consumption and investment.

The power of the phase diagram is that it indicates the necessary directions for these policies, depending on the state of the economy. Figure 15–6 provides such a policy map. It shows the same external and internal balance lines as in the previous diagrams but adds a new element: four policy quadrants, I to IV, within which the policy prescription always is the same.

Take, for example, point 1, which has been placed on the external balance line but in the inflationary zone. For many years, Brazil was in this situation, with buoyant exports and balance in foreign payments but chronic inflation running from 40 to well over 100 percent a year. Because the demand for nontradables exceeds supply, we know that one necessary correction is a reduction in real absorption, monetary and fiscal *austerity*, that would reduce demand and move the economy due west from point 1. But, if that is the only policy taken, the economy would not reach internal balance until point 4, in the zone of external surplus. One imbalance is exchanged for another. To avoid generating a surplus, reduced absorption needs to be accompanied by an appreciation of the exchange rate, a move due south from point 1. The result would be a move approximately toward the equilibrium or bliss point, 0.

Note three things about this result. First, this combination of policies, austerity and appreciation, would work from any point within quadrant I to return the economy to equilibrium. That is, the same combination is needed whether the economy had inflation with a moderate external surplus or inflation with a moderate deficit, either just above or just below the EB line. If the economy starts just below external balance, with a moderate deficit, it may seem strange (*counterintuitive*) to recommend an appreciation that, on its own, would worsen the deficit. But the reduction in absorption, needed to reduce inflation, also reduces the deficit because it also lowers the demand for tradables. Indeed, it reduces the demand for tradables too much and throws the economy into surplus; this is the reason an appreciation is needed. Of course, the relative intensity of each policy is different, depending where in quadrant I the economy starts. But the basic principle holds: Anywhere in quadrant I, the right combination of policies is austerity and appreciation, the combination that moves the economy toward point 0.

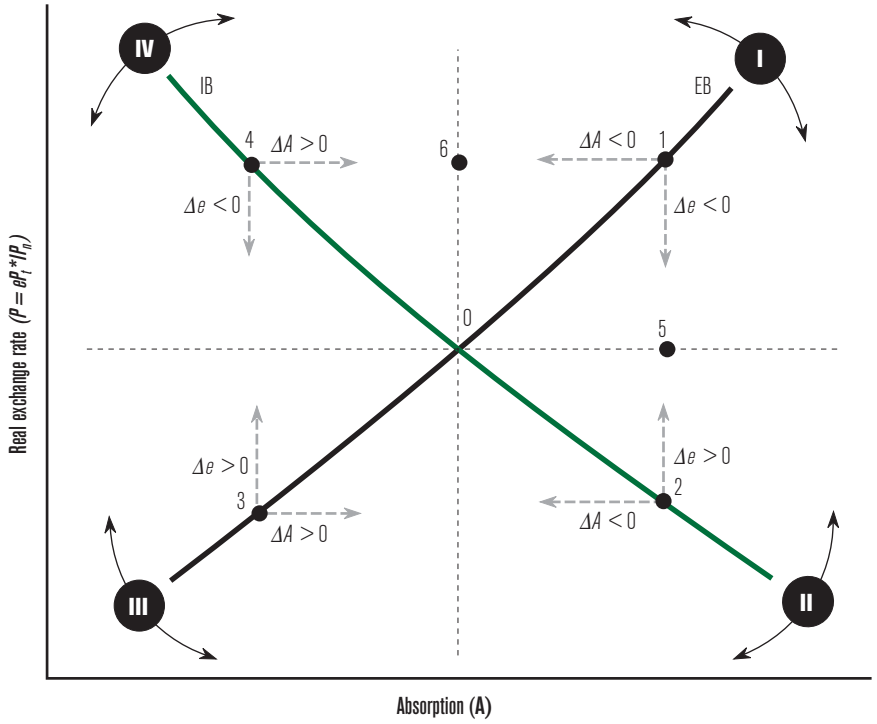


FIGURE 15-6 Policy Zones

From any position of disequilibrium, two policy adjustments generally are needed to restore internal and external balance. In each quadrant (I–IV) a particular combination of exchange rate and absorption policy is prescribed.

Second, in general, two policy adjustments are required to move toward equilibrium. This is a simple example of the general rule enunciated by Dutch economist Jan Tinbergen: To achieve a given number of policy goals, it generally is necessary to employ the same number of policy instruments. Here we have two goals, internal and external balance, and need adjustments in both absorption (austerity) and the real exchange rate (appreciation) to reach them both. It is not always necessary to use two goals, however. If the economy lies just to the east of equilibrium at point 5, then a reduction in absorption achieves internal and external balance simultaneously. And, if the initial situation is point 6, due north of 0, then appreciation alone does the job.

Third, we could view the policy prescription in either of two ways. Austerity is needed to reduce inflation (move west) and appreciation is used to avoid surplus (move south). Or appreciation can be targeted on internal balance (move south toward point 2) but alone would cause a deficit, so that austerity then is required to restore external balance. Therefore, no logic in macroeconomics suggests that one particular policy should be assigned to one particular goal. Economic institutions often do this anyway. In practice, the central bank might use the exchange rate to achieve external

balance while the finance ministry uses the budget for internal balance. But if these two approaches are not coordinated, they may well fail to reach equilibrium.

With these principles established for quadrant I, it is fairly routine to go around the map in Figure 15–6 and see what policy responses are required:

- In quadrant II at a point like 2, with an external deficit but internal balance, exchange-rate devaluation is needed to restore foreign balance but, taken alone, would push the economy into inflation. Fiscal and monetary austerity also are needed to avoid inflation and reach equilibrium. We could reverse this assignment of policies and use austerity to achieve external balance and devaluation to stimulate the economy. Many African countries have been in this situation right up to the present, with low inflation but an insufficiency of export earnings and foreign investment to pay for the imports required for economic development.
- In quadrant III at point 3, an expansionary fiscal or monetary policy eliminates unemployment but at the cost of a foreign deficit, so devaluation is needed to reach equilibrium. Or devaluation stimulates employment and so requires expansion to eliminate the resulting surplus. This is the situation of a mature industrialized economy during a recession, with unemployed labor and capital, but it is not so common in developing countries.
- In quadrant IV at point 4, exchange-rate appreciation can eliminate the external surplus while fiscal expansion prevents unemployment. Or fiscal expansion can end the surplus while appreciation prevents a resulting inflation. A few countries in Asia, such as Taiwan and Malaysia in the 1980s, have been in this situation.

So the principles of macroeconomic stabilization are simple: If policy makers know where to place their economy on this map, they know how to move toward equilibrium. But how do policy makers know where they are? The answer lies partly in measurement, partly in art. Regularly available data on the balance of payments, changes in reserves, and inflation can help locate an economy with respect to the external and internal balance lines. Data on the nominal and real exchange rates, the budget deficit, and the money supply can indicate movements from one policy quadrant to another. Some kinds of data, such as private sector short-term borrowing abroad, however, may not be readily available to policy makers. Such was the case in Korea at the beginning of that country's financial crisis in 1997. In principle, barring such surprises as an unknown large short-term foreign debt that has to be repaid immediately, econometric models can locate the economy and indicate the policies needed to balance it. In practice, especially but not only for developing economies, such models can be too imprecise and too unstable to be wholly dependable. The art of stabilization policy comes in knowing just how hard to push on each component of policy and how long to keep pushing. In this, experience in managing a particular economy is as important a guide as the models estimated by economists.

APPLICATIONS OF THE AUSTRALIAN MODEL

Throughout this book we refer to different kinds of economic problems that are associated with developing countries, including the Dutch disease, debt crises, terms-of-trade shocks, foreign-exchange shortages, destructive inflation, and droughts or other natural catastrophes. The Australian model and its phase diagram can be used to show how these and other shocks affect macroeconomic balance and how they should be handled.

DUTCH DISEASE

In Chapter 18 we will discuss the strange phenomenon of the Dutch disease, in which a country that receives higher export prices or a larger inflow of foreign capital may end up worse off than without the windfall. The Dutch disease was first analyzed by Australian economists Max Corden and Peter Neary, using a version of the open-economy model.[6] Figure 15–7 traces the impact of a windfall gain using the phase diagram. (Box 18–1 provides an alternative exposition of Dutch disease.)

An economy in equilibrium at point 1 suddenly begins to receive higher prices for its major export or is favored by foreign aid donors or foreign investors. All the oil producers, from Saudi Arabia to Indonesia to Mexico, were in this position in the 1970s, as were coffee (and many other commodities) exporters during the boom of the mid-1970s. Egypt and Israel were rewarded with large aid programs by the United States after the Camp David accord of 1978, as was Ghana by the World Bank and others during its stabilization of the 1980s (Box 15–3). Both Chile in the late 1970s and Mexico after its stabilization in the late 1980s received large inflows of private capital, much of it a return of previous flight capital. Foreign exchange windfalls are more frequent than sometimes is supposed. In some cases, these windfalls result from new discoveries of natural resources, such as the major offshore oil reserves discovered by Ghana in 2007.

When the windfall occurs, the supply of tradable goods rises at any given price. This can be shown as a rightward shift in the supply curve in Figure 15–2a. In the phase diagram of Figure 15–7, there is a rightward shift in the EB curve. At point 1, for example, which had been in external equilibrium along EB_1, the economy now is in surplus, so the new EB curve must be to the right—for example, at EB_2. The economy cannot remain at point 1 because the inflow of reserves increases the money supply; this adds to demand and, because the windfall increases private income and government revenue, leads to greater expenditure. So absorption rises, a move

[6] W. Max Corden and J. Peter Neary, "Booming Sector and Deindustrialisation in a Small Open Economy," *Economic Journal* 92 (1982), 825–48.

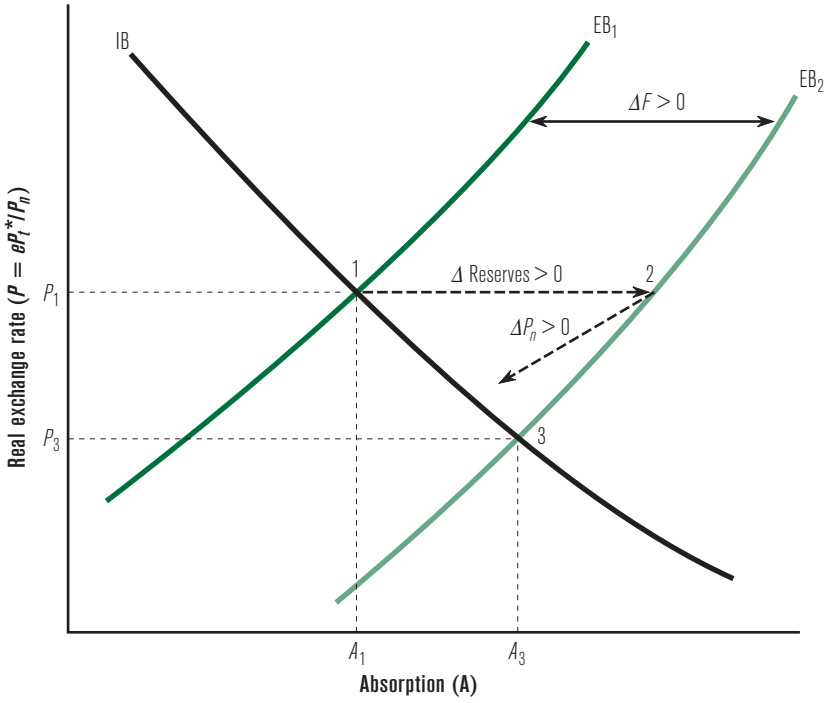


FIGURE 15–7 The Dutch Disease

An export boom or capital inflow shifts the EB curve rightward and leaves the economy at point 1 in surplus. As reserves accumulate and the money supply rises (or as the government and consumers spend the windfall), absorption rises and the economy moves eastward, into inflation. As nontradable prices rise, the real exchange appreciates. At the new equilibrium (point 3), because P is lower, the supply and demand is balanced with less production of T goods and more output of N goods than before. The loss of tradable output is what makes this a disease.

from point 1 toward point 2. This moves the economy off its internal balance, into inflation.[7]

The resulting rise in P_n has two effects: a reduction in real absorption that partially corrects the initial rise in A and, assuming the official rate is fixed, a real appreciation of the exchange rate. (The real rate also appreciates if the nominal rate is floating because the greater supply of foreign currency drives down the price of foreign currency.) Therefore, the economy first moves from point 1 toward point 2 in Figure 15–7, then begins to head in the general direction of the new equilibrium, point 3. In this case, market forces are likely to be sufficient to reach the new equilibrium, unless the

[7]If the windfall is an inflow of capital, this treatment is precise. In the case of a rise in export prices, however, the move from point 1 to point 2 is an approximation. Strictly speaking, a rise in export prices should raise P_t^*, a depreciation of the real exchange rate that moves the economy upward from point 1, after which the economy moves east toward EB_2.

BOX 15-3 RECOVERING FROM MISMANAGEMENT: GHANA, 1983–91

In 1983, after a decade of economic mismanagement, Ghana's gross domestic product (GDP) was 20 percent below its 1974 peak, investment was only 4 percent of GDP, exports had sunk to 6 percent of GDP, and inflation rocketed to 120 percent for the year. After a decade of economic decline, Ghana's military government, headed by Flight Lieutenant Jerry Rawlings, was ready to undertake drastic measures to stabilize the economy and restart economic development.

Working closely with the International Monetary Fund (IMF), Ghana focused on three deep-seated problems: exchange-rate reform, fiscal adjustment, and monetary policy. At first, the government maintained its fixed exchange rate but drastically devalued the cedi from 2.75 to the dollar in 1983 to 90 to the dollar by 1986. In 1986, Ghana adopted a restricted floating currency, using periodic auctions to determine the rate. The official exchange market was broadened in 1988, when many foreign exchange bureaus were authorized to trade currencies and virtually absorbed the parallel market in currency; by 1990, the banks were empowered to trade in an interbank currency market. This completed the move to a floating rate regime. By the end of 1992, the cedi traded at 520 per dollar.

In 1983, with fiscal revenues less than 6 percent of GDP, the urgent need was to restore revenues and control expenditures. The deficit was cut from 6.2 to 2.7 percent of GDP in the first year of austerity, and by 1985, the government had begun a major public investment program to stimulate growth. By 1988, the government had restored total expenditures to 15 percent of GDP, 20 percent of which was investment, and was running a surplus of nearly 4 percent of GDP.

Throughout the period, the money supply was constrained but inflation remained stubbornly above 20 percent a year until 1991, when it was reduced to 16 percent and real interest rates finally became positive. Because food prices play a large role in the consumer price index, investment in food production was seen as an important component of any long-run attack on inflation.

The aid donors responded handsomely to Ghana's stabilization and the accompanying economic reforms: The sum of net official transfers and net long-term capital rose from just over $100 million in 1983 to $585 million in 1991.

Stabilization helped restore economic growth. From the depression of 1983–91, GDP grew by 5.1 percent a year and investment rose to 17 percent of GDP. The improvement, although dramatic in relation to the early 1980s, still left Ghana with a lot to be done: In 1991, income per capita remained 25 percent below its 1973 level.

Source: This account is based on Ishan Kapur et al., *Ghana: Adjustment and Growth, 1983–91* (Washington, DC: International Monetary Fund, 1991).

authorities prevent appreciation and maintain real absorption and so keep the economy in an inflationary posture like point 2.

What, then, is the problem? The economy is at a new equilibrium, its terms of trade improved, its currency appreciated and so citizens have more command over foreign resources, people spending and consuming more without having to work any harder. There are two flaws in this otherwise idyllic picture. First, such windfalls generally are temporary. When export prices fall or the capital inflow dries up, the EB curve shifts back and a costly adjustment is necessary. We analyze that process in the next section.

The second problem is that, in shifting from the old to the new equilibrium, adjustments in the economy must be made. The real exchange rate P is lower, so S_t has fallen, while S_n has risen. Because the booming export sector does not retrench, nonboom tradables bear the brunt of the adjustment. Frictions in the labor market are likely to mean at least temporary unemployment as workers switch from tradable to nontradable production. If the tradable sector includes modern manufacturing, then long-term development may be set back because manufacturing is the sector likely to yield the most rapid productivity growth in the future. And if tradable industries close, it is more difficult to make the inevitable adjustment back toward point 1 when the windfall is over. This decline in nonboom-tradable production turns a foreign exchange windfall into a "disease."

What can be done to cure the disease? The government could try to move the economy back toward the old (and probably future) equilibrium at point 1. Its tools are the official exchange rate, which would have to be devalued against the tendencies of market forces, and expenditure, which would have to be reduced through restrictive fiscal and monetary policies that also reduce inflation (lower P_n or at least its growth). The resulting buildup of reserves and bank balances have to be sterilized through monetary policy so they are held as assets and not spent. It is a neat political trick to manage an austere macroeconomic policy in the face of a boom because all the popular pressures are for more spending. Not too many countries have managed it. Indonesia is among the few that have.

DEBT REPAYMENT CRISIS

When Mexico announced in 1982 that it no longer could service the debt it acquired during the oil boom of the 1970s, many other developing countries followed Mexico's lead, and the financial world entered a decade of debt crisis (Chapter 13). Most Latin American countries largely have overcome their debt problems, but many African countries continue to struggle to repay the money they borrowed, mostly from aid agencies. Although debt service insolvency encroaches gradually on an economy and can be foreseen, it often appears as a national crisis because economic management has been inept.

The formal analysis of a debt crisis is similar to that of another common phenomenon, a **decline in the terms of trade** that leads to a foreign exchange shortage, which in turn is simply the reverse of the Dutch disease. Therefore, the oil exporters, such

as Indonesia, Nigeria, and Venezuela, faced a similar kind of crisis once oil prices began falling in the 1980s. We can understand the similarity between a debt crisis and a decline in the terms of trade more clearly by seeing them in their common context in the balance of payments. A decline in the terms of trade implies deterioration in the balance of trade (in which the excess of imports over exports increases). All else equal, a declining trade balance adds directly to the current account deficit. As detailed in the appendix to this chapter, one of the few ways in which countries can finance current account deficits is by borrowing abroad.[8] Indeed, many developing countries financed chronic current account deficits by borrowing abroad, in the process accumulating enormous stocks of debt (often to levels greater than their GDP). Debt crises ensue (as discussed in Chapter 13) when current account deficits become unsustainable and lenders want to be repaid.

Figure 15–8 captures this process. An economy in balance at point 1 needs to find additional resources to repay its foreign debt or needs to adjust to falling terms of trade. The supply of tradables therefore shifts to the left in Figure 15–2a; in the phase diagram, the EB curve also shifts leftward to EB_2.[9] If the crisis leads to debt relief or additional foreign aid, the curve moves less far and might settle at EB_3.

Now in foreign deficit, the economy begins losing reserves. If the government has to repay some of the debt or falling export prices cut into its revenues, the government needs to reduce its expenditures as well. Both cause a reduction in absorption. These actions move the economy toward external balance but also into unemployment. To gain the new equilibrium at point 3, it is also necessary to devalue the currency. This could be done by the central bank under a fixed rate or by the foreign exchange market under a floating rate. At the new equilibrium, the country produces more and consumes fewer tradables because P has risen. This, of course, is a loss of welfare for the populace. The surplus of S_t over D_t is used to repay the debt or simply compensates for reduced export prices.

Debt crises and the hardships they cause are not an inevitable consequence of borrowing to finance development, as was discussed in Chapter 13. If the borrowed resources are invested productively, they increase the potential output of both tradables and nontradables. Added production increases income and generates the capacity to repay the debt out of additional income, without a crisis and an austerity program. Countries such as Korea and Indonesia have been large international borrowers, but before the financial crisis of 1997–99, they escaped debt crises.

[8]The only other ways (beyond foreign borrowing) to finance a current account deficit are to attract foreign investment and/or to run down the central bank's stock of foreign reserves. See the appendix to this chapter for a concise summary of balance of payments accounting.

[9]Strictly speaking, we cannot analyze the fall in export prices this way, but it is a reasonable approximation for many situations. See note 7.

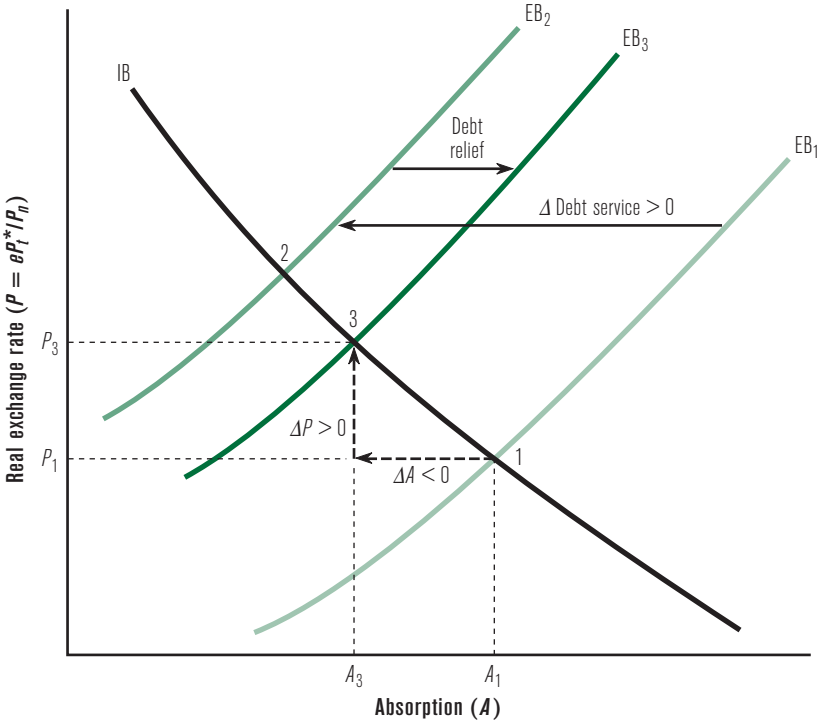


FIGURE 15–8 Debt Crisis or Declining Terms of Trade

An economy in equilibrium at point 1 suddenly needs to repay its debt (or faces falling export prices). External balance shifts from EB_1 to EB_2, although debt relief or increased foreign assistance might reshift the balance line back to EB_3. If policies accommodate the fall in reserves and income, absorption declines. A devaluing exchange rate, via central bank action or market forces, helps the economy move to its new equilibrium at point 3. With more tradables produced and less consumed, the surpluses can be used to repay the debt.

STABILIZATION PACKAGE: INFLATION AND A DEFICIT

External shock is not the only way an economy gets into trouble. Reckless or misguided government policies often are to blame. Impatient with sluggish development or intent on benefiting its constituencies, a government expands its spending and incurs a budget deficit. Unable to finance the deficit by borrowing from the public, the ministry of finance sells short-term bills to the central bank; this adds to the money supply. The economy drifts into inflation and a foreign deficit, at a point like point 1 in Figure 15–9, far from equilibrium at point 2 on the economy's original external balance curve EB_1. When economies become unstable in this way, private investors get skittish and try to invest in nonproductive assets like land or, more often, invest abroad; this deepens the external deficit. The government, recognizing the error of its ways or just hoping for some outside help to avoid painful adjustment, calls in the IMF.

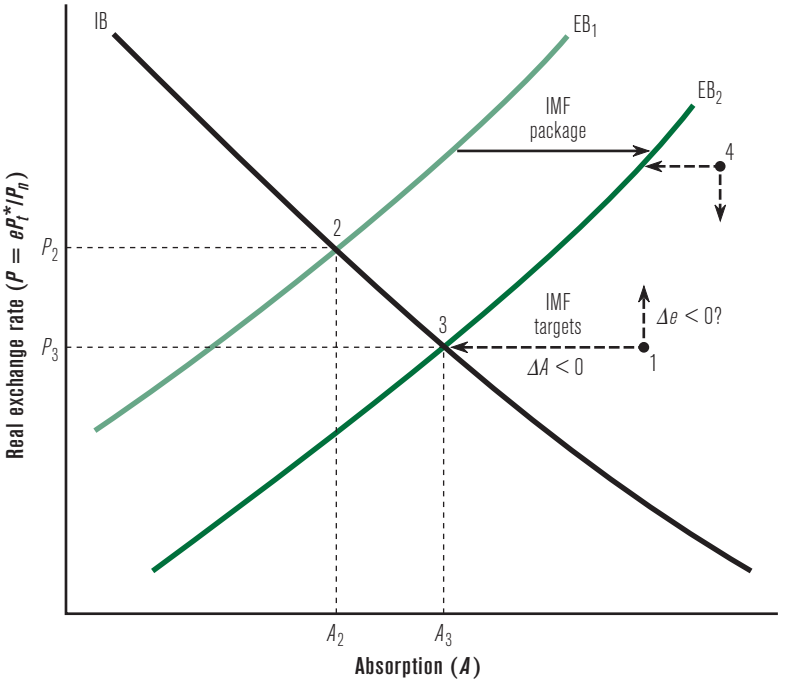


FIGURE 15–9 Stabilization from Inflation and a Deficit

An economy at point 1, far from equilibrium at point 2, above all needs to reduce absorption through austerity: reduced budgetary deficits and slower growth of the money supply. An International Money Fund (IMF) and donor package of aid might bring equilibrium closer by shifting the external balance to EB_2, but the aid package is conditional on the austerity program. Whether any exchange rate action is required depends on the precise initial position, point 1.

The core IMF stabilization program consists of a reduction in the government's budget deficit and programmed targets for domestic credit that, in effect, cap the growth of the money supply. Together, these measures reduce absorption in the economy and move it westward from point 1, closer to external and internal balance. IMF packages frequently include an exchange-rate devaluation as well. Whether this is needed or not depends on the precise location of the economy (point 1) relative to equilibrium (point 2). In some cases, the reduction in absorption is sufficient to reach both internal and external balance. As pictured in Figure 15–9, a small devaluation is needed to reach point 2 and avoid unemployment.

However, IMF programs usually come with substantial aid attached, not only from the fund, but from the World Bank and bilateral donors. The aid package, by adding to the economy's capacity to buy tradables, shifts the EB curve to the right, to EB_2 in the diagram, and moves equilibrium to point 3. Note two things about this aid package. First, it reduces the need for austerity to some extent, as A_3 is greater than A_2. Second, it reduces the need for devaluation of the exchange rate. Indeed, as shown,

there is little or no need to devalue to move from 1 to 3. Donors and the IMF nevertheless frequently insist on devaluation. Sometimes, that may be a requirement just to reach a point like 3. In other cases, donors and the IMF may have in mind a self-sustaining stabilization that will be valid even after aid is reduced and the external balance curve moves back toward EB_1. Whatever the motive, it is important to realize that aid itself is a partial substitute for both devaluation and austerity. In essence, the aid does what higher production of tradables otherwise must do and it finances expenditures that otherwise must be cut. Ghana's experience, which fits this description, is discussed in Box 15–3. More recently, Greece has suffered a major debt crisis, the analysis of which draws on elements of both Figures 15–8 and 15–9 (Box 15–4).

BOX 15-4 THE GREEK DEBT CRISIS OF 2010–12

Problems of unsustainable debt are not limited to the poorest countries. Starting in 2010, Greece faced a debt crisis so severe as to threaten the stability of the European Union (EU). As of September 2011, it remained uncertain whether Greece would be forced to default on its foreign debts or whether other members of the EU would provide its second bailout package for Greece in two years. In May 2010, the EU and the International Monetary Fund (IMF) provided a package of loans and balance of payments support worth €750 billion (approximately $938 billion). In return, the government of Greece committed itself to an austerity program that included severe reductions in expenditures and wages, termination of tens of thousands of government jobs, and extensive tax increases. This agreement sparked widespread civil unrest in Greece, which contributed to a lack of confidence in the country's ability to implement the promised reforms and service its foreign debts.

Greece's debt problem accumulated over many years of current account deficits, the magnitude of which ballooned during the late 2000s. Greece's current account deficits as a share of GDP began mounting in the mid-1990s. Deficits on the order of 3 percent of GDP in the late 1990s more than doubled relative to GDP during 2000–05. By 2005, Greece's current account deficit was equivalent to 7.5 percent of GDP, a proportion that doubled again to nearly 15 percent by 2008. Between 2008 and 2010, Greece succeeded in reducing its current account deficit to 10.5 percent of GDP, mainly by reducing imports. Borrowing by the Greek government to finance its deficits had also grown rapidly, as the government budget deficit as a share of GDP increased from 5.7 percent in 2006 to 15.4 percent in 2009. The government's total debt shot up from €183.2 billion in 2004 (equivalent to 99 percent of GDP, or approximately $228 billion)

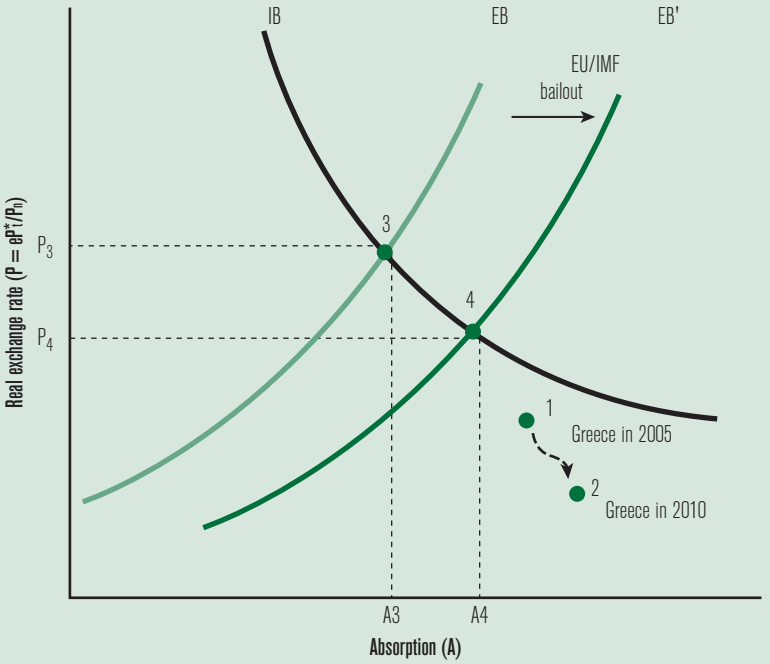
to €330.4 billion in 2010 (equivalent to 144 percent of GDP, or approximately $438 billion). The country's general economic stagnation was further reflected in its unemployment rate, which after fluctuating around 10 percent for a decade, fell to 7.5 percent in May 2008, only to double by March 2011. Greece's GDP shrank by 6.6 percent in 2010.

By the fall of 2011, the government of Greece found itself between a rock and a hard place, as the EU debt relief package was imperiled by a threatened cutoff in response to Greece's apparent inability to meet its austerity commitments. Yet daily strikes and mounting political pressure within Greece were preventing the government (led by a socialist party, with a slim parliamentary majority) from fully implementing the promised cuts. A disorderly default was a distinct possibility, along with an exit by Greece from the euro zone.

The Australian model lends itself well to a depiction of the Greek debt crisis. The analysis shown in the figure combines elements of Figures 15–8 and 15–9. By 2005, Greece was suffering from both a balance of payments deficit and unemployment, conditions indicated by point 1. By 2010, both unemployment and the balance of payments deficit had substantially worsened, moving Greece toward point 2 (farther from both internal and external balance and far from equilibrium at point 3). It seems straightforward, based on the phase diagram, that Greece needed to reduce absorption and depreciate its real exchange rate, but there's a catch. As a member of the EU, Greece was a member of the euro zone and was thus unable to devalue its currency as a means of reducing its balance of payments deficit. The only way for Greece to induce the necessary real depreciation was to reduce P_n sufficiently. In short, austerity presented itself as virtually the only tool available to the Greek government. Pushing down wages—a central element of the Government's austerity program—would reduce P_n (because labor can be counted as a key nontradable) and help depreciate the real exchange rate. Austerity was also critical to restoring external balance because borrowing to finance the current account deficit was no longer an option for Greece.

The bailout package offered by EU member states in concert with the IMF would effectively shift the EB curve out to *EB′*, thus reducing the amount of both real depreciation and spending cuts required to reach the new equilibrium at point 4. Herein lay the impasse encountered in the fall of 2011. Greece was widely seen as incapable of making the cuts necessary to reach equilibrium at point 3. Without the bailout package, default seemed unavoidable. Yet even the degree of austerity required to reach the postbailout equilibrium at point 4 was becoming politically impossible for the Greek government. Absent a credible commitment by the Greek government to impose the necessary spending and public employment cuts (along with substantial tax increases), the EU was

unwilling to provide the bailout. The brinksmanship between Greece's government and its debtors continued into early 2012. Greece's next round of debt repayments were due in on March 20, 2012, and, absent a new bailout agreement, default was inevitable. In late February, just weeks before the prospective default, EU and Greek negotiators agreed on a second bailout program worth €130 billion ($169 billion) in return for renewed promises of severe austerity and greater EU influence over Greek budgets. The deal was intended to keep the Greek government solvent until 2014. Yet, Greece's continuing deep recession and political uncertainty suggested that the crisis was far from over.



Another kind of stabilization also can be illustrated with Figure 15–9, **rapid inflation** (or hyperinflation). In Bolivia's hyperinflation of the mid-1980s or the chronic inflations in the past in Brazil and Argentina, external balance is a secondary consideration or not a major problem. Point 4 in the diagram depicts this situation. Austerity still is required to move toward equilibrium at point 2 or 3 (if there is an aid package), but devaluation only intensifies inflation. Instead, the currency must be appreciated, which also dampens inflation. One way to achieve this would be to fix the nominal rate and let the continuing (if decreasing) inflation in nontradable prices (P_n) appreciate the real rate *P*. This is the *exchange-rate anchor*, a device used often in Latin America, especially in Chile during the late 1970s, in Bolivia during the

mid-1980s, and in Argentina during the 1990s. It has the disadvantage that a lower real rate discourages export growth. Yet investment in new exports may be part of a strategy to open the economy, diversify exports, and move the external balance curve to the right.

DROUGHT, HURRICANES, AND EARTHQUAKES

The human tragedy of drought, earthquakes, and other natural disasters in places such as Ethiopia, the West African Sahel, and Haiti in 2010 dwarfs issues of macroeconomic management. But the adept management of an economy racked by natural disaster is essential to reduce the misery of starving or displaced people. Drought, for example, reduces a country's capacity to produce food, export crops, and in some countries, generate electricity from hydropower. At the same time, income is lower because farmers and others have less product to sell. Government then needs to provide social safety nets; this means spending more on the provision of food, transportation, health services, and sometimes shelter. Foreign governments often provide financial, food, and technical aid under these situations.

The macroeconomic reflection of a drought or earthquake is depicted in Figure 15–10. The economy begins in equilibrium at point 1. Drought or earthquake reduces the economy's capacity to produce both nontradables (some foods, hydroelectricity, water supplies) and tradables (export crops, importable foods, some manufactures). We show this as a leftward shift in both the IB and EB lines: Reduced output of S_n at any given price means a larger zone in which D_n exceeds S_n; this is inflationary. Similarly for S_t, and this enlarges the area of deficit. The new external balance curve, EB_2, may be augmented (shifted back to the right) by foreign aid to EB_3, in which case, the new equilibrium is point 3.

The economy, still at point 1, is inflationary. The fall in incomes creates a tendency for absorption to shrink on its own and move the economy leftward toward the new equilibrium. At the same time, the government tries to spend more to relieve hunger, disease, and other problems. The outcome depends on the relative force of these tendencies. The impact of most natural disasters is temporary, typically lasting a year, although some African droughts have been much longer, and the Haiti earthquake of 2010 is likely to last for several years or more. It is appropriate to try to ride out such shocks with minimal adjustment, especially if foreign aid can bear much of the burden. Therefore, for example, even if an exchange-rate adjustment is called for to reach equilibrium, it is unlikely to work very well during the natural disaster and probably should be resisted. This could be said for fiscal austerity too, except that the rise in prices can deepen the suffering of those already hurt by the disaster. If the government is able to shift its expenditures so that a greater portion goes into alleviating the impact of the disaster, it may be able to relieve the worst suffering while restricting the rise of total expenditures and containing inflation.

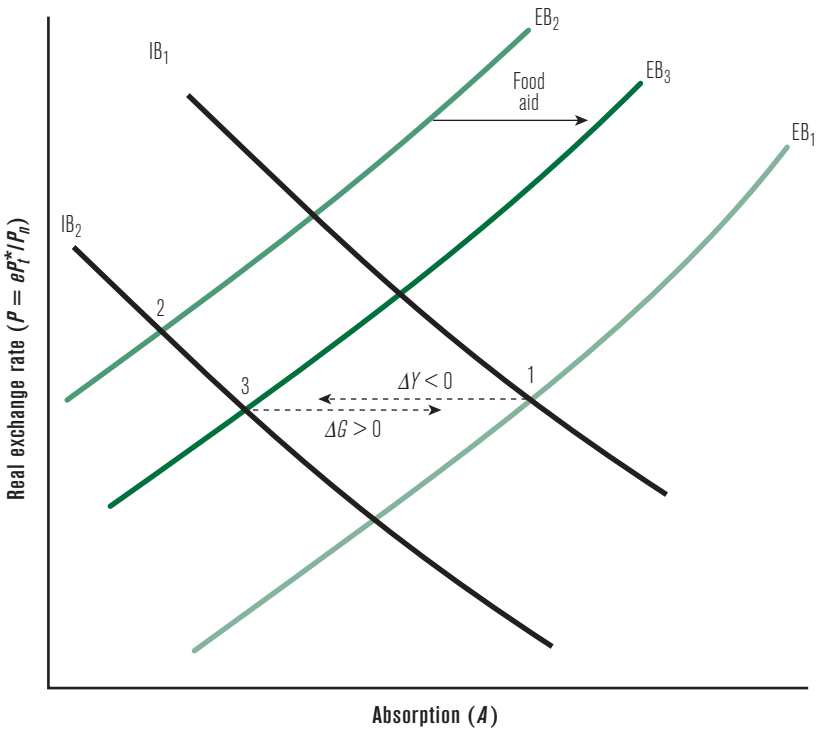


FIGURE 15–10 Drought

Drought or another natural disaster reduces the capacity to produce both nontradables and tradables, so the curves shift to the west. Disaster relief from abroad augments the external balance curve and shifts it to EB_3, with equilibrium at point 3. Remaining temporarily at point 1, the economy becomes inflationary. The reduction in output and therefore in income reduces absorption, but the government's need to spend more on relief tends to offset this move toward equilibrium. The outcome could be continued inflation.

SUMMARY

- Open economies may be exposed to a variety of shocks, both positive and negative, either of which may require policy responses to balance potential negative effects. In many cases, we have seen that a country's vulnerability to external shocks may be increased by its own history of poor policy choices.
- One common problem among developing countries arises from extended periods of spending more than they earn and financing the resulting current account deficits with foreign borrowing. For many developing countries, these persistent current account deficits become unsustainable, typically requiring both spending reductions and currency devaluation.

- The Australian model provides a convenient and versatile framework within which to diagnose common macroeconomic imbalances. Equilibrium in the Australian model requires balance in the markets for both tradable and nontradable commodities (the relative price of which is one expression of the real exchange rate). Equilibrium in the tradables markets is known as external balance; equilibrium in the nontradables market is known as internal balance.
- Disequilibrium in the tradables market implies either a balance of payments deficit (in the case of excess demand) or a balance of payments surplus (in the case of excess supply). Disequilibrium in the nontradables market implies either a higher rate of inflation (in the case of excess demand) or higher unemployment (in the case of excess supply).
- Considered together, these disequilibrium conditions may arise in four combinations: balance of payments deficit with inflation, balance of payments deficit with unemployment, balance of payments surplus with inflation, and balance of payments surplus with unemployment.
- The main policy tools available to regain macroeconomic equilibrium are fiscal and monetary policy interventions to change the level of domestic spending (absorption) and changes in nominal exchange rate policy aimed at altering the real exchange rate.
- Recommendations to use some combination of these policy tools to regain equilibrium must take account of the economy's natural tendencies toward equilibrium. The functionality of these natural tendencies will vary with each country's economic and policy circumstances.

APPENDIX TO CHAPTER 15: NATIONAL INCOME AND THE BALANCE OF PAYMENTS

Many macroeconomic problems and crises in developing countries have their roots in the balance of payments. Chapter 10 opens the discussion of balance of payments by describing the role of foreign savings in financing investment. Chapter 13 discusses debt crises in developing countries, and Chapter 15 provides a model for diagnosing and responding to macroeconomic crises—including debt crises. Balance of payments accounting provides a framework within which to understand many of the opportunities and risks created by opening economies to trade and financial flows.

This appendix provides a brief overview of some key aspects of balance of payments accounting.[10]

We begin by placing the balance of payments in the broader context of national income accounting. The value added from all goods and services produced domestically sums to Gross Domestic Product (GDP). In a closed economy, this exactly equals gross national expenditure, which consists of private consumption on final goods and services (C), government consumption and investment in final goods and services (G), and private investments to expand the country's capital stock (I). That is, GDP = C + I + G in a closed economy. In Chapter 15, we refer to gross national expenditure as **absorption** (A), or A = C + I + G. Adopting this notation, we can say that in a closed economy, GDP = A. In addition, in a closed economy all of gross national expenditures (or absorption) is paid to domestic entities, so GDP also equals gross national income (GNI).

When we allow for trade with other countries, it is no longer required that gross domestic product exactly equals absorption or gross national income. This is where the balance of payments comes in. Balance of payments accounting records all of the (home) country's transactions with the rest of the world, including both trade and financial flows. The main components of the balance of payments are the **current account**, the **financial account**, and the **capital account**.

We begin by incorporating the current account into national income. The current account records all foreign trade in goods (the trade balance) and services (net factor income from abroad), along with net unilateral transfers (such as foreign aid or remittances from residents working abroad). The current account also includes interest payments on foreign debts (debt service payments, but not repayment of principal on the debt). In an open economy, part of absorption may be spent on imports (M), and must be subtracted from A; conversely, some domestic production may be exported abroad (X), and the resulting payments to domestic firms are added to A. In this case, we can write the national income accounting identity as

$$GDP = C + I + G + X - M \qquad \text{[A15–1]}$$

Substituting absorption (A) in for gross national expenditure, we can re-write equation A15–1 as

$$GDP = A + X - M \qquad \text{[A15–2]}$$

[10]This discussion draws on Robert C. Feenstra and Alan M. Taylor, *International Macroeconomics*, 2nd edition (New York: Worth Publishers, 2012), Chapter 5 (to which the reader is referred for a more detailed treatment), as well as the IMF's *Balance of Payments and International Investment Position Manual*, 6th edition (Washington, DC: International Monetary Fund, 2011).

That is, gross domestic product equals domestic expenditure plus net exports. Rearranging the terms of equation A15–2, we can see that A − GDP = M − X. In other words, if domestic spending exceeds GDP, then our imports exceed our exports and we run a negative trade balance. The trade balance is typically the major component of the current account. As noted above, net factor income from abroad (NFIA), or income from the net export of services (e.g., wages paid to a home country resident working abroad for a foreign firm), is also part of the current account and is included in gross national income (GNI). Thus, GNI = GDP + NFIA. The difference between gross national income (GNI) and gross national disposable income (Y) is that the latter also includes the final component of the current account, net unilateral transfers (NUT). Putting these steps together, we can write

$$Y = A + (X - M) + NFIA + NUT = A + CA \qquad \text{[A15–3]}$$

where CA is the current account, which equals the sum of net exports, net factor income from abroad, and net unilateral transfers. Equation A15–3 is important. It implies that if the home country's total expenditures exceed its total receipts, the difference must be provided by running a deficit on the current account. From equation A15–3, we can see that $A - Y = -CA$. The current account balance thus serves as an indicator of whether a country's spending exceeds its income.

As discussed in Chapter 10, we can also approach the current account from the perspective of national savings. National savings is what is left over from total receipts net of private and government consumption, or $S = Y - C - G$. If we expand A in equation A15–3 to write $Y = C + I + G + CA$ and subtract $C + G$ from both sides, we arrive at the **current account identity**,

$$S = I + CA \qquad \text{[A15–4]}$$

which says that we can only save more than we invest if there is a CA surplus, and we can only invest more than we save if there is a CA deficit.

Domestic savings has two components—private savings (S_p) and government savings (S_g). Income in the private sector must be consumed, saved, or paid to the government as taxes. Thus,

$$S_p = Y - C - T \qquad \text{[A15–5]}$$

Similarly, government savings is the difference between the tax revenue collected by the government (T) and its expenditures (G), or

$$S_g = T - G \qquad \text{[A15–6]}$$

which is the government's fiscal surplus (or deficit, if $G > T$). The sum of these two equations equals our previous observation that $S = Y - C - G = S_p + S_g$, or total national savings is the sum of private savings and government savings. Substituting this idea into equation A15–4 and rearranging, we can see that

$$CA = S_p + S_g - I \qquad \text{[A15–7]}$$

Stated slightly differently, if investment is to exceed national savings, the difference must be financed from abroad, and that difference equals the current account deficit. In this sense, we can think of the current account as foreign savings. In addition, if we substitute equation A15–6 into equation A15–7, we see that

$$CA = (S_p - I) + (T - G) \qquad \text{[A15–8]}$$

which tells us that if investment exceeds private savings and/or the government runs a fiscal deficit, the current account must be in deficit by an equivalent amount.

The rest of the balance of payments story is about how the current account is financed. As noted above, in addition to the current account, the balance of payments includes the financial account (FA) and the capital account (KA). The financial account records foreign transactions involving financial assets such as stocks, bonds, and real estate ownership. The major components of the financial account (using the IMF definition) are foreign direct investment, portfolio investment, and (importantly) reserve assets (which is the stock of foreign currency—usually U.S. dollars, euros, or yen) held by the central bank. The capital account refers to the transfer of assets such as gifts to and from abroad.[11]

A fundamental principle of balance of payments accounting is that bills must be paid. This implies that the balance of payments must sum to zero, or

$$CA + FA + KA = 0 \qquad \text{[A15–9]}$$

This is the **balance of payments identity**. It implies that there are only three ways in which a country can finance a current account deficit: 1) it can borrow from abroad (from foreign governments, banks, or international lending bodies), 2) it can attract capital inflows (in the form of direct or portfolio investment), or 3) it can run down its stock of foreign reserves. If the current account is in deficit, then it must be balanced by a surplus in the financial and capital accounts (which potentially includes running down the country's stock of foreign reserves). If the balance of payments must also equal zero, what is meant when we hear about a "balance of payments deficit?" In common use, a balance of payments deficit refers to a situation in which a CA deficit is not fully balanced by a surplus in FA + KA *where FA is defined narrowly to exclude the foreign reserve assets of the central bank*. In such a case, the total balance of payments must be brought to zero by a change in foreign reserves (running them down, in this instance, to finance the remaining CA deficit).

Running a surplus in the (non-reserve) financial and capital accounts means that the home country is either borrowing from abroad or selling domestic assets to foreigners. In either case, the home country is increasing the claims by foreigners

[11]This narrow definition of the capital account (and broad definition of the financial account) is in accordance with the approach now taken by the IMF, along with OECD and the UN System of National Accounts. Previously, the accepted tradition was to define the capital account broadly enough to include everything that the IMF now assigns to the financial account, with the exception of the reserve assets of the central bank.

on domestic residents. Conversely, if the home country is running a current account surplus, it must also be running a deficit on the financial and capital accounts, thus increasing the claims by residents on foreigners. Since the home country's stock of net foreign assets equals the difference between claims by residents of the home country on foreigners and claims by foreigners on home country residents, the current account equals the change in net foreign assets.

PART FOUR

Agriculture, Trade, and Sustainability

16 Agriculture and Development

During the years 2006 through 2008 the world experienced its worst food crisis since the early 1970s. Over that period, the price of corn on world markets nearly tripled, followed closely by the prices of other basic cereals. During the six months from October 2007 to April 2008 the world price of rice tripled, increasing from $335 to over $1,000 per ton. The results were immediate and sometimes violent. Food riots broke out in Bangladesh, Burkina Faso, Cameroon, Egypt, Haiti, Mozambique, and Senegal. A report from Cairo, Egypt, by Al Jazeera on March 13, 2008, recounted, "thousands of people have resorted to violence due to shortages of basic food commodities and rising food prices. At least 10 people have died . . . in riots that erupted at government subsidized bakeries." A few weeks later, on April 18, the *New York Times* reported from Port-au-Prince, Haiti, that, "Hunger bashed in the front gate of Haiti's presidential palace. Hunger poured onto the streets, burning tires and taking on soldiers and the police. Hunger sent the country's prime minister packing. Haiti's hunger, that burns in the belly that so many here feel, has become fiercer than ever in recent days as global food prices spiral out of reach, spiking as much as 45 percent since the end of 2006 and turning Haitian staples like beans, corn, and rice into closely guarded treasures." Grain prices fell in late 2008, but once again rose steeply in 2010–11.

When food prices rise dramatically, what does this imply for developing countries and the welfare of the poor? If many of the poor are farmers, do they benefit from the kinds of food price increases that have occurred in recent years? Some may benefit, if they produce and sell food in sufficient quantities; yet, most poor households are net consumers of food. Similarly, most poor countries are net importers

of food. Increases in world food prices may be quite harmful. According to the Food and Agricultural Organization (FAO) of the United Nations, high food prices in 2008 pushed an additional 40 million people into hunger, raising the total to an estimated 963 million. In addition, the FAO estimates that the food import bill for developing countries increased by nearly 75 percent in 2008 as a consequence of the food price shock.[1] The implications of such events for developing countries are complex and raise central questions about the role of agriculture in both economic growth and poverty alleviation.

Agriculture has substantial implications at both the macroeconomic and microeconomic levels in most developing countries. Among countries categorized by the World Bank as low income, agriculture accounted for 25 percent of the gross domestic product (GDP) in 2008, making it the largest single sector in many countries. In addition, over 70 percent of the population of these low-income countries lived in rural areas. Even though not all rural households earn their primary income through farming, the simple observation that nearly three-quarters of the population of low-income countries was sharing one-quarter of the income suggests that poverty in the poorest countries tends to be disproportionately concentrated in rural areas. The large size of the agriculture sector, both in its share of GDP and employment, along with the likely concentration of poverty in rural areas, points to the unique opportunities that agricultural development provides. This chapter addresses the role of agriculture in economic growth and poverty alleviation; Chapter 17 looks more closely at the specific policies and institutions that governments can use to maximize those contributions.

UNIQUE CHARACTERISTICS OF THE AGRICULTURAL SECTOR

Economists Peter Timmer, Walter Falcon, and Scott Pearson identified five characteristics of the agriculture sector that distinguish it from other sectors of most developing economies: the agricultural sector's share of GDP, the agricultural sector's share of the labor force, special characteristics of the agricultural production function, that much of the agricultural sector's output is directly consumed by its producers, and agriculture's role as a resource reservoir.[2]

Agriculture is a dominant sector in many of the world's poorest countries. Table 16–1a summarizes trends in the agricultural share of GDP for major areas of

[1]FAO, *Crop Prospects and Food Situation*, 2008, www.fao.org/giews/, and A. Mittal, "The 2008 Food Price Crisis: Rethinking Food Security Policies," G-24 Discussion Paper No. 56, United Nations Conference on Trade and Development (UNCTAD), Geneva, June 2009.

[2]Peter Timmer, Walter Falcon, and Scott Pearson, *Food Policy Analysis* (Baltimore, MD: Johns Hopkins University Press, for the World Bank, 1983).

TABLE 16–1 The Share of Agriculture in GDP and Rural Population Share

	1965–75	1976–85	1986–95	1996–2008
(a) Share of agriculture in GDP (%)				
East Asia and Pacific	35.3	29.3	22.8	14.4
Latin America and Caribbean	13.3	10.8	8.9	6.4
South Asia	41.3	34.0	28.9	22.3
Sub-Saharan Africa	20.4	18.8	18.5	17.1
High-income countries	5.7	4.1	2.7	1.8
(b) Rural population share (%)				
East Asia and Pacific	81.1	78.0	70.8	61.2
Latin America and Caribbean	42.5	34.8	29.1	23.8
South Asia	81.1	77.8	75.0	72.2
Sub-Saharan Africa	80.4	76.1	71.6	66.4
High-income countries	33.1	29.4	26.6	23.7

GDP, gross domestic product.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

the developing world since 1965. While it is clear that agriculture plays an important role in generating national income, it is also clear that the magnitude of that role has trended downward over time (an observation pursued in greater depth in the following section). Nonetheless, the challenge of accelerating economic growth becomes much more difficult if a large sector such as agriculture is left to lag behind the rest of the economy.

Table 16–1b demonstrates the even greater share of each region's population that lives in rural areas. Not all rural dwellers are farmers, and many farmers earn at least some of their income outside of farming. Yet, most nonfarm rural activities depend substantially on the existence of a vibrant agricultural sector. Agriculture also plays a substantial role in the consumption side of the economy. It is common among poor households in developing countries for food expenditures to make up 50 to 70 percent of total household expenditures.

The agricultural sector is also distinguished from other sectors by both the sheer number of participants and by the degree of decentralization of those participants. A farming sector may consist of hundreds of thousands, or even millions, of individual production units, all operating independently, yet all allocating resources in response to the same broad set of incentives created by government policy.

Key characteristics of the agricultural production function are also distinctive. These features include seasonality, geographic dispersion, the sources of risk, and the sources of technical change. Agricultural production is uniquely sensitive to seasonality. Most countries have distinct growing seasons, usually defined by rainfall patterns. From an economic point of view, we can think of farmers' use of seasonality as cost minimization (for example, it is much cheaper to use natural sunlight

than to use electric grow lights). Yet, with the reliance on seasonality also comes strict requirements for the timing of operations. Late planting, for instance, might substantially reduce yields by failing to take advantage of the best natural growing conditions. Strict timing requirements imposed by seasonality may lead to labor shortages at critical times, such as planting and harvesting, despite widespread rural underemployment at other times of the year. If inputs such as seeds and fertilizer are not available at the right time, they may become useless. These considerations, plus the fact that agriculture is highly geographically dispersed due to its unique use of surface area as an input to production, underscore the need for efficient marketing of both inputs and outputs as a critical ingredient for agricultural development (a point discussed in more detail in Chapter 17).

Risk and uncertainty also affect production in ways that are unique to agriculture. The primary sources of risk for farmers are prices and weather. While it is true that all productive sectors must worry about the price of their output, agricultural prices tend to be particularly variable, both within and across seasons. Farmers are confronted with challenging forms of price risk, as they must commit resources (and potentially take on debt) at the time of planting despite not knowing the price they will receive for their output until months later, when they harvest their crops. This uncertainty about prices is related to weather being an additional source of risk that is unique to agriculture. Most poor farmers (especially outside of Asia) do not have access to modern irrigation systems and thus depend on rainfall. In sub-Saharan Africa only 4 percent of cropland is irrigated and in drought-prone regions, such as Africa's Sahel, weather-related risk is substantial. Too much rain can also be catastrophic, as the destruction of nearly one-fifth of Pakistan's cultivable area by the monsoon floods of 2010 demonstrated.

Productivity growth is essential for sustained agricultural development and is a critical component of overall economic development. In the long run, growth in agricultural productivity is driven by technical change. Creating and disseminating technical change in agriculture is a special challenge for policy makers. The diversity and complexity of agricultural production systems around the developing world make agricultural technologies highly context specific. A wheat variety that thrives in the Pakistani Punjab may fail miserably in Argentina. Thus substantial effort and resources are required not only to create agricultural technologies but also to adapt them to particular agroecological settings. The highly atomized and dispersed nature of agriculture in most countries then presents additional challenges for disseminating new technologies, the impact of which are nil if they are not widely adopted by farmers.

Agriculture is also unique in that producers in developing countries consume a large share of their own output. This highlights the complexity of decision making within agricultural households. Production and consumption decisions are inextricably linked. Households must consider the trade-offs inherent in allocating a mother's time to working in the field or to childcare or in deciding how much time to send

children to school versus putting them to work (either on the family's own farm or as hired off-farm labor). Such decisions may have large impacts on the welfare of poor households.

Finally, agriculture differs from other sectors in having traditionally been seen largely as a reservoir of inexpensive resources (principally labor and capital) available for extraction and use in modern industry and services. In many cases, this perspective justified the neglect of agriculture by policy makers, as agriculture was perceived to be a traditional, low-productivity sector with little to contribute to industrialization. Although later approaches have challenged this view of agriculture's role in economic growth, as economies grow, agriculture tends to account for a declining share of both GDP and employment. This robust empirical regularity is part of the broader concept called **structural transformation**.

STRUCTURAL TRANSFORMATION

Structural transformation refers to the systematic changes in sector proportions as economies grow. Initially, at low levels of income, agriculture dominates both as a share of GDP and as a share of employment. As economies transition toward middle levels of income per capita, agriculture accounts for a smaller share of both GDP and employment (replaced by industry and services). This pattern is a central feature of economic development and both a cause and an effect of economic growth. In advanced economies, the farming population accounts for a negligible share of the total population. As Timmer observes, "there are more lawyers in the United States than farmers, more dry cleaning establishments than farms."[3] The relative decline of agriculture and rise of industry and services that characterize structural transformation imply an additional characteristic—namely, a substantial migration of labor from the rural to the urban economy. Finally, structural transformation is also characterized by a demographic transition (discussed in Chapter 7) in which the high birth and death rates of traditional societies are replaced by low birth and death rates (as health conditions improve with income growth).

From the perspective of the demand side of the economy, this process of structural transformation is driven by **Engel's law**, the observation by nineteenth-century statistician Ernst Engel that the proportion of income spent on food declines as income rises (that is, the income elasticity of food is positive, but less than 1). This implies that income grows faster than the demand for food, resulting in the decline of agriculture as a share of national income.

[3]C. Peter Timmer, "A World without Agriculture: The Structural Transformation in Historical Perspective," Wendt Lecture, American Enterprise Institute, Washington, DC, October 30, 2007, p. 5.

Structural transformation is an empirically robust phenomenon. Figure 16–1 shows the (smoothed) patterns of rural population shares and agricultural shares of GDP by region as a function of income per capita. (The same patterns are apparent in Table 16–1, which includes the average levels of these variables by decade.) It is clear from Figure 16–1 that there is a persistent, though declining, gap between the rural population share and the share of agriculture in GDP. This gap reflects the persistent, though declining, concentration of poverty in the rural economy. However, as the gap falls, labor productivity and wages across sectors tend to converge. The data in Table 16–1 show that significant gaps remain, on average, in the major regions of the developing world. Among those regions, only Latin America has approached low levels (and convergence) of rural population share and agricultural share of GDP that characterize the advanced economies. The particular path followed by any given country will vary from these norms, depending on local circumstances. Yet, the broad pattern is remarkably robust.

We can begin to consider the potential contributions of agriculture to economic growth by looking more closely at the process of structural transformation. History reflects that structural transformations broadly consist of four distinct phases.[4] The first phase begins with an increase in agricultural output per worker (average labor

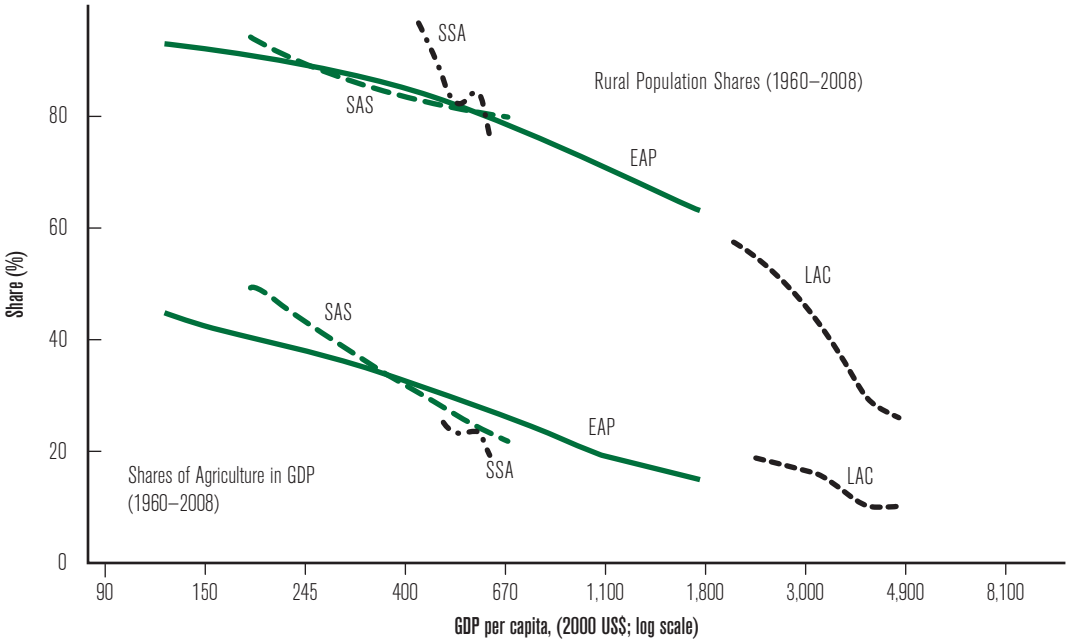


FIGURE 16–1 The Structural Transformation

EAP, East Asian and the Pacific; GDP, gross domestic product; LAC, Latin American and the Caribbean; SAS, South Asia; SSA, sub-Saharan Africa.

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

[4]C. Peter Timmer, "The Agricultural Transformation," in H. Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 1, (Amsterdam: Elsevier Science, 1988).

productivity). This rising level of output per worker creates a surplus in the rural economy. The second phase of the structural transformation consists of the transfer of that surplus from the agricultural to the nonagricultural sector. This transfer can be implemented either directly (through taxation) or indirectly (through other forms of government intervention). This process of resource transfer is facilitated by the increasing integration of agricultural and nonagricultural factors and product market linkages, with intersectoral trade and labor migration being supported by improved infrastructure. The progressive integration of the agricultural and nonagricultural sectors is the third phase of structural transformation. Finally, the fourth phase occurs when the agricultural sector is fully integrated into the macroeconomy, at which point much of the economic distinctiveness of agriculture will have faded. This is the situation that we currently observe in the advanced economies, where agriculture and nonagriculture operate in essentially the same labor and capital markets.

The concentration of poverty in rural areas that it is implied by the gap between rural employment shares and agriculture's share in GDP also implies that average labor productivity tends to be higher in nonagricultural sectors. Indeed, it may be reasonable to believe that poverty is concentrated in rural areas *because* the level of agricultural productivity is lower than the level of productivity in nonagriculture. Williams College economist Douglas Gollin calculated the ratio of nonagricultural output per worker to agricultural output per worker in 1999–2000 for a number of countries.[5] His results are summarized in Table 16–2. Gollin's calculations confirm that average labor productivity is substantially higher in nonagricultural sectors than in agriculture. In general, the ratio of average labor productivity in nonagriculture relative to agriculture is on the order of 2.5 to 7. Countries in this range include Brazil, Côte d'Ivoire, Ghana, Indonesia, Mexico, and Pakistan. Yet, for more than a few countries, such as Burkina Faso, Burundi, China, and Thailand, this ratio is greater than 10.

If average labor productivity is often substantially higher in nonagriculture than in agriculture, then the process of structural transformation in which labor flows from agriculture to nonagriculture creates a potentially important source of economic growth. We can think of part of the benefit of structural transformation as arising from the transfer of labor from relatively low-productivity to relatively high-productivity employment. Columns 3 to 5 of Table 16–2 present Gollin's calculation of the contributions of sectoral shifts of labor to growth in total output per worker for selected countries.[6] In China, average output per worker in nonagriculture in

[5]Douglas Gollin, "Agricultural Productivity and Economic Growth," in Prabhu Pingali and Robert Evenson, eds., *Handbook of Agricultural Economics*, vol. 4 (Amsterdam: Elsevier, 2010).

[6]Gollin calculates the contribution of sectoral reallocation of labor to growth in aggregate output per worker as a residual. This residual is the growth in output per worker that is not explained by adding the separate contributions to growth in output per worker coming from growth in agricultural and nonagricultural output per worker. Accurate accounting for these distinct sectoral contributions requires that each sectors' contribution be weighted by the share of GDP coming from each sector. If the weighted sum is different from the actual growth in aggregate output per worker for a given country, then the difference (that is, the residual) provides an estimate of the contribution of the reallocation of labor between sectors. This type of calculation is thus quite similar in its approach to the growth accounting analysis presented in Chapter 4.

TABLE 16-2 **Ratio of Sectoral Labor Productivity, Agriculture, and Nonagriculture (1999–2000) and the Contribution of Sectoral Labor Shifts to Growth (1960–2000)**

COUNTRY	NONAGRICULTURAL OUTPUT TO AGRICULTURAL OUTPUT*	GROWTH RATE†	GROWTH FROM		
			AGRICULTURE	NONAGRICULTURE	SECTORAL SHIFTS
Burundi	13.91	0.005	−0.011	0	0.016
Mexico	6.28	0.01	−0.001	−0.002	0.013
Burkina Faso	29.07	0.013	0.002	0.006	0.006
Côte d'Ivoire	3.03	0.014	0.004	0	0.01
Brazil	3.37	0.017	0.002	−0.005	0.021
Ghana	2.43	0.025	0.014	0.007	0.004
Indonesia	5.07	0.028	0.001	−0.001	0.027
Pakistan	2.54	0.03	0.007	0.005	0.018
Thailand	13.07	0.043	0.004	0.007	0.032
China	11.25	0.053	0.014	−0.015	0.054

*Ratio of nonagricultural output per worker to agricultural output per worker.
†Output per worker

Source: Douglas Gollin, "Agricultural Productivity and Economic Growth," in Prabhu Pingali and Robert Evenson, eds., *Handbook of Agricultural Economics*, vol. 4. (Amsterdam: Elsevier, 2010), tables 4 and 5.

1999–2000 was more than 11 times greater than average output per worker in agriculture. Output per worker from 1960 to 2000 grew at 5.3 percent per year, essentially all of which was explained by shifts of workers from agricultural to nonagricultural employment. In Thailand, the labor productivity ratio was over 13. In that case, 3.2 percent of the country's 4.3 percent growth in output per worker during the four decades was explained by sectoral shifts of labor. Intersectoral labor shifts accounted for about three-quarters of the growth of output per worker in Thailand.

Our awareness of the structural transformation comes from historical observation. Yet, the notion of economies consisting of two sectors (agricultural and nonagricultural, traditional and modern, etc.) and the understanding of development as a process centered around the transfer of labor (and capital and other resources) from agriculture to nonagriculture is firmly grounded in the theoretical tradition of two-sector (or dualistic) growth models.

TWO-SECTOR MODELS OF DEVELOPMENT

Although single-sector growth models, presented in Chapter 4, have the great advantage of simplicity, they do not explore production in different sectors such as agriculture, industry, or services (such as banking or tourism); the allocation of capital,

labor, and land across these different activities; or the implications for growth. Like the one-sector models, two-sector models recognize the prime importance of labor and capital in the growth process. Two-sector models can also explore differences in the levels and growth rates of productivity in different activities and the implications for relative wages (and returns on capital investment); the allocation of labor and capital across the two sectors; and the potential for migration of labor from rural (agricultural) to urban (industrial) areas. We present two-sector models here to emphasize their value for understanding the interactions between agriculture and nonagriculture during the course of development, rather than as growth models per se.

THE LABOR SURPLUS MODEL

Two-sector models have a long tradition in economic thinking. The best-known of the early models appeared in David Ricardo's *The Principles of Political Economy and Taxation*, published in 1817. In his model, Ricardo included two basic assumptions that have played an important role in two-sector models ever since.

- He assumed that agricultural production was subject to **diminishing returns** because crops require land and the supply of arable land is limited. To increase production, Ricardo felt, farmers would have to move onto poorer and poorer land, and therefore each new acre of land matched with the same amount of labor would produce less grain.
- Ricardo formulated a concept that today is called **labor surplus**. Britain in the early nineteenth century still had a large agricultural workforce, perhaps more than was necessary to produce sufficient food for all consumers. Ricardo believed that the industrial sector could draw away surplus labor from the farms without reducing total agricultural production or causing a rise in wages in either urban or rural areas.

Labor surplus, to the extent that it exists, is closely related to concepts such as rural unemployment and underemployment or disguised unemployment. Very few people in rural areas of developing countries are unemployed in the strict sense. Yet, agriculture's seasonality implies that rural employment may fluctuate from nearly full employment during planting and harvesting times to substantially lower levels of employment at other times, even though the total number of workers remains more or less constant. Economists call this **underemployment** or **disguised unemployment**.

The two-sector models we examine here focus on employment and are designed to answer several questions. How does surplus labor (or very low productivity labor) in agriculture affect industry? Can workers move to industry without causing a fall in agricultural production and thus expand total economic output? Will accelerated population growth help or make matters worse? And, what is the effect of agricultural productivity growth on intersectoral labor flows and economic growth?

The modern version of the two-sector labor-surplus model was first developed by economics Nobel laureate W. Arthur Lewis in 1954.[7] Lewis, like Ricardo before him, pays particular attention to the implications of surplus labor for income distribution and growth. Our concern here, however, is with the relationship between industry and agriculture, and for that, we use a version of Lewis's model formulated by economists John Fei and Gustav Ranis in 1964.[8]

Our starting point is the agricultural sector and the **agricultural production function**. We assume two inputs, labor and land, produce an output, such as grain. The production function in Figure 16–2 is similar to but differs slightly from the standard neoclassical production function with continuously diminishing returns to labor. Instead of showing output as a function of capital per worker, agricultural output is shown as a function of labor per unit of land. Because any increase in labor must be combined with the existing stock of land (or perhaps new land of decreasing quality), the production function exhibits diminishing returns.

The labor surplus model, however, takes diminishing returns to its extreme: It assumes that at some point, further additions of labor make zero (or even negative) contribution to output. In Figure 16–2, a rise in the labor force from *a* to *b* leads to an increase in output from *d* to *e*; an equal increase in labor from *b* to *c* leads to a smaller

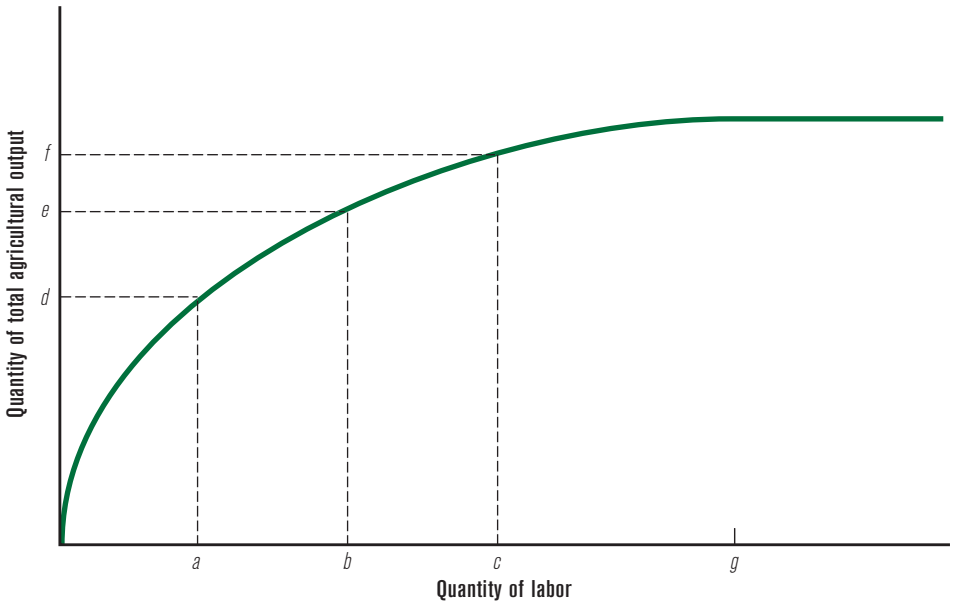


FIGURE 16–2 **The Agriculture Production Function**

[7]W. A. Lewis, "Economic Growth with Unlimited Supplies of Labor," *The Manchester School of Economic and Social Studies* 22 (1954), 139–91.

[8]John C. H. Fei and Gustav Ranis, *Development of the Labor surplus Economy: Theory and Policy* (Homewood, IL: Richard A. Irwin, Inc., 1964).

rise in output. At point *g*, however, further increases in the amount of labor used lead to no rise in output at all. Beyond point *g*, the **marginal product of labor (MPL)** is zero or negative, so additional labor causes no increase or reduction in output. This could happen if all arable land were fully used and so many workers were available already that adding new ones would not result in more grain being produced.

The next step is to show how rural wages are determined. The standard assumption in all labor surplus models is that rural wages do not fall below a minimum level, regardless of how many workers are available. More specifically, the usual assumption is that rural wages do not fall below the **average product** of farm labor. The logic behind this view is that a member of a farm household will not look for work outside the household unless he or she can earn at least as much as he or she would receive by staying at home. At home, total food production would be divided equally among all members of the household so each person consumes the average product of household production. A slightly different, but comparable concept is that wages are not allowed to fall below a **subsistence level**. In this view, no one would look for work off the farm for wages that were below the amount needed for a minimum level of subsistence. The minimum wage, however determined, sometimes is called an **institutionally fixed wage** to contrast it with wages determined by market forces.

If the MPL falls to zero while wages remain at some minimum level, a wedge emerges between the MPL and the wage rate. This is the key characteristic that distinguishes labor surplus models from standard neoclassical models with perfectly competitive markets (examined in the next section) in which the MPL equals the wage rate. Labor surplus models include not just the possibility that the MPL falls to zero but situations in which the MPL is above zero but less than the rural minimum wage.

These concepts are presented in Figure 16–3, which is derived directly from Figure 16–2, but with several changes. To begin with, the horizontal axis is flipped, so that moving to the right represents a decline in the number of agricultural workers. At the origin, the horizontal axis represents the point where the entire labor force works in agriculture, with no one working in industry. Next, whereas the vertical axis in Figure 16–2 represents the total agricultural product, in Figure 16–3, it is converted to represent the marginal product per unit of labor. Thus when moving to the right, as the number of agricultural workers declines, the MPL begins to increase (corresponding to Figure 16–2 in which increases in the number of workers leads to diminishing returns to labor).

The minimum or subsistence wage is represented by the dotted line *hi*. Agricultural wages remain at this level until the MPL (represented by the solid curve) rises above this minimum, which occurs at point *i*. Thereafter, agricultural wages rise, following the marginal product curve as more labor is drawn away from the sector. This curve plays a dual role: It shows both the agricultural wage and the minimum amount that industry must pay to lure workers off the farm. To hire workers away from the farm, factories have to pay at least as much as the workers are earning on the farm. Therefore, the line *hij* in Figure 16–3 can be thought of as the **supply**

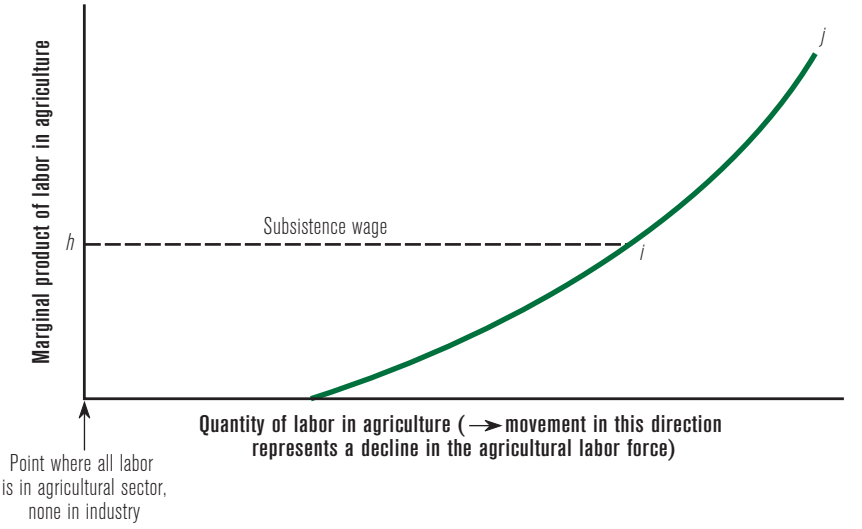


FIGURE 16–3 Modern Product of Labor in Agriculture

As the quantity of agricultural labor decreases, the marginal product increases.

curve of labor facing the industrial sector. Actually the usual assumption is that the supply curve of labor in industry is a bit above the line *hij* because factories must pay farmers a bit more than they receive in agriculture to get them to move.

The key feature of this supply curve of labor is that, unlike more common supply curves, it does not rise continuously as one moves from left to right but instead has a substantial horizontal portion. Formally, this means that the supply curve of labor up to point *i* is **perfectly elastic**. **Elasticity** is a measure of responsiveness, equal to the percentage change in one variable (in this case, the supply of labor) arising from a percentage change in another variable (in this case, wages).[9] The elasticity becomes very large when small changes in wages induce very large changes in the supply of labor. Perfect elasticity occurs when the ratio of these two percentages equals infinity. From the point of view of the industrial sector, this means that the sector can hire as many workers as it wants without having to raise wages, at least until the amount of labor is increased beyond point *i*. To the right of this point, sometimes called the *turning point*, industrial wages rise as firms draw more workers from the agricultural sector.

Figure 16–4 shows the supply and demand for labor for the industrial market. The supply curve *kk′* is taken directly from Figure 16–3 and shows the wages that industry has to pay to draw workers from agriculture. The amount 0*k* on the vertical

[9]More formally, this elasticity is the ratio of the percentage change in the supply of labor $(\Delta L/L)$ to the percentage change in the wage rate $(\Delta W/W)$: Elasticity $= (\Delta L/L)/(\Delta W/W)$. In the case of perfect elasticity, this ratio approaches infinity, which implies a flat supply curve along which employers can employ as much labor as they choose at that fixed wage.

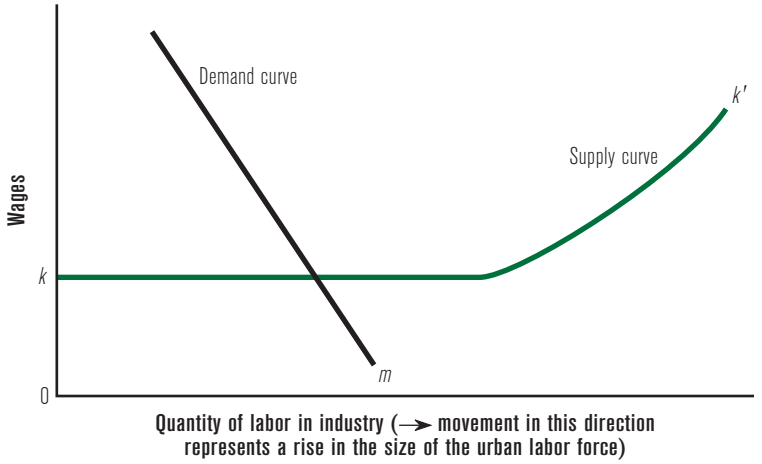


FIGURE 16–4 The Supply and Demand for Industrial Labor

The supply curve, *kk′*, is drawn directly from Figure 16–3. Demand, *m*, is derived from the industrial production function.

axis in Figure 16–4 is assumed to be slightly higher than the subsistence wage in Figure 16–3, as discussed previously. The supply curve turns up when the withdrawal of labor from agriculture no longer can be accomplished without a decline in the agricultural output (when the MPL rises above 0) because, at that point, the relative price of agricultural produce rises and this necessitates a commensurate rise in urban wages. In other words, because after the turning point agricultural production is falling, the price of food rises and industry must pay its workers more to compensate for the higher price of food. The demand curve for labor in industry, *m,* displays the usual downward-sloping quality and shows the wages that the industrial sector is willing to pay for different quantities of labor. This demand curve is determined by the marginal product of labor in industry and can be derived from the industrial production function. To simplify our exposition, we do not show the details of this derivation.

The final step is to combine Figures 16–2, 16–3, and 16–4 into a complete version of the model, which is shown in Figure 16–5. Figure 16–5a is the agricultural production function of Figure 16–2 with the horizontal axis flipped. An increase in the number of agricultural workers is shown as a movement from right to left from the origin (0 workers in agriculture) to point *p*, which is the initial size of the total labor force. Many versions of this model use total population rather than the labor force, and this switch has little effect if the labor force is closely correlated with total population. Figure 16–5b shows the MPL curve from Figure 16–3, and Figure 16–5c shows the supply and demand curves for labor in the industrial sector from Figure 16–4. In all three panels, a movement from left to right represents both a decline in the agricultural labor force and a rise in the industrial labor force—that is, a transfer of labor from agriculture to industry.

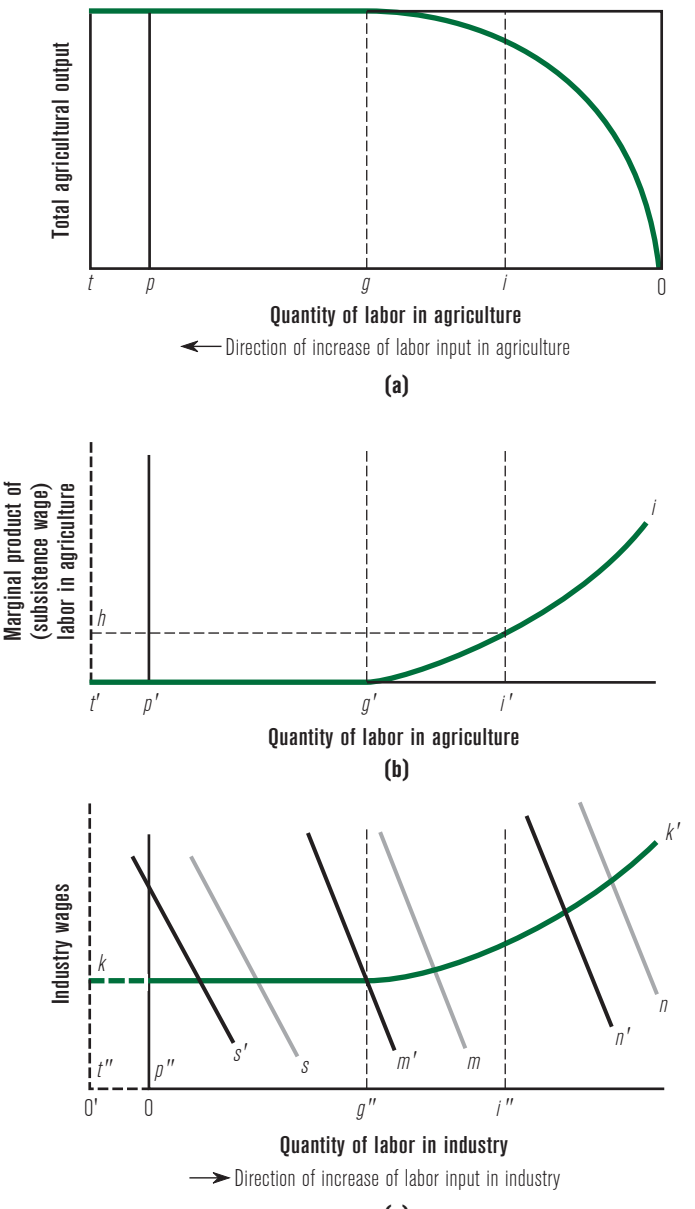


FIGURE 16–5 The Two-Sector Labor Surplus Model

(a) Agricultural production function. (b) Rural (agricultural) labor market. (c) Industrial labor market. The limit imposed by the country's population (0 to *p* in panel a), coupled with the agricultural production function, allows us to analyze the effects of industry wages on the mix between agricultural and industrial labor.

If a labor surplus economy starts with its entire population in agriculture, it can remove a large part of that population (*pg*) and move it to industry or other employment with no reduction in farm output. Industry must pay those workers a wage a bit above subsistence (the difference between the vertical distance *p″k* in Panel c

and $p'h$ in panel b) to get the workers to move. As long as there is some way of moving the food consumed by this labor from the rural to the urban area, industrialization can proceed without reducing agricultural output, implying an increase in total GDP because the marginal product of these workers in industry is greater than zero.

As industry continues to grow, however, it eventually exhausts the supply of surplus labor. Further removal of labor from agriculture leads to a reduction in farm output. A shift in industrial labor demand to m in Figure 16–5c forces industry to pay higher wages to compensate workers for the higher price of food. The rise in the price of agricultural output relative to the price of industrial output (which has not changed) sometimes is described as the terms of trade between industry and agriculture turning against industry and in favor of agriculture. The rising price of food as workers move to industry—that is, the shift in the terms of trade against industry—accounts for the rise in the supply curve of labor between g'' and i'' in panel c.

The Fei-Ranis model can be used to explore the implications of population growth and a rise in agricultural productivity, among other things. To simplify, if one assumes a close relationship between population and the labor force, then an increase in population from p to t increases the length of the horizontal axis in all three panels. Note, however, that additional workers (or a larger population) do not increase agricultural output at all. The elastic portion of both the urban and rural labor supply curves are extended by $p't'$ and $p''t''$, respectively, thus postponing the day when industrialization causes wages to rise.[10]

Most important, if the population rises with no increase in food output, the average amount of food available per capita falls. From the standpoint of everyone but a few employers who want to keep wages low and profits high, population growth is an unqualified disaster. Wages actually may fall in the urban areas, and the welfare of the great mass of farmers certainly falls. It is a model such as this, even if only imperfectly understood, that people often have in mind when they speak of population growth in wholly negative terms.

Britain's economy displayed labor surplus characteristics during Ricardo's time. In the middle of the twentieth century China, India, Indonesia, and some countries in Africa appear to have had surplus labor, but there are few such situations today. The most recent major example of a clear application of the model is China from the 1950s through the 1970s (Box 16–1), but China's surplus labor was fully absorbed by the 1980s. More common is a situation in which a withdrawal of labor leads to a small decline in agricultural production, which brings us to the neoclassical two-sector model.

[10]In the industrial labor supply and demand part of Figure 16–5c, it is also necessary to move the labor demand curves to the left because the 0 point on the horizontal axis has been moved to the left. These new demand curves, s', m', and n', therefore really are the same as s, m, and n. That is, the quantity of labor demanded at any given price is the same for s' as s, and so on.

BOX 16-1 SURPLUS LABOR IN CHINA

In China, by the 1950s most arable land already was under cultivation, and further increases in population and the labor force contributed little to increases in agricultural output. Urban wages rose in the early 1950s but then leveled off and remained unchanged for 20 years, between 1957 and 1977. If allowed to do so, tens of millions of farm laborers would happily have migrated to the cities, despite urban wage stagnation. Only legal restrictions on rural–urban migration, backed by more than a little force, held this migration to levels well below what would have been required to absorb the surplus. Population growth that averaged 2 percent a year up until the mid-1970s continued to swell the ranks of those interested in leaving the countryside. In short, during this period China was a labor surplus country.

China invested in agriculture but only enough to maintain, not to raise, per capita food production. The rural–urban migration that occurred was not fast enough to eliminate the agricultural labor surplus, but it was enough to require farmers to sell more of their production to the cities. Thus the prices paid to farmers for their produce were gradually raised, while the prices paid by farmers for urban products remained constant or fell—that is, the rural–urban terms of trade shifted slowly but markedly in favor of agriculture.

To get out of this labor surplus situation, the Chinese government after 1978 had to accelerate the transfer of workers from agricultural to urban employment, take steps to keep the agricultural pool of surplus labor from constantly replenishing itself, and at the same time invest in agriculture to raise farm output. Accelerating the growth of urban employment was accomplished by encouraging production of labor-intensive consumer goods (textiles, electronics, etc.) and service industries (restaurants, taxis, etc.). To feed this increase in urban population, the government increased domestic food production by raising investment in the agricultural sector, increased food imports, and allowed a further improvement in rural–urban terms of trade that further encouraged growth in food output.

To keep the rural pool of surplus labor from replenishing itself, the Chinese government attacked the surplus at its source through a massive (and controversial) effort to bring down the birth rate. By 1980, China's population growth rate had slowed from 2 to 1.2 percent a year, and after 1997 the rate fell below 1.0 percent a year, where it has remained ever since. Rapid industrial and services growth in and around the cities and in the countryside itself has absorbed roughly 10 million new workers a year. In 1991, the workforce in agriculture reached its peak and has declined steadily since then. The agricultural labor force that accounted for 70.5 percent of total employment in 1978 reached a

key milestone in 1997 when it fell to 49.9 percent, signifying that for the first time in Chinese history, less than half the people employed in the country were farmers. By 2009 the share of employment in agriculture had fallen further to under 38 percent of the total workforce, and people between the ages of 18 and 40 were mainly employed in the cities. Due to migration restrictions imposed by the government, however, their below-working-age children often remained in the rural areas with their grandparents who made up most of the over-age-40 population and most of the agricultural workforce.

THE NEOCLASSICAL TWO-SECTOR MODEL

The neoclassical model, formalized by Harvard economist Dale Jorgenson, differs from the labor surplus model in two key ways: (1) The MPL in agriculture is never zero and (2) there is no institutionally fixed minimum wage, so wages always equal the MPL.[11]

A simple neoclassical model is presented in Figure 16–6, showing the same three panels as Figure 16–5. The agricultural production function in panel a is never flat, and the marginal product of labor curve in Panel b is always rising. Correspondingly, the supply curve of labor to industry in Panel c no longer has a horizontal section. At every point, the removal of workers from agriculture reduces agricultural production and increases the MPL for those remaining in agriculture. Industry must pay an amount equal to that marginal product (plus a premium) to get workers to move. Thus, whereas in the labor surplus model industrial production can increase without reducing agricultural production (leading to an unambiguous increase in GNP), in the neoclassical model an increase in industrial production can take place only alongside a decrease in agricultural production. As labor is removed from agriculture, farm output falls, and to extract enough food from the agricultural sector to pay its workers, industry must pay higher and higher prices for good. Only if industry is in a position to import food from abroad will it be able to avoid these worsening terms of trade. If imports are not available, rising agricultural prices will lead to a higher value of output and hence higher wages for workers in agriculture. As in the labor surplus case, industry will have to pay correspondingly higher wages to attract labor. The total value of GNP rises only to the extent that industrial production rises more than agricultural output falls.

[11]Dale W. Jorgenson, "The Development of a Dual Economy," *Economic Journal* 71 (1961), 309–34.

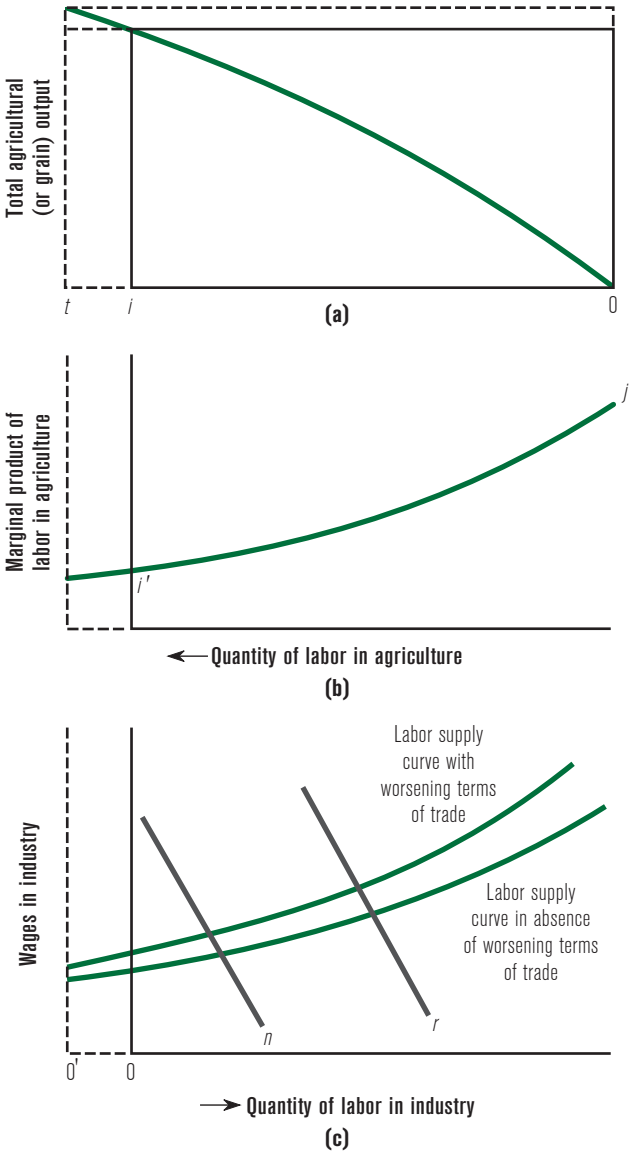


FIGURE 16–6 Neoclassical Two-Sector Model

(a) Agricultural production function. (b) Marginal product of labor in agriculture. (c) Industrial labor market. The key difference between this figure and Figure 16–5 is the agricultural production function agricultural sector, but the curve never flattens out—that is, the marginal product of labor never falls to a minimum subsistence or institutionally fixed wage in Figure 16–6b.

The implications of population or labor force growth in the neoclassical model are different from what they were in the labor surplus model. An increase in population and labor in agriculture raises farm output (dashed line *t* in Figure 16–6a)

because the agricultural production function continues to rise with more labor and never flattens out. Thus, in a neoclassical model, population growth is not such a wholly negative phenomenon. The increase in labor is much less of a drain on food availability because that labor is able to produce much or all of its own requirements, and there is no surplus of labor that can be transferred without a consequent reduction in agricultural output.

If industry is to develop successfully, simultaneous efforts must be made to ensure that agriculture grows fast enough to feed workers in both the rural and urban sectors at ever-higher levels of consumption and prevent the terms of trade from turning against industry. A stagnant agricultural sector with little new investment or technological progress causes the wages of urban workers to rise rapidly and thereby cuts into profits and the funds available for industrial development. In other words, investments in improved agricultural technology or rural roads (which reduce the cost of transporting farm inputs and outputs, including food) help the industrial sector by reducing its costs. Whereas in the labor surplus model policy makers can ignore agriculture until the surplus of labor is exhausted, in the neoclassical model there must be a balance between industry and agriculture.

These dual sector models provide a clear framework in which to consider the economywide effects of productivity growth in agriculture. The best example is an extension of the two-sector model proposed by economists Mukesh Eswaran and Ashok Kotwal.[12] Eswaran and Kotwal begin by considering an economy that is closed to trade and consists of two sectors, textiles and agriculture. In this stylized economy, textiles and food are the only goods produced or consumed. Textiles are assumed to require only labor, while food production requires labor and land. Eswaran and Kotwal do not require surplus labor (and thus come closer to the neoclassical model described in Figure 16–6) but add an assumption that consumer preferences are such that people begin to consume textiles only once they are able to afford at least a subsistence level of food. It is assumed that most of the workers have not reached this subsistence level and thus will spend all incremental income on food until they reach that subsistence level of consumption. In contrast, landlords (the only other actors in this model besides workers) are richer and consume both food and textiles. The assumption that Eswaran and Kotwal share with Jorgenson, that labor markets are competitive, implies that wages are determined by the productivity of labor. Further, labor is mobile between sectors, which implies that the wage is the same in both sectors and that inducing labor to migrate from agriculture to textiles requires wages to increase. Because at an early

[12]M. Eswaran and A. Kotwal, "A Theory of Real Wage Growth in LDCs," *Journal of Development Economics* 42 (December 1993), 243–69, and "The Role of Service Sector in the Process of Industrialization," *Journal of Development Economics* 68 (2002), 401–20.

stage of development agricultural productivity is typically quite low, so is the wage in this economy. As wages are set in the agriculture sector, the low wage in agriculture that results from low agricultural productivity also keeps wages low (and poverty high) in the textiles sector.

Now, suppose that agricultural productivity increases. This results in both higher wages for agricultural workers and increased profits for landlords (as land also becomes more productive). This increases both the supply of and the demand for food. Landlords will also increase their demand for textiles, bidding up both textiles prices and wages. Labor will be willing and able to leave agriculture to produce those added textiles. Thus agricultural productivity growth in this simple model results in the transfer of labor from agriculture to industry while also reducing poverty (through the increased wages).

If instead the productivity increase occurs in textiles, rather than in agriculture, each industrial worker can produce more textiles. The price of textiles (in this closed economy) will fall. Yet, if most workers have yet to attain the subsistence level of food consumption (and hence consume no textiles), they receive no benefit from this change and there is no incentive for labor to migrate to industry. The wage and poverty remain unchanged, and the only beneficiaries of productivity growth in industry are the landlords, who can now consume more textiles. The implication, at least for closed economies, is that it is productivity in agriculture, not industry, that matters for poverty reduction. Whether or not surplus agricultural labor exists in any country (and therefore which version of the two-sector model is most appropriate) has long been a subject of debate (Box 16–2).

BOX 16–2 DEBATES OVER SURPLUS LABOR

The most famous debate over the existence of surplus labor was an exchange between two Nobel economics laureates, Theodore Schultz and Amartya Sen.[a] Schultz reasoned that if any country was likely to have surplus rural labor, it was India. To test whether there was surplus labor in rural India, Schultz took advantage of a sort of natural experiment—the global influenza pandemic of 1918–19 which killed approximately 50 million people, including 8.3 percent

[a]Theodore Schultz's original analysis was in his book *Transforming Traditional Agriculture* (New Haven, CT: Yale University Press, 1964). The subsequent round of exchanges between Schultz and Sen was published in *The Economic Journal* 77, no. 305 (March 1967).

of India's agricultural labor force. Schultz's test of the surplus labor hypothesis was based on the hypothesis that if there were surplus labor, the reduction in the labor force would not reduce India's agricultural output (or, by extension, the area planted). He compared the area planted just before the pandemic with the area planted just after the pandemic and found a 3.8 percent reduction in the postpandemic area planted. This reduction (which was close to Schultz's prediction of what would have happened in the absence of surplus labor) appeared to support his argument against the existence of surplus labor in agriculture.

Amartya Sen took issue with Schultz's approach. Sen made the point that surplus labor could exist within families. He gave the example of peasant families with five working members, each of whom was 20 percent idle. In that case, withdrawing one member would leave output unchanged, indicating the existence of surplus labor. But, Sen continues, if the same 20 percent reduction in labor occurred because one in five entire households disappeared while the other four remained then output would fall (at least in the short run), even if there were surplus labor within families. It is this latter scenario that Sen argues is a more apt description of the influenza pandemic. Indeed, Sen argues that the effects of the pandemic were unevenly distributed over families with a given region and between different regions in a given province, concluding that Schultz's test was inconclusive regarding the existence of surplus labor.

This debate did not end with Schultz and Sen. More recently economist Mark Rosenzweig presented rigorous microeconomic evidence of an upward-sloping supply curve for labor in developing countries (in contrast to the surplus labor model's assumption of a perfectly elastic labor supply curve, as seen in Figure 16–5, implying surplus labor).[b] Yet, economist Gustav Ranis (co-author of the Fei-Ranis model) later defended the surplus labor model. Ranis concedes that the number of contemporary developing countries with surplus labor has been declining but claims relatively recent turning points, at which certain countries ceased to have surplus labor—Taiwan in 1968, South Korea in 1973, and Thailand in 1993. Ranis contends, however, that as of 2004, China and India, as well as "other parts of South Asia, much of Central America, the Caribbean, parts of Latin America, and even some countries in sub-Saharan Africa"[c] remain in conditions that qualify as surplus labor.

[b]Mark Rosenzweig, "Labor Markets in Low-Income Countries," in H. Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 1 (Amsterdam: Elsevier, 1988).

[c]Gustav Ranis, "Labor Surplus Economies," Center Discussion Paper No. 900, Economic Growth Center, Yale University, New Haven, CT, December 2004, p. 17.

EVOLVING PERSPECTIVES ON THE ROLE OF AGRICULTURE IN ECONOMIC GROWTH AND POVERTY ALLEVIATION

AGRICULTURE AND ECONOMIC GROWTH

If agriculture becomes a smaller share of output and employment as economies grow, how can agriculture contribute to broader economic growth? A meaningful answer to this question must delve beyond mere national income accounting relationships. In accounting terms, we can think of a country's rate of income growth as the weighted average of the growth rates of various sectors in the economy. If we adopt a dualistic perspective, in which an economy consists of agriculture and nonagriculture, the *direct* effect of growth in the agriculture sector may be large initially, but soon declines. That is, if value added in the agricultural sector is Y_A, we can write the share of agriculture in national output as $\alpha = Y_A/Y$ (where α is between 0 and 1). If the only sectors are agriculture and nonagriculture, then the share of value added from nonagriculture (Y_{NA}) in national income is $(1 - \alpha)$, and we can express the growth rate of national income as the weighted average,

$$\frac{\Delta Y}{Y} = \alpha\frac{\Delta Y_A}{Y_A} + (1 - \alpha)\frac{\Delta Y_{NA}}{Y_{NA}} \qquad [16\text{–}1]$$

where Δ indicates change. Dividing the change in the level of a variable by its level gives us its growth rate. So equation 16–1 simply states that the growth rate of national income is the weighted average of the growth rates of income in agriculture and nonagriculture, where the weights are the share of each sector's value added in national income.

We can see from this equation that the growth rate of the agricultural sector matters in a purely arithmetic accounting sense. Borrowing data for South Asia from Table 16–1, during the period 1965–75, agriculture's share of GDP (α in equation 16–1) was 41.3 percent. If governments in that region had a goal of seeing their economies grow at, say, 5 percent per year per capita, then a sector that accounted for nearly half of GDP would need to contribute. Maybe both sectors would grow at 5 percent per year per capita, in which case that would also be the weighted average growth rate. Yet, if the agriculture sector was allowed to languish because of the perception that only industrialization mattered and (taking the extreme case) thus did not grow at all (that is, if $\Delta Y_A/Y_A = 0$), then reaching a weighted average of 5 percent would require that the nonagriculture sector grow at the annual rate of 8.5 percent per year per capita, a much higher hurdle.

This simple arithmetic highlights the role of agricultural growth early in the development process when agriculture represents a large share of the economy but also

suggests that agriculture's role diminishes as the sector becomes smaller over time. All we need to do is look at Table 16–1 to see that this arithmetic problem becomes less important over time. In South Asia, agriculture's share of GDP fell from 41.3 percent during 1965–75 to only 22.3 percent by the period 1996–2008. This is the structural transformation in action. The deeper questions are how does this structural transformation happen, and what is the role of agricultural growth in that process? In terms of equation 16–1, this amounts to asking whether the growth rate of nonagriculture depends to some extent on the growth rate of agriculture. Even if agriculture's direct arithmetic contribution to growth diminishes over time, agricultural growth might contribute indirectly to economic growth through its effect on growth in nonagriculture. This possibility requires us to think in greater detail about **intersectoral linkages**. Early interpretations of the dualistic development paradigm tended to ignore agriculture's potential indirect contributions to growth.

The Lewis and Fei-Ranis assumption of a very low or zero marginal product of labor in the rural economy was the dominant perspective to emerge historically from dualism. With that perspective, the post–World War II development paradigm maintained that the agricultural sector had little to contribute to the process of economic growth. This was not Lewis's own interpretation. Indeed, in introducing his model, Lewis famously wrote, "industrialization is dependent upon agricultural improvement; it is not profitable to produce a growing volume of manufactures unless agricultural production is growing simultaneously. This is also why industrial and agrarian revolutions always go together, and why economies in which agriculture is stagnant do not show industrial development."[13] It is inherent in the Lewis and Fei-Ranis models (albeit assuming closed economies) that in the absence of increasing agricultural productivity, the withdrawal of labor from agriculture eventually (in Fei-Ranis) or immediately (in the neoclassical version) results in reduced food supply, increased food prices, and thus lower real wages in industry. To continue to attract labor from agriculture under such circumstances, industrialists would have to pay continually higher nominal wages, thus undermining the incentives for continued industrial expansion. However, this was not the lesson as understood by most policy makers and early development thinkers.

The more dominant early interpretation of classical dualism was that of political economist Albert Hirschman, who wrote in 1958, "agriculture certainly stands convicted on the count of its lack of direct stimulus to setting up of new activities through linkage effects—the superiority of manufacturing in this respect is crushing."[14] This view was reinforced by the then-current assertion that primary products, such as agriculture, were doomed to suffer declining terms of trade relative to manufactured goods in international markets and thus presented few growth opportunities for poor countries. (This is the Prebisch-Singer hypothesis, discussed in more

[13]Lewis, "Economic Growth with Unlimited Supplies of Labor," p. 433.

[14]Albert O. Hirschman, *The Strategy of Economic Development* (New Haven, CT: Yale University Press, 1958).

detail in Chapter 18.) Widely seen, then, as a stagnant pool of underutilized labor and capital, agriculture was usually neglected in early development strategies.

The evolution of thought regarding the role of agriculture in economic growth is best understood in terms of intersectoral linkages. This perspective begins with the Lewis linkages of labor and capital flows from agriculture to nonagriculture. These are direct linkages across sectors that operate via factor markets. Ideally, labor and capital are released from agriculture as a consequence of increasing agricultural productivity. However, as discussed in Chapter 17, agricultural productivity growth is not automatic. It requires substantial investment in research and development, systems to extend new techniques to farmers, and functioning input and output markets with appropriate production incentives. In the context of early development paradigms in which agriculture was viewed as a stagnant resource reservoir, the need for such investments was not well understood. Instead, government policies often sought merely to extract labor and capital from the agricultural sector. If the development paradigm was one in which rural labor was assumed to have a zero marginal product (that is, if the production function for agricultural labor eventually became flat), labor could be withdrawn from agriculture with no reduction in output.

By the early 1960s, economists' perspectives on intersectoral linkages and agriculture's potential contributions to growth began to evolve. In a particularly influential article, agricultural economists Bruce Johnston and John Mellor listed five intersectoral linkages.[15] In contrast to the direct Lewis linkages, the Johnston-Mellor linkages were indirect. Their list of agriculture's contributions is (1) increased supplies of food for domestic consumption, (2) released labor for industry, (3) a domestic market for industrial output, (4) a supply of domestic savings, and (5) a source of foreign exchange. While the second and fourth linkages in this list built on the Lewis tradition, Johnston and Mellor set the stage for a richer understanding of agriculture's potential contributions to growth.

The first Johnston-Mellor linkage, increased food supply for domestic consumption, underscores the logic, noted earlier, that was implicit in the Lewis and Fei-Ranis approaches. A vibrant and productive agricultural sector that could both release labor and supply increasing quantities of food for domestic consumption could facilitate industrialization by helping keep food affordable for industrial workers. While affordable food is an important part of agriculture's potential contribution to poverty alleviation, it also benefits urban employers by reducing labor costs. There is a big difference, however, between government policies that keep food cheap by controlling food prices or by heavily taxing agricultural producers and government policies that keep food cheap by investing in agricultural productivity. Johnston and Mellor highlight linkages that operate not only forward, in the sense of agriculture supplying inputs and foreign exchange to nonagriculture, but also backward. The third Johnston-Mellor

[15] B. Johnston and J. Mellor, "The Role of Agriculture in Economic Development," *American Economic Review* 51 (1961), 566–93.

linkage, in which agriculture provides a market for domestic industrial output, can be understood as a backward linkage, one that underscores the consumption externalities that become possible when agriculture is profitable and the sector itself is growing.

Awareness of these intersectoral linkages, arising from the early contribution of Johnston and Mellor, led to numerous efforts to quantify the effects of agricultural growth on broader economic growth. These efforts generally sought to measure agriculture's contributions to growth in terms of **growth multipliers**. A growth multiplier is a number that answers the question, If a country's agricultural value added increases by $1, what is the ultimate impact on that country's aggregate GDP? If GDP is the sum of value added in agriculture and nonagriculture, then an additional $1 of agricultural output automatically adds $1 to national output. The question is to what extent intersectoral linkages operate in such a way that an additional $1 of agricultural income stimulates economic activity in nonagriculture, thus adding indirectly to national income. Growth multipliers seek to measure the sum of these direct and indirect contributions to national income.

Efforts to estimate agricultural growth multipliers based on household-level data demonstrate that it is Johnston and Mellor's backward linkage, the demand from within agriculture for domestic goods produced by the nonagriculture sector, that dominates the multiplier. The intuition is that when new resources are brought into production in the agriculture sector, farmers (and landlords) spend that incremental income. While some of those expenditures will be on additional food, some portion will also be spent on the purchase of nontradable goods produced domestically in the nonagriculture sector. The multiplier effect created by this additional rural demand will be greater under certain circumstances than under others. In particular, agriculture must make up a large share of total employment. In addition, the multiplier effect will be greater when the benefits of agricultural growth are widely shared among rural producers. This is more likely when land ownership is relatively evenly distributed. Demand-driven multiplier effects will also be larger when consumption patterns are such that large portions of the incremental demand are spent on domestically produced labor-intensive goods and services. Finally, multiplier effects will be greater when the supply of the additional goods demanded is highly elastic, which implies that the increased demand will bring forth increased supply without substantially increasing prices.[16] In practice, the origin of such income shocks, at least in Asia, was the technological improvements developed for agriculture as part of the green revolution (discussed in more detail in Chapter 17).

Efforts to quantify agricultural growth multipliers date back to the early 1980s in Asia. Economist C. Rangarajan calculated that the agricultural growth multiplier for India was 1.70, while economists Clive Bell, Peter Hazell, and Roger Slade found the

[16]These conditions and more technical details regarding the estimation of growth multipliers are explained in C. Delgado, J. Hopkins, and V. Kelly, "Agricultural Growth Linkages in Sub-Saharan Africa," International Food Policy Research Institute Research Report No. 107, Washington, DC, 1998.

multiplier in Malaysia to be 1.83.[17] The interpretation of these numbers is that when $1 of additional income is generated in the agricultural sector, it stimulates an additional $0.70 (or $0.83 in Malaysia) of income in nonagriculture, for a total addition of $1.70 to national income. Using similar modeling techniques, a subsequent series of studies focused on West African cases. They found agricultural growth multipliers to be even greater in Africa than in Asia, ranging from 1.96 in Niger to 2.88 in Burkina Faso, though other studies have found growth multipliers in Africa generally smaller than in Asia (more on the order of 1.30 to 1.50).[18] Applying a simpler methodology, Tufts University economist Steven Block estimated an agricultural growth multiplier for Ethiopia to be 1.54, compared to an industrial growth multiplier of 1.22 to 1.34.[19]

The Johnston-Mellor linkages also paved the way for more recent perspectives on agriculture's potential contributions to economic growth. There is strong empirical support for two additional channels through which agriculture can contribute to economic growth: a stability linkage, which connects stability in food prices with investment; and a nutritional linkage, which operates through the health of the labor force (Box 16–3).[20]

BOX 16-3 THE NUTRITION LINKAGE TO ECONOMIC GROWTH

The nutritional linkage from agriculture to economic growth is based on evidence, not from contemporary developing countries, but rather from research on eighteenth- and nineteenth-century Britain and France by Nobel Prize–winning economic historian Robert Fogel.[a] His research on the secular decline in mortality rates in western Europe combines historical detective work with nutrition, human physiology, and economic analysis. By reconstructing the distribution of

[a]Robert Fogel, "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy," *American Economic Review* 84, no. 3 (June 1994), 369–95.

[17]C. Rangarajan, "Agricultural growth and industrial performance in India," International Food Policy Research Institute, Research Report 33, Washington, DC, 1982; C. Bell, P. Hazell, and R. Slade, *Project Evaluation in Regional Perspective* (Baltimore, MD: Johns Hopkins University Press, 1982).

[18]Delgado, Hopkins, and Kelly, "Agricultural Growth Linkages in Sub-Saharan Africa," in S. Haggblade, P. Hazell, and T. Reardon, eds., *Transforming the Rural Nonfarm Economy* (Washington, DC: International Food Policy Research Institute, 2009).

[19]S. Block, "Agriculture and Economic Growth in Ethiopia: Growth Multipliers from a Four-Sector Simulation Model," *Agricultural Economics*, 20 (1999), 241–52.

[20]For a more detailed exposition of these types of linkages, see C. Peter Timmer, "Agriculture and Economic Development," in B. Gardner and G. Rausser, eds., *Handbook of Agricultural Economics*, col. 2A, (Amsterdam: Elsevier Science, 2002).

caloric intake in late-eighteenth-century Britain and France, Fogel found that the levels of caloric intake for the bottom 20 percent of the distribution were insufficient to support the physical requirements of much (if any) manual labor. In addition, he found that undernutrition of the labor force was also associated with high rates of disease and physical disability (which themselves contributed back to the problem of undernutrition). This type of chronic undernutrition also results in stunted growth (indeed, the average adult male in late-eighteenth-century Britain was about 3.5 inches shorter than in late-twentieth-century Britain). Thus chronic undernutrition in the late eighteenth century was associated, not only with increased rates of illness and disability but also with stunted growth and increased risk of death. (Because height is in part a reflection of long-term nutritional status, shorter people, on average, face a higher likelihood of premature death.)

As agricultural development in Britain and France proceeded and food supplies increased, Fogel found that nutritional status of workers also improved. The caloric intake of the bottom 20 percent of the consumption distribution increased sufficiently to permit them to work. At the same time, their improved nutritional status improved their health status, delaying illness and disability until later in life, further enhancing their ability to engage in productive labor. In short, Fogel concluded that improvements in gross nutrition account for roughly 30 percent of the growth in per capita income in Britain between 1790 and 1980. With current rates of child (under 5 years) stunting greater than 40 percent in much of sub-Saharan Africa and South Asia, this nutritional linkage may have important implications for today's developing countries. Economist Carmel Nadav provides more recent evidence, based on a sample of 97 countries, that improved nutrition contributes to economic growth.[b]

[b]C. Nadav, "Nutritional Thresholds and Growth," Department of Economics, Ben-Gurion University, Israel, 1996.

The stability linkage operates primarily at the macroeconomic level. One reason is that large fluctuations in food prices can provoke broader crises, which may undermine both political and economic stability, as the food riots of 2008 reflect. Political violence in response to rising food prices resulted in 10 deaths and nearly 500 injuries in Mozambique's capital city, Maputo, in September 2010, following a 20 percent increase in the price of bread over the course of that year. This type of instability undermines incentives for investment. (Recall from Chapter 4 the important role of investment in theories of economic growth.)

The stability of food prices may also impact national investment through another channel. In poor countries it is common for consumers to spend 50–70 percent of their income on food. In such settings, unstable food prices may directly act to destabilize savings and investment. While consumer theory might suggest that this is true for any commodity, the singularity of food in this regard derives from its large share of expenditures. Highly unstable food prices can lead farmers to save in a precautionary way to guard themselves from sudden drops in the price of their output, while consumers similarly save to guard themselves against sudden increases in food prices. This type of precautionary savings must be readily accessible (even held in cash), and those resources are thus withheld from productive investment. Food price instability can also destabilize demand for other goods and services in the economy, further undermining incentives for productive investment in the broader economy. This spillover instability arises because the demand for food is price inelastic. When food prices increase, food expenditures increase as well, reducing demand for other goods and services in the economy. The opposite occurs when food prices fall. The resulting instability in prices throughout the economy may further confuse potential investors who rely on price signals to guide their actions or lead them to allocate more investment in speculative, rather than productive, investment opportunities. As a result, the *quality* of investment may also suffer when food prices are unstable.

Empirical support for these arguments comes from several sources. Economist David Dawe presents evidence from cross-country data for the period 1970–85.[21] He finds wide fluctuations in export prices, on balance, lower investment, and slow the rate of economic growth. More specific evidence comes from Indonesia, where rice was a dominant economic sector during the 1960s and 1970s. During that time, the government of Indonesia was heavily involved in efforts to stabilize rice prices. Indonesia's rice price stabilization program during that country's first five-year plan (1969–74) generated nearly one percentage point of additional economic growth each year (or about one sixth of the country's growth over that period). As Indonesia's economy continued to grow, and its structural transformation led the rice sector to play a diminishing role, the macroeconomic benefits of rice price stabilization also diminished. Yet, the net result of these efforts over the period 1969–95 was to increase Indonesia's rate of economic growth by an average of 0.5 percentage points per year, resulting in a net increase of 11 percent greater income over that period.

AGRICULTURE AND POVERTY ALLEVIATION

Economic growth tends to reduce poverty (measured either as a head count index or a poverty gap); yet the effectiveness with which growth reduces poverty may vary from one economic setting to another. Recent research has shown that the *sectoral composition of growth* matters: On average, agricultural growth is more effective in

[21]D. Dawe, "A New Look at the Effects of Export Instability on Investment and Growth," *World Development* 24, no.12 (December 1996), 1905–1914.

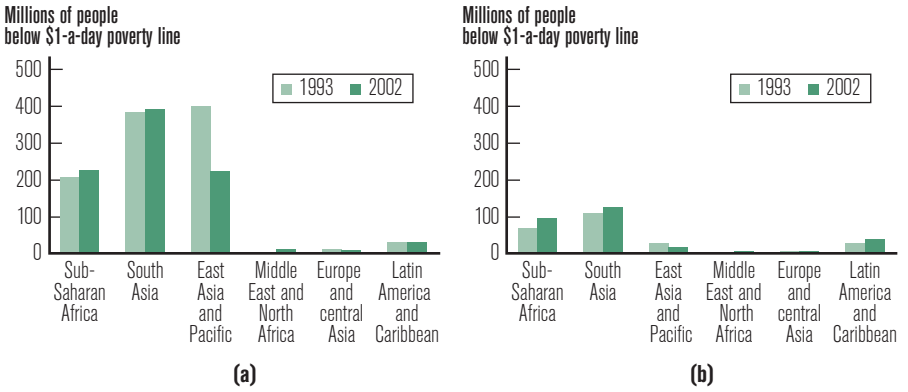


FIGURE 16–7 Number of Poor, 1993 and 2002

(a) Rural poverty. (b) Urban poverty. The number of poor rose in South Asia and sub-Saharan Africa from 1993 to 2002, measured as millions of people below the $1-a-day poverty line.

Source: Figure 1 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

reducing poverty than is industrial growth. This conclusion about the special potency of agricultural growth for alleviating poverty rests in part on the observation that poverty in many regions of the developing world is disproportionately concentrated in rural areas. Figure 16–7 compares the numbers of rural and urban poor in different regions for the years 1993 and 2002. For sub-Saharan Africa, South Asia, and East Asia, the numbers of rural poor vastly exceed the numbers of urban poor. Among these regions, only East Asia shows substantial reduction in the number of rural poor between 1993 and 2002; however, even in that region, the absolute number of rural poor in 2002 was about eight times greater than the number of urban poor. In countries where agriculture accounts for a large share of employment and poverty is disproportionately rural, the extent to which growth is pro-poor depends substantially on the extent to which agriculture is included in growth.

The effectiveness of the mechanisms through which agriculture helps reduce poverty also depends on the incidence of poverty among various groups in society. In reviewing recent evidence, University of California at Berkeley economists Alain de Janvry and Elisabeth Sadoulet note that for producers, the effect of agriculture in reducing poverty depends on access to assets (notably land), while for rural workers the benefits depend on their access to increased employment opportunities in agriculture and rural nonfarm activities.[22] For consumers, greater supplies of food help to reduce poverty by lowering food prices. Consumers of food might or might not also be producers. The most obvious beneficiaries of reduced food prices are pure consumers—the urban poor and rural landless. Yet, in many countries, the

[22]A. de Janvry and E. Sadoulet, "Agricultural Growth and Poverty Reduction: Additional Evidence," *The World Bank Research Observer* 25, no. 1 (February 2010).

majority of smallholder farmers (often the largest group among the poor) are also net consumers of food. They are either incapable of satisfying all of their food needs through their own production, or they sell some of their output for cash and buy food (and other goods). Thus, they too benefit from lower food prices (provided that the gains from reduced spending outweigh any losses to income). Lower food prices could, of course, hurt net sellers, unless productivity gains outpace declining prices. This consideration points to the importance of agricultural productivity growth, both as a driver of reduced food prices and as a source of increased income for producers. The broad association between poverty reduction and growth in agricultural productivity is reflected in Figure 16–8, which compares the experiences of South Asia and sub-Saharan Africa. In South Asia, increasing cereals yields during the 1990s accompanied decreasing incidence of poverty, while in sub-Saharan Africa relatively stagnant cereals yields accompanied a lack of progress in reducing the incidence of poverty.

While the ability of agriculture to reduce poverty is greater in settings where poverty is concentrated in rural areas, this ability also depends on certain conditions within those rural areas. The effectiveness of agriculture in reducing poverty is much stronger in settings where land ownership is more evenly distributed and where production techniques are more labor-intensive, and thus generate more employment.

How does agriculture compare with nonagriculture in its ability to reduce poverty? A number of studies have attempted to quantify the effects of growth originating in agriculture versus nonagriculture. The general consensus among these studies, based on cross-country evidence, is that growth originating in agricultural growth is on the order of three times more effective at reducing poverty than growth originating

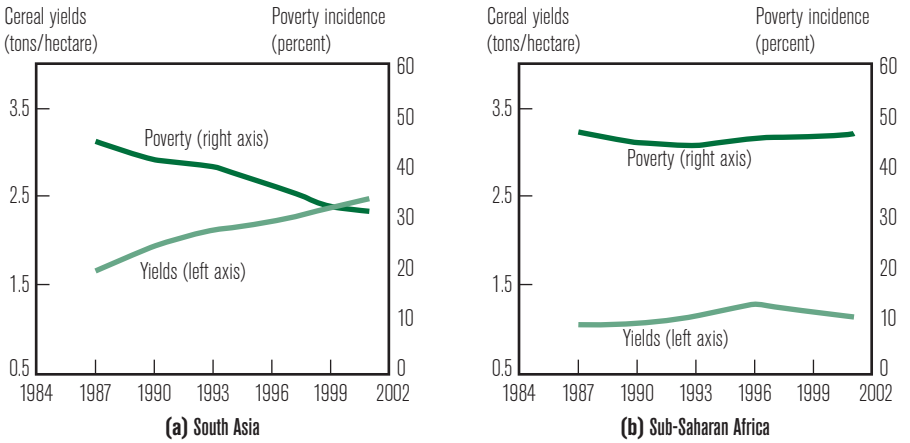


FIGURE 16–8 Cereal Yields and Poverty in South Asia and Sub-Saharan Africa, 1984–2002

Cereal yields rose and poverty dropped in South Asia but both indicators remained unchanged in sub-Saharan Africa.

Source: Figure 1.1 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

in nonagriculture.[23] The World Bank's *World Development Report, 2008*, which focuses on the role of agriculture in development, presents the results of a study by University of California at Berkeley economists Ethan Ligon and Elisabeth Sadoulet, who compared the effects of a 1 percent increase in GDP originating in agriculture versus nonagriculture on the expenditures of each decile of the income distribution. Based on cross-country data from 42 countries over the period 1978–2003, they estimated that growth originating in agriculture was at least three times as effective in reducing poverty as growth in nonagriculture. They also found that among the poorest people (those in the bottom decile of the income distribution), the effect of agricultural growth was approximately six times greater, whereas income growth originating in nonagriculture actually resulted in small reductions in income among the poorest households. Of course, these cross-country results do not describe the experience of any particular country, where the results will depend on a range of conditions.

AGRICULTURAL GROWTH AS A PATHWAY OUT OF POVERTY

Having established some of the theoretical connections between agricultural growth and poverty alleviation and also having considered some empirical evidence, we now turn to the question of mechanisms. For households living in poverty, what specific opportunities does agricultural growth provide for pathways out of poverty? There are three broad avenues: One possibility is for poor, subsistence-oriented farmers to increase their on-farm output for sale; a second pathway can be to engage in rural labor markets, especially for off-farm rural work; and, a third is to migrate to the urban economy. As the *World Development Report* of 2008 notes, these pathways are complementary. Nonfarm earnings can be used to invest in improved farming, and linkages within the rural economy are such that agricultural growth can both generate off-farm employment and facilitate migration. Migration, in turn, can be an important source of remittances back to rural areas that can support household consumption or investment, as has been the case in the Philippines in particular.

The 2008 *World Development Report* describes various success stories. In Tanzania the farmers that were mostly likely to exit poverty were those who diversified their production to include vegetables, fruits, and vanilla for export, in addition to traditional food crops. Similar diversification of farm output to higher-value commodities contributed to poverty reduction in Uganda and Malawi. In contrast, India and Indonesia are cited as examples in which labor markets provided an important pathway out of rural poverty for households engaged in nonfarm rural activities, such as rural services and small-scale industry. In China and Nepal, migration to cities has played a larger role. A more detailed examination of rural poverty reduction in Vietnam highlights the complementarity of these pathways out of poverty.

[23]These studies are summarized in de Janvry and Sadoulet, "Agricultural Growth and Poverty Reduction," and in *World Development Report, 2008, Agriculture for Development* (Washington, DC: World Bank, 2007).

Table 16–3 summarizes the experience between 1992–93 and 1997–98 of three categories of Vietnamese farming households who earn more than 50 percent of income from agriculture (a category that makes up 47 percent of Vietnam's rural population). This was a period in which value added in Vietnamese agriculture was growing at the rapid pace of 4.1 percent per year. The table distinguishes between subsistence-oriented farmers (6 percent of rural farming households, who sell less than 10 percent of their output), market entrant households (13 percent of farming households, who began the period as subsistence farmers but ended the period as market-oriented farmers), and market-oriented farmers (28 percent of farming households, who sell more than 25 percent of their output).

Table 16–3 shows that the market-oriented households enjoyed the greatest reduction in poverty, with rates falling from 64 percent in 1992–93 to 37 percent in 1997–98 (a reduction of 27 percentage points, or 42 percent reduction relative to the base year). This group diversified away from agriculture as a source of income, and within their agricultural earnings they also diversified away from staple crops into higher-valued crops. Market entrants tended to be poorer than the market-oriented

TABLE 16–3 Changing Market Participation among Farming Households in Vietnam

HOUSEHOLD CHARACTERISTICS	SUBSISTENCE ORIENTED (6%)*		MARKET ENTRANT (13%)*		MARKET ORIENTED (28%)*	
	1992–93	1997–98	1992–93	1997–98	1992–93	1997–98
Assets						
Land owned (hectares)	0.37	0.43	0.50	0.57	0.60	0.72
Education of household head (years)	4.6		6.3		6.3	
Context						
Market in community (%)	31		40		47	
Outcomes						
Real income per capita (1998 dong 1,000)	893	1,702	1,138	2,042	1,359	2,978
Share of agricultural income in total income (%)	80	62	83	66	83	73
Share of households below the poverty line (%)	86	62	73	48	64	37
Share of gross agricultural output by crop type						
Staple crops (%)	78	73	70	61	63	54
High-value crops (%)	14	13	21	31	29	39

*Number in parentheses is percent of Vietnam's rural population. These three categories of farmers collectively make up the 47 percent of the Vietnam's rural population that earns most of its income from farming.

Source: Table 3.1 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development*, 2008. Reprinted with permission.

households but also experienced significant reductions in their poverty rate (from 73 percent to 48 percent, a reduction of 34 percent). They too diversified away from agriculture as a source of income and away from staple crop production. Even those households that remained subsistence-oriented producers participated in agricultural growth. Yet, by comparison, the subsistence households experienced the smallest reductions in poverty rates (24 percentage points, or a 28 percent reduction) and remained heavily reliant on staple agricultural production. Agriculture thus provides a variety of pathways out of poverty. In reviewing these results, de Janvry and Sadoulet make the important point that public policies designed to reduce rural poverty (a focus of Chapter 17) must be tailored to specific types of households.[24] Market-oriented households, they note, will benefit from improved competitiveness, whereas market entrants will benefit from access to assets and markets, and subsistence-oriented farmers require improved production opportunities and access to rural labor markets. In this sample, it is clear that the beneficiaries of the greatest rate of poverty reduction among these households were the market-oriented households, whose lower initial poverty rate likely reflects greater initial access to markets. These examples from Vietnam highlight the importance of income diversification within the rural economy.

While nearly all rural households participate in agriculture, they vary widely in the shares of their income coming directly from agriculture itself. Figure 16–9 illustrates this variability across countries and regions. Nearly all rural households in Ghana, Guatemala, Vietnam, and Nepal participate in agriculture. Yet, the share of household income coming from agriculture varies widely across these illustrative countries, from a high of nearly 60 percent in Ghana, to approximately 40 percent in Vietnam. Looking within these countries, Figure 16–10 reveals fairly consistent patterns of income diversification in regard to the income distribution. In general, the poorest households are the most directly dependent on agriculture as a source of income. Agricultural wage labor also declines sharply as a share of household income as income increases, while nonagricultural income shares increase substantially. These types of income diversification may reflect strategies for coping with various risks associated with farming. It is likely that relatively wealthier households have more options for diversification than relatively poor households. Poor households, lacking land or other assets, may be forced into low-wage nonfarm employment, whereas relatively better-off rural households may have access to the capital required to take advantage of local business opportunities.

The ability of the nonfarm rural economy to provide a pathway out of poverty depends on the rate of economic growth, but also on population density.[25] India and China are examples of countries in which urban crowding, high rents, and improving infrastructure have contributed to the growth of opportunities in the nonfarm rural

[24]de Janvry and Sadoulet, "Agricultural Growth and Poverty Reduction."

[25]Haggblade et al., eds., *Transforming the Rural Nonfarm Economy.*

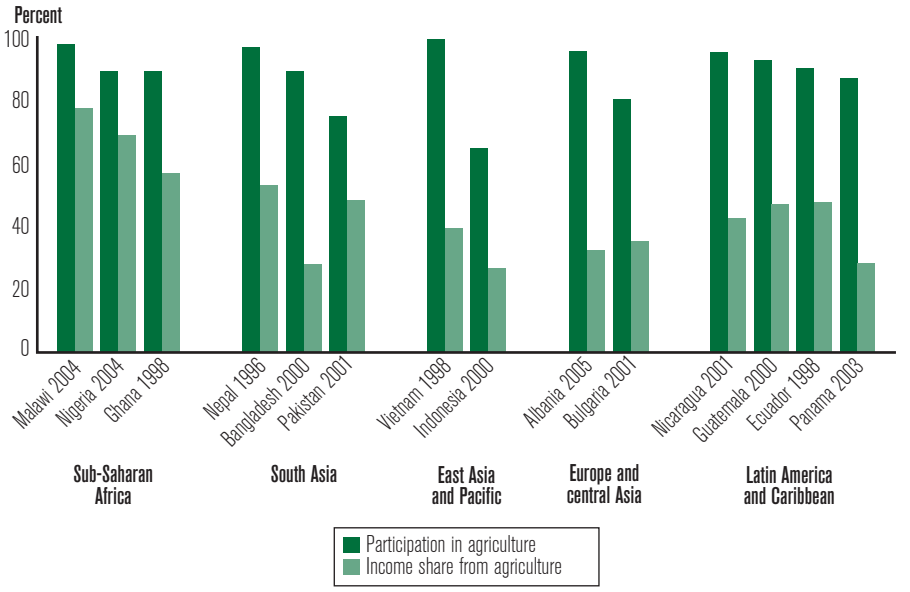


FIGURE 16–9 Percent of Households Participating in and Deriving Income from Agriculture

Source: Figure 3.2 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

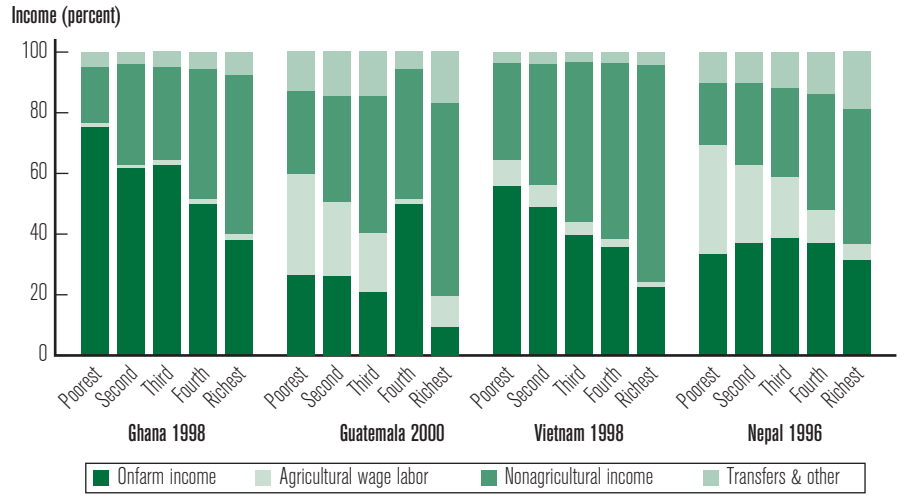


FIGURE 16–10 Sources of Income

Source: Figure 3.3 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

economy in a way that is much less common in sub-Saharan Africa. In most countries, the nonfarm rural economy is characterized by a high degree of heterogeneity in the scale of operations and the types of activities undertaken. Services tend to dominate manufacturing. A key point regarding the viability of nonfarm rural employment as a pathway out of poverty is that it depends on the viability of agriculture itself. Many opportunities in the nonfarm rural economy involve the processing and transportation of food, selling agricultural inputs, servicing agricultural machinery, and providing other services for which the demand arises from agriculture.

Similarly, out-migration from rural areas, although a potentially important pathway out of poverty, can reflect either a severe lack of livelihood opportunities in the rural economy or the real promise of greater welfare in urban employment. Some migration is temporary and follows agricultural seasonality, while some is permanent. From the perspective of rural development, migration can be a two-edged sword. Often, those who leave the rural economy are younger, better educated, and more skilled. Although this might reduce average labor productivity in rural areas, migrants may also be an important source of remittance income, which can contribute to improved productivity by increasing access to agricultural inputs and other productive assets.

SUMMARY

- A variety of characteristics differentiate agriculture from other sectors of the economy. These characteristics include its large share of GDP and employment, unique features of the agricultural production function, a large share of the sector's output consumed directly by its producers, and agriculture's role as a reservoir of resources.
- One of the defining characteristics of economic development is the structural transformation, during which an economy's sector proportions systematically evolve. At early stages of economic development, agriculture dominates both GDP and employment shares in the economy. Structural transformation consists of the shift of sector proportions away from agriculture toward industry and services.
- The decline in the relative size of the agricultural sector is driven, in part, by Engel's law, an implication of which is that the demand for food grows more slowly than income.
- Growth in agricultural productivity plays a critical role in facilitating the release of labor and capital from agriculture for employment in industry and services.

- Dual-sector growth models provide a stylized depiction of the interactions between the agricultural and nonagricultural sectors of the economy, specifying the conditions under which labor and capital migrate from agriculture to nonagriculture. The dominant dual-sector model, developed by Lewis and elaborated by Fei and Ranis, assumes that the defining characteristic differentiating agriculture from nonagriculture is the existence of surplus labor with low or zero marginal product.
- The Lewis and Fei-Ranis models were widely misinterpreted by policy makers in developing countries as a justification for ignoring agriculture's potential contributions to growth and concentrating investment on industrial development.
- Subsequent research demonstrated a more nuanced role for agriculture in economic growth by considering intersectoral linkages. Building on the Lewis linkages, through which agriculture supplies labor and capital to nonagriculture, Johnston and Mellor compiled a broader list of linkages, which includes agriculture's role in supplying food for nonagricultural laborers, agriculture's importance in earning the foreign exchange necessary to finance industrialization, and agriculture's role as a source of demand for the output of the nonagriculture sector.
- Intersectoral linkages also include the benefits to investment from agricultural price stabilization and agriculture's contribution to growth by supporting improved nutrition (and health) of the labor force. Agriculture provides a potent vehicle for addressing poverty, much of which is concentrated in the rural population. These contributions are direct and indirect and include both those employed in agriculture and net food consumers for whom agricultural growth reduces the price of food, often the major share of the household expenditures.
- On average, a 1 percent increase in aggregate GDP originating in agriculture has about three times the effect in reducing national poverty as a 1 percent increase in aggregate GDP originating in nonagriculture.
- Three mutually reinforcing avenues, through which agriculture can provide a pathway out of poverty, are increased on-farm output, increased engagement in rural labor markets and nonfarm rural employment, and migration to cities.
- Although many middle-income countries have grown out of their dependence on agriculture, it is clear that agriculture still has a central role to play in the development of some of today's most challenging regions, large parts of sub-Saharan Africa in particular.

17 Agricultural Development: Technology, Policies, and Institutions

The previous chapter discussed the role of agriculture in economic development, highlighting the sector's potential contributions to both economic growth and poverty alleviation. We turn now to the question of what policies and institutions are needed to maximize those contributions. Addressing that question requires a more detailed focus on the agricultural sector itself. What kinds of policies should governments adopt to promote agricultural development? Which kinds of investments are most effective? And, what kinds of institutions and rules are necessary to make the most of the policies and investments that are chosen? In short, what is required to get agriculture moving?

This chapter approaches that question by addressing the broad constraints that have limited agricultural development in general, and food production in particular, in many parts of the developing world. These constraints take many forms, and their severity varies widely across regions and countries. Some of the constraints to increasing agricultural productivity may be technological. Before the 1950s, agriculture in nearly all developing countries relied on traditional technologies and crop varieties. Traditional agriculture, while stable and well-suited to its environment, was typically characterized by low levels of productivity. Growth in agricultural output relied almost entirely on expansion of the areas under cultivation. By the 1960s, large public investments in crop science had begun to pay off in the form of seeds for improved crop varieties and packages of complementary inputs, the combination of which led to substantial increases in agricultural productivity. Yet, the benefits of this **green revolution** have been concentrated in Asia and Latin America, with comparatively little impact on agricultural development in sub-Saharan Africa.

Other constraints on agricultural development arise from unfavorable government policies (relating in particular to food prices) or from inadequate institutional support for agriculture. Both of these categories of potential constraints relate to the incentives faced by farmers. If food prices are too low, farmers may have limited incentive to expand production. In the long run, low food prices might also be a disincentive for investments in agriculture. Governments in developing countries face a fundamental dilemma in this respect: while higher food prices might provide an important incentive for farmers to increase their output, higher food prices might simultaneously increase the problems of hunger and malnutrition that are endemic in many developing countries. Farmers' incentives are also determined by institutions (that is, by the rules of the game that govern economic relations in society). Institutions governing land rights are of particular concern because farmers who own their land may have greater incentive to invest in maintaining its quality. Before exploring these issues in greater detail, we first set the scene for agriculture in developing countries by characterizing traditional agriculture and farming systems.

CHARACTERISTICS OF TRADITIONAL AGRICULTURE AND AGRICULTURAL SYSTEMS

Farms and farmers in developing countries differ widely from one another, both within and between countries. They differ in size, technology, environmental conditions, crop mix, and degree of commercialization to name but a few categories. Many farmers in poor, densely populated, countries such as Bangladesh might own no land at all and work as day laborers on other people's farms. A farmer in sub-Saharan Africa, most likely a woman, may own and cultivate several small plots totaling one to two hectares,[1] grow five or six different crops, and keep a few chickens. The current model of agricultural development in Brazil involves commercial farms of thousands of hectares, often growing soybeans for sale, coexisting with a multitude of small family farms. Given such diversity, generalizations about **traditional agriculture** are necessarily stylized. Most farms in developing countries are small-scale family farms. They consume a large portion of their own food production, though few farmers are purely subsistence farmers.[2] The livelihood, health, and productivity of farm families depend directly on the efficiency with which they allocate their scarce resources.

Traditional farms tend to be small in area, typically less than three hectares, but intensively cultivated. The demand for labor fluctuates with the agricultural seasons.

[1]One hectare equals 2.47 acres.

[2]This section draws on the discussion in chapters 7 and 8 in George W. Norton, Jeffrey Alwang, and William A. Masters, *Economics of Agricultural Development, World Food Systems and Resource Use*, 2nd ed. (New York: Routledge, 2010).

Labor may be in short supply during planting and harvesting time, but underemployed in other seasons. Seasonality may also bring with it wide fluctuations in food prices, as prices are at their lowest point immediately after the harvest but increase continually from that time until the next harvest. This seasonal fluctuation arises in part because traditional farms often lack adequate facilities to store their crops and must sell them shortly after harvest. The months preceding the next harvest, when prices reach their seasonal high, are often referred to as *the lean season*.

Traditional farms use traditional technologies, which typically rely on few purchased inputs (such as fertilizer, insecticides, and pesticides), grow locally indigenous crop varieties, and operate at low levels of productivity. It is widely accepted that traditional farmers allocate their resources rationally (as defined by economists) and thus do the best that they can with the few resources at their disposal. Hence the famous description by economist Theodore W. Schultz that traditional farmers are "poor but efficient."[3] Traditional farmers may also earn a significant portion of their livelihood by engaging in nonfarm rural activities, such as providing services and engaging in petty trade.

AGRICULTURAL SYSTEMS

Farming systems, characterized by technology, mix of crops and livestock, and the physical environment, vary widely across the developing world. Farming systems tend to be dominated by a small number of crops—typically a staple cereal such as rice, wheat, or corn—and many minor crops or livestock. The major types of farming systems (each of which includes a variety of subsystems) are **shifting cultivation**, **pastoral nomadism**, and **settled agriculture**. Shifting cultivation describes a system in which producers cultivate one area until its fertility is exhausted and then migrate to another plot of land. If the new areas must first be cleared of brush and the remaining brush then burned to clear the fields and increase the nutrient content of the soil, it is termed **slash and burn** agriculture. Economists George Norton, Jeffrey Alwang, and William Masters estimate that shifting cultivation is still practiced on about 15 percent of the world's cultivated area, mostly in Latin America and sub-Saharan Africa.[4] They note that shifting cultivation has been linked to soil erosion.

Pastoral nomadism, as the name implies, describes a farming system in which producers travel more or less continuously. This mobility requires that their production system be based on livestock, which the nomads shepherd across grazing areas. This system can function only in areas of low population density and is thus most frequently found in arid and semiarid agroecological zones, such as in much of the Sahel region of Africa or in the Indian state of Rajasthan. Nomadic groups typically

[3]Theodore W. Schultz, *Transforming Traditional Agriculture* (Chicago: University of Chicago Press, 1964), p. 38.

[4]Norton et al. *Economics of Agricultural Development*," p. 150.

include five or six families traveling together with medium-size herds of livestock, perhaps 25 to 60 sheep and goats or a smaller number of camels or cattle. Increasing population densities and global climate change are particular threats to this type of farming system, which is also associated with environmental damage, resulting from overgrazing.

Settled agriculture includes a variety of systems, which as a group represent the best potential for productivity growth. Some settled agricultural systems are the following:

- *Intensive annual crops*, the most dominant system in terms of total cultivated area, typically concentrating on production of staple cereal crops (wheat, rice, and corn)
- *Mixed farming*, which builds on the interactions of crops and livestock production to manage risk and maintain soil fertility
- *Perennial crops*, generally tree crops such as bananas, coffee, cocoa, along with sugarcane, that produce for a period of years (often in combination with annual food crops)
- *Livestock systems*, producing both dairy and meat products, either through grain feeding (intensive production) or exclusively by grazing (extensive production)

Settled agriculture forms the core of efforts to modernize agriculture. The challenges are numerous, complex, and ever evolving. With cereals demand in developing countries projected to increase by nearly 50 percent between 1997 and 2020, the challenge of getting agriculture moving is urgent.[5] The starting point is a framework for diagnosing the constraints to increased productivity.

DIAGNOSING THE CONSTRAINTS TO AGRICULTURAL DEVELOPMENT

Increasing agricultural output is a high priority for most developing countries. In addition to the obvious benefit of providing greater access to food for domestic consumers, increasing agricultural output also enhances the livelihoods of those working in agriculture and related industries. In many developing countries, the agricultural

[5]Projections of future demand for food take into consideration not only projections of how many people will need to be fed but also projected income levels, which shape both the quantity and the type of food demanded. It is widely documented that demand for meat increases with income, and meat production implies additional demand for grain as feed. Much of this demand will be in China. Mark Rosegrant, Michael Paisner, Siet Meijer, and Julie Witcover, *Global Food Projections to 2020, Emerging Trends and Alternative Futures* (Washington, DC: International Food Policy Research Institute, 2001), p. 58.

labor force is both the largest and the poorest segment of the population. Agriculture also may involve a significant proportion of exports, as we see in such examples as cocoa exports from Ghana and coffee exports from Ethiopia. Yet, despite the substantial efforts (and ample rhetoric) that many developing countries devote to the task of increasing agricultural output, progress has often been halting and difficult. A particularly useful framework for diagnosing the constraints to increasing agricultural productivity was proposed by agricultural economist Arthur Mosher.[6]

The **Mosher framework** compares the *actual* performance of agricultural producers with their *potential* performance. Performance in this framework is represented by the output per hectare (or **yield**) of a particular commodity, often a staple cereal. Yield, even for a particular cereal, may vary widely from farm to farm in a given country. Some farmers may have better-quality land, a more conducive climate, or more regular access to water; other farmers may be closer to roads and have greater access to markets for both inputs and their output or more exposure to agricultural extension agents who teach them improved techniques. Some farmers may simply be better at farming than other farmers. The Mosher framework begins by recognizing these variations and observing the actual distribution of yields across farms, ranking them from highest to lowest. Given this distribution, the approach then addresses the question of why some yields are not higher.

What are the constraints that prevent farmers from producing more output per unit of land? The Mosher framework considers two categories of explanations: technical and economic. Figure 17–1 illustrates the sense in which we can think of technical and economic constraints as ceilings on the levels of crop yield that farmers might achieve.[7] The horizontal axis in Figure 17–1 measures the distribution of land cultivated to a particular cereal crop, and the vertical axis measures yield per hectare for that crop. Curve *a* illustrates the distribution of yields across farms, or the **achievement distribution**. These are the yields actually obtained by farmers, ordered from highest to lowest, as observed in field surveys. (This distribution would be flat if all farmers had the same yields.) Because of the differences across farms and farmers noted earlier, the achievement distribution typically slopes downward.[8]

The technical ceiling, represented by curve *t* in Figure 17–1, describes the maximum biologically possible yield given available technology. This is defined by the yield achieved by experiment stations in various regions where agricultural scientists grow the crop using the best technologies under controlled conditions.

[6]Arthur Mosher, *An Introduction to Agricultural Extension* (Singapore: Singapore University Press for the Agricultural Development Council, 1978).

[7]Figure 17–1 originated with Mosher, *An Introduction to Agricultural Extension*. Our discussion of this framework draws on C. Peter Timmer, Walter Falcon, and Scott Pearson, *Food Policy Analysis* (Baltimore: Johns Hopkins University Press, for the World Bank, 1983), chap. 3.

[8]Because the height of the achievement distribution measures output per hectare and the horizontal axis indicates the land area achieving each yield, the area beneath the achievement distribution measures total production.

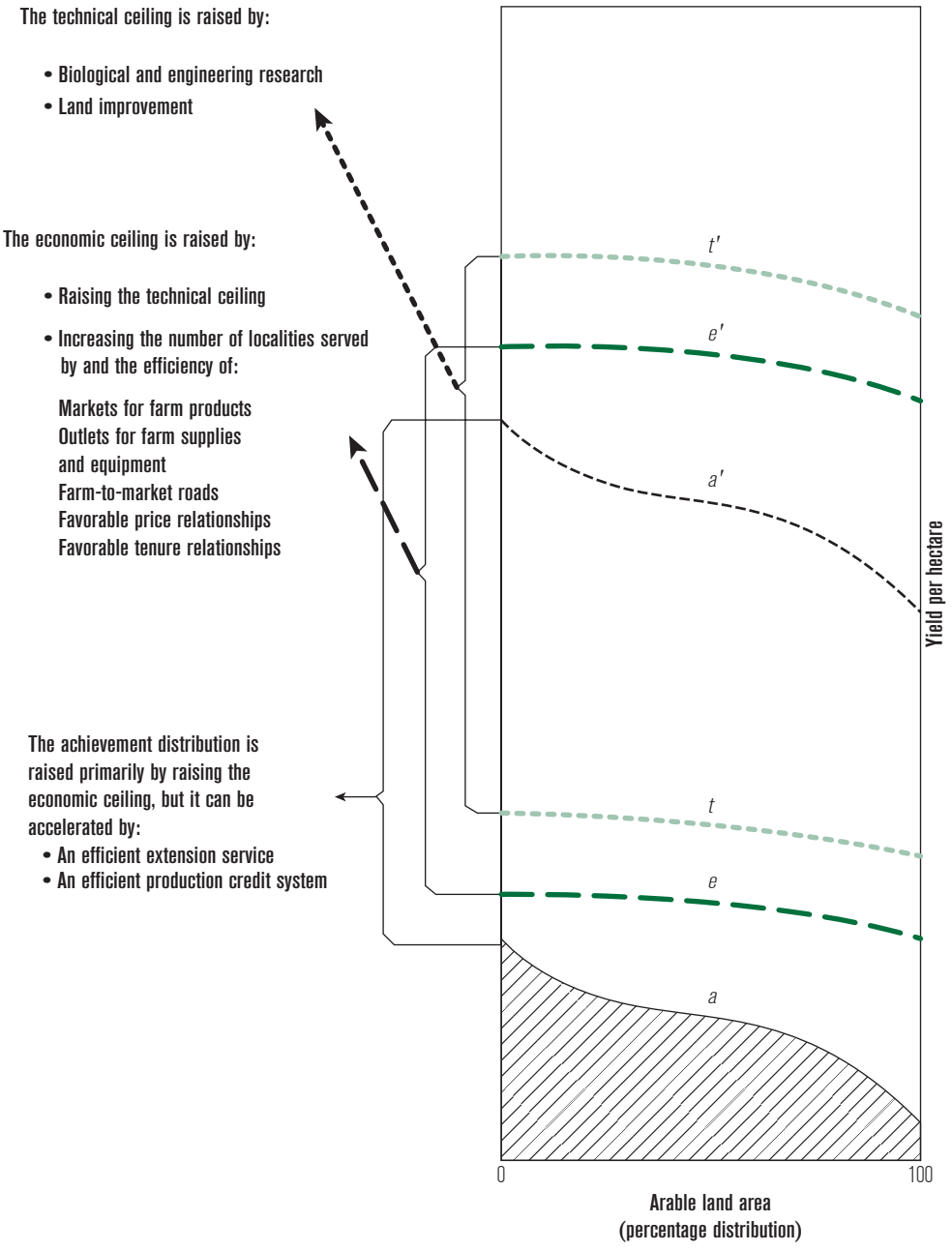


FIGURE 17–1 The Mosher Framework

Source: Figure 3.9 from The International Bank for Reconstruction and Development. The World Bank: *Food Policy Analysis,* 1983. The Johns Hopkins University Press for The World Bank. Reprinted with permission.

Farmers can do no better than the technical ceiling because it represents the maximum yield physically possible in a given environment. The economic ceiling, represented by curve *e*, describes the distribution of yields that are possible when all farmers maximize their profits given the opportunities available to them in a given

policy environment. The economic ceiling cannot lie above what is physically possible (the technical ceiling); how far below the technical ceiling the economic ceiling lies is determined by the policy environment. Limited access to credit and the risks of uncertain weather can keep actual achievements well below the economic ceiling. As a result, the true economic ceiling (unlike the technical ceiling) cannot be observed from field data.[9] The achievement distribution, in turn, can be no higher than the economic ceiling.

Figure 17–1 also describes scenarios under which the achievement distribution and the technical and economic ceilings might be raised, thus creating space in which the achievement distribution might be raised. Improved agricultural technologies, such as improved seed varieties, can raise the technical ceiling from t to t'. This could create the possibility of raising the economic ceiling from e to e' by improving price incentives or improving the connections between farmers and markets or reforming land tenure arrangements. With the economic ceiling raised from e to e' it may then be possible to raise the achievement distribution from a to a'. Improved agricultural extension services or access to credit can contribute to raising the achievement distribution by helping farmers take maximum advantage of their technical and economic circumstances.

Figure 17–2 illustrates various relationships between the technical and economic ceilings and the achievement distribution, which may be observed in practice. *Area A* illustrates a low-productivity environment in which the achievement distribution is tightly constrained by the economic ceiling and in which the economic ceiling itself is tightly constrained by the technical ceiling. In such a setting, farmers cannot be expected to do much better, even if economic policies improve. The binding constraint in this case is the technical ceiling, which may be low owing to a lack of improved seeds. In contrast, *Area B* describes a situation in which the technical ceiling is far from binding but in which farmers are close to doing the best that they can, given what might be an unsupportive economic policy environment or poor institutions for land tenure. In *Area C*, neither the technical nor the economic ceilings are binding for most farmers, but the achievement distribution indicates that only a few farmers may be benefiting from the best available technologies. This diagnosis might point to the need for improved agricultural extension services. In *Area D*, the technical and economic ceilings are high, to the benefit of most farmers. Additional increases in yield under this scenario may require long-term investments in agricultural research to raise the technical ceiling even further.

The Mosher framework is not intended to provide a basis for detailed recommendations regarding technological development or economic policies. This framework

[9]Timmer et al., *Food Policy Analysis*, suggest that a general placement of the economic ceiling is possible by comparing the marginal product of fertilizer (the additional output of cereal resulting from using an additional unit of fertilizer) with the ratio of the grain price to the fertilizer price. If an additional unit of fertilizer will produce five units additional cereals output but the cost of that additional unit of fertilizer is six times the price of grain, then a rational farmer will not use additional fertilizer.

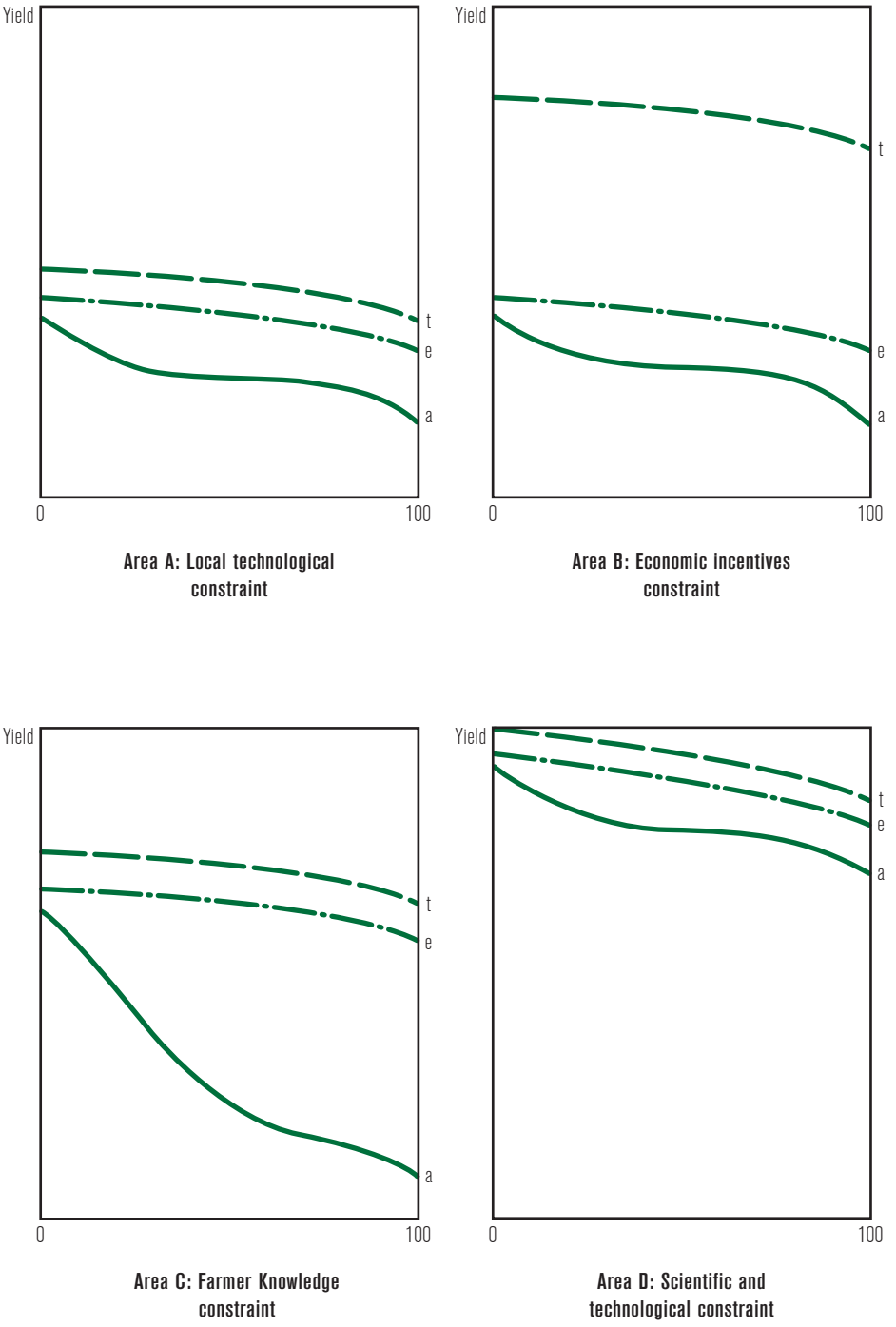


FIGURE 17–2 Possible Scenarios in the Mosher Framework

Source: Figure 3.10 from The International Bank for Reconstruction and Development. The World Bank: *Food Policy Analysis,* 1983. The Johns Hopkins University Press for The World Bank. Reprinted with permission.

provides a flexible diagnostic tool for prioritizing potential initiatives to support agricultural development and for understanding the relationships between those initiatives. The categorization of constraints as pertaining either to technology or to economic policy also provides a springboard from which to consider the challenges presented by both categories in greater detail. We begin by considering possible ways to raise the technical ceiling.

RAISING THE TECHNICAL CEILING

The height of the technical ceiling is defined in part by agroecological conditions and the agricultural potential of different areas. This is of particular concern when considering rain-fed agriculture, in which production is especially sensitive to differences in soil quality and temperature, in addition to rainfall. According to the World Bank, two-thirds of the rural population in developing countries (over 1.8 billion people) live in areas with favorable agroecological potential: about 40 percent within irrigated areas and an additional 26 percent live in areas with reliable moisture.[10] One-third of the population of the developing world (over 820 million people) live in areas with less-favorable potential, where the irrigation is scarce and rainfall is unreliable. These less-favorable areas account for 54 percent of the agricultural area but produce only 30 percent of the total value of agricultural production. Investing in irrigation is one (albeit expensive) way to raise the technical ceiling. A broader and more flexible approach to raising the technical ceiling lies in improving agricultural technology.

In the short and medium run, average technology might improve because more farms adopt improved technologies (a process known as **technology diffusion**), because the least productive farms ceased to exist, or because the best available technologies themselves improve. In the long run, however, average technology can improve only if the best technology is also improving. In this section, we consider various approaches to raising the technical ceiling on agricultural production. We examine the sources of new agricultural technologies and their impact on productivity in developing-country agriculture as well as a model of technological innovation. There has been substantial progress in developing and diffusing improved agricultural technologies, resulting in impressive gains in productivity, yet equally substantial challenges remain. Most of the gains in agricultural productivity over the past 50 years have resulted from the green revolution.

[10]World Bank, *World Development Report, 2008: Agriculture for Development*, (Washington, DC: World Bank, 2007), p. 54.

THE GREEN REVOLUTION[11]

The term **green revolution** refers broadly to the science-based innovations in crop breeding and farming practices, with roots dating from the 1940s, that transformed global agriculture during the second half of the last century. Agronomist Norman Borlaug is often referred to as the Father of the Green Revolution for his early scientific work in Mexico on developing **semidwarf disease resistant wheat** varieties. These new wheat varieties were an early example of **modern crop varieties (MVs)**, or **high-yielding crop varieties**, that were to become the centerpiece of the green revolution. Plants are vulnerable to a wide range of pests, diseases, and parasitic weeds that can severely reduce yields. By cross-breeding appropriate strains of wheat, Borlaug and colleagues were able to fortify the wheat then grown in Mexico against major diseases. In addition, existing wheat varieties tended to grow tall and thin as each plant competed with its neighbors for sunlight. Efforts to increase wheat yields through the application of chemical fertilizers increased the grain production of plants, but the tall thin wheat varieties tended to fall over under the added weight of grain. The creation of semidwarf wheat varieties, which had shorter and thicker stems than the traditional varieties as well as improved responsiveness to fertilizers was a solution to that problem. By the mid-1960s, similarly improved varieties of rice had also been developed and released, along with MV wheat, to farmers in Latin America and Asia.

These research programs were housed at new international agricultural research centers (IARCs). The International Maize and Wheat Improvement Center (known by its Spanish acronym, CIMMYT) was established in Mexico in 1963 (with Borlaug as its founding director), and the International Rice Research Institute (IRRI) was established in the Philippines in 1960. Both centers were supported by large grants from the Rockefeller and Ford Foundations. These were the first two centers of what is now the Consultative Group on International Agricultural Research (CGIAR), a global network of 15 research institutes. While popular accounts of the green revolution tend to focus on the early advances with MVs in the 1960s, a comprehensive review by economists Robert Evenson and Douglas Gollin suggests the green revolution remained in full swing until at least the year 2000.[12]

Part of this extended history of the green revolution is the recognition that the early successes were largely limited to wheat and rice for use in Latin America and Asia. Green revolution advances ultimately included the release of more than 8,000 MVs for 11 major crops, yet progress was uneven across both commodities and countries. Advances for certain semiarid crops, such as sorghum, millet, barley, and root crops such as cassava, did not come until the 1980s. The adoption of modern

[11]This section draws on R. E. Evenson and D. Gollin, "Assessing the Impact of the Green Revolution, 1960 to 2000," *Science* 300 (2003); World Bank, *World Development Report, 2008*.

[12]Evenson and Gollin, "Assessing the Impact of the Green Revolution."

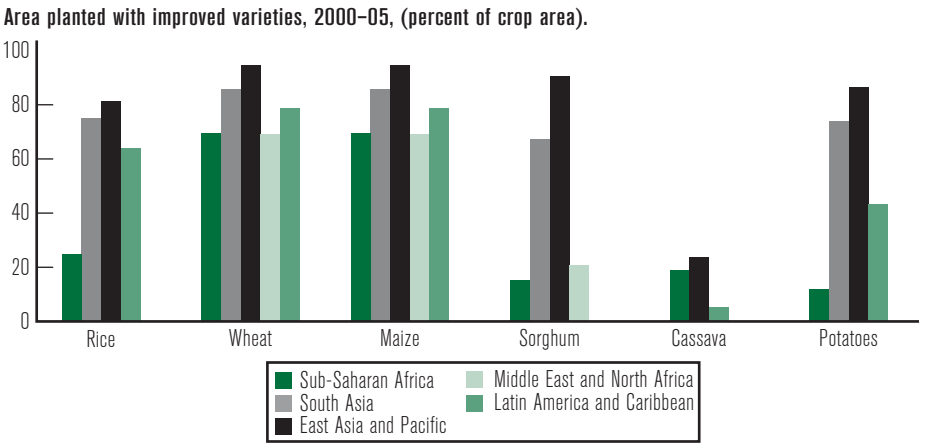


FIGURE 17–3 Modern Variety Diffusion by Crop and Region

Note: Data are provided for the period 2000–05, except for maize in some Sub-Saharan African countries where data are from 1997.
Source: Figure 7.1 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

varieties was widespread but unevenly distributed over time and space. Figure 17–3 illustrates the diffusion of MVs by crop and region. By the early 2000s, adoption rates across commodities in East Asia were uniformly high (exceeding 80 percent, with the exception of cassava, a root crop that plays only a small role in Asian diets). MVs for major commodities had also diffused widely through South Asia and Latin America. Yet, with the exception of wheat, adoption of MVs in sub-Saharan Africa by the early 2000s lagged far behind other developing regions. Adoption of modern varieties for rice and sorghum in sub-Saharan Africa are particularly low in comparison with other regions, even though these two crops are widely grown in the region.

Lagging adoption of MVs in sub-Saharan Africa has several explanations. One problem has been that agroecological conditions in Africa are extremely diverse—sometimes even within a single country. Another problem, particularly before the 1980s, was that the new varieties were engineered to be highly responsive to purchased inputs, such as chemical fertilizers, or (especially in the case of rice) thrived best in irrigated fields. These complementary inputs have been less common in Africa than in Asia or Latin America, limiting the ability of African farmers to benefit from green revolution innovations. Figure 17–4 demonstrates the stark disparities in input intensity across regions. By the year 2002, nearly 40 percent of cropland in South Asia was irrigated, compared to only 4 percent of the cropland in sub-Saharan Africa. Similarly, fertilizer application per hectare in sub-Saharan Africa remained but a small fraction of average fertilizer applications in other regions.

Most of the MVs were the result of genetic research performed at the international research centers. Yet the real measure of success in agricultural research is not the number of varietal releases, but their impact in farmers' fields. In the language of

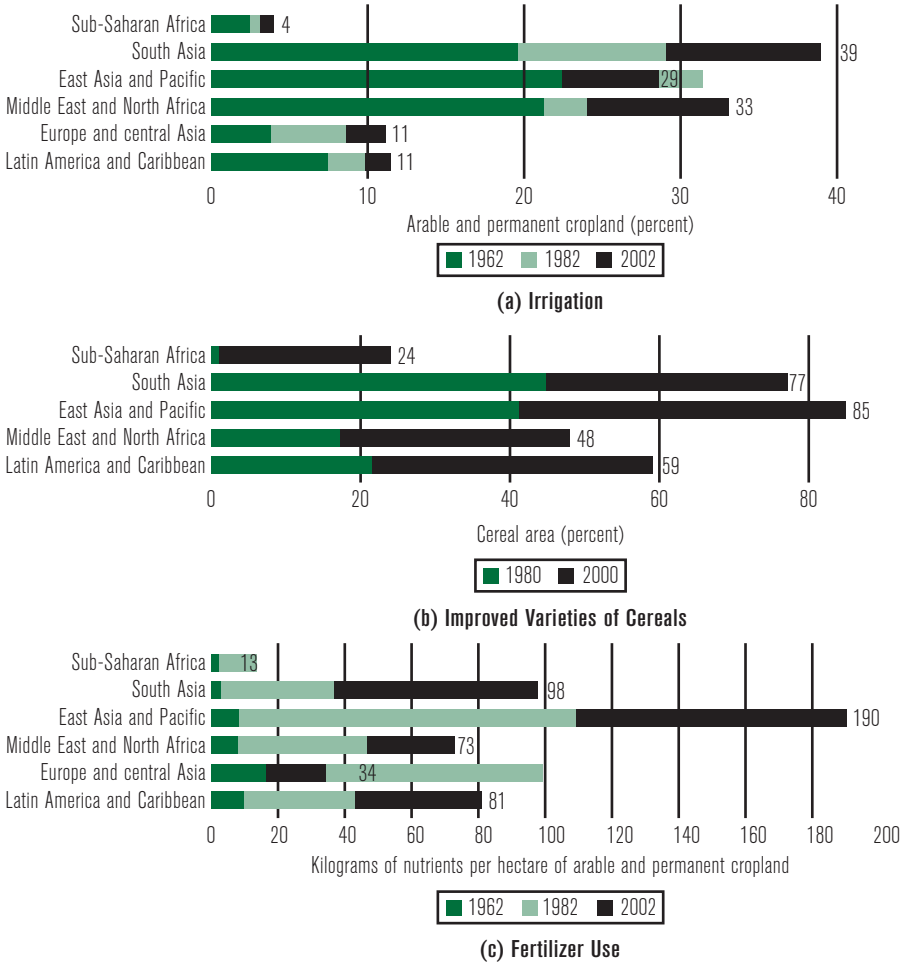


FIGURE 17–4 Use of Agricultural Inputs by Region

Source: Figure 2.2 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

the Mosher framework, new agricultural technologies may raise the technical ceiling, but it raises the achievement distribution only if those technologies are widely adopted and properly used by farmers. Table 17–1 describes growth rates by production, area, and yield for food crops by region, distinguishing between the early and late green revolution periods. The growth rate of output declined significantly from the early to the late green revolution period in every region of the developing world except for sub-Saharan Africa. These changes in the rates of output growth were driven primarily by changes in the growth rates of agricultural area, which fell in all regions except Africa.

Table 17–1 distills the growth rate of crop yield into the contributions of MVs and the contributions to yield of increases in the use of other inputs. During the

TABLE 17–1 Average Annual Growth Rates of Food Production, Area, Yield, and the Contribution of Modern Varieties

	EARLY GREEN REVOLUTION (1961–80)	LATE GREEN REVOLUTION (1981–2000)
Asia		
Production	3.65	2.11
Area	0.51	0.02
Yield	3.12	2.09
MV contributions to yield	0.68	0.97
Other input per hectare*	2.44	1.12
Latin America		
Production	3.08	1.63
Area	1.47	−0.51
Yield	1.59	2.15
MV contributions to yield	0.46	0.77
Other input per hectare*	1.12	1.38
Middle East and North Africa		
Production	2.53	2.12
Area	0.95	0.61
Yield	1.56	1.51
MV contributions to yield	0.17	0.78
Other input per hectare*	1.39	0.72
Sub-Saharan Africa		
Production	1.70	3.19
Area	0.52	2.82
Yield	1.17	0.36
MV contributions to yield	0.10	0.47
Other input per hectare*	1.07	−0.11
All developing countries		
Production	3.20	2.19
Area	0.68	0.39
Yield	2.50	1.81
MV contributions to yield	0.52	0.86
Other input per hectare*	1.98	0.95

*Refers to such factors as fertilizer and machinery. Growth rates of other inputs are taken as a residual.
MV, modern crop variety.

Source: R. E. Evenson and D. Gollin, "Assessing the Impact of the Green Revolution, 1960 to 2000," *Science* 300 (2003).

early years (before 1980), as revealed by the table, MVs played a large role in driving yield growth in Latin America and Asia but made little contribution in other regions. For example, between 1961 and 1980 in Latin America, crop production grew at an annual rate of just over 3 percent, of which nearly 1.6 percent was the result of growth in yield. MVs explained over one-fourth of that 1.6 percent growth in yield. Over the same period, however, MV's accounted for only 0.097 percentage points of the

1.17 percent per year growth rate of yield in sub-Saharan Africa (thus explaining only 8 percent of the growth in Africa's crop yields).

For the developing world as a whole, crop production grew at an average rate of 3.2 percent, of which just under 0.7 percent was the result of growth in an agricultural area and 2.5 percent the results of growth in yield. The use of MVs accounted for 0.5 percentage points of the growth rate in yield, or just over one-fifth. Between 1981 and 2000, the rate of production growth for all developing countries (on average) fell to about 2.2 percent per year, whereas yield growth fell to 1.8 percent. Although yields grew more slowly after 1981, it is notable that yield growth accounted for a larger share of growth in total output post-1981, and the share of yield growth explained by MVs increased to nearly one half. Sub-Saharan Africa was the only region in which food production accelerated in the late green revolution, in part as a result of the substantially increased contribution of MVs in that period.

What would the world have been like by the year 2000 had the green revolution *not* occurred? Economists at the International Food Policy Research Institute (IFPRI; the economic policy center of the CGIAR) used a simulation model to answer this hypothetical question.[13] If no MVs had been released in developing countries, agricultural yields in developing countries would have been around 20 percent lower by the year 2000 than their actual level in that year. Cropped area in both developed and developing countries would have been 3 to 5 percent greater than their actual level, and crop prices would have been 35 to 66 percent higher. Crop production in developing countries would have been 16 to 19 percent lower than they actually were by the year 2000. The consequences of the green revolution for human welfare were enormous. The IFPRI simulation exercise indicates that had the green revolution *not* occurred, the percent of malnourished children in developing countries would have been 6 to 8 percent greater than it actually was (a difference of 32 to 42 million preschool children), and calorie consumption per capita in developing countries would have been about 14 percent lower than it actually was by the year 2000. Norman Borlaug was awarded the Nobel Peace Prize in 1970 for his pioneering contributions to the green revolution.

RECENT TRENDS IN AGRICULTURAL PRODUCTIVITY

Figure 17–5 illustrates trends in cereals yields, first comparing developed and developing countries and then distinguishing between developing regions. It is clear from Figure 17–5a that cereals yields in the developed countries are higher and have grown faster than in the developing countries. Among regions of the developing world, the differences are also striking. Although the various regions depicted in Figure 17–5b attained similar cereal yields in the early 1960s, they diverged quite dramatically

[13]This simulation exercise is summarized by Evenson and Gollin, "Assessing the Impact of the Green Revolution."

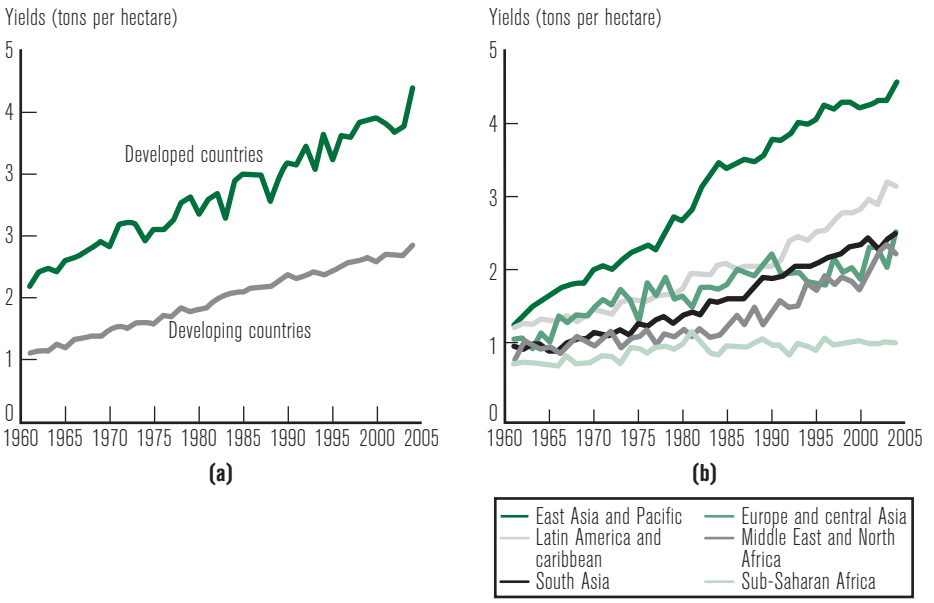


FIGURE 17–5 Cereal Yield Trends by Region

Source: Figure 2.1 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

by 2005. Consistent with Table 17–1, the flat trend in cereals yields for sub-Saharan Africa contrast sharply with the rapid growth experienced in East Asia, where by 2005, average yields were more than four times the level found in Africa.

While the levels of yields in cereals illustrated in Figure 17–5 increased in most regions, the *growth rates* of those increases, while remaining positive, have slowed over time. Figure 17–6 captures this disturbing global trend: The growth rate of yields in the world's major cereal crops (wheat, rice, and maize) has trended strongly downward since the 1980s. This slowdown in productivity growth may reflect that the easy gains of the green revolution's early years have been exhausted. It may also represent declining public expenditures for agricultural research and development (R&D) during the 1990s. In sub-Saharan Africa, for example, public agricultural R&D spending declined in nearly half of the 27 countries with available data, as did the R&D share of agricultural gross domestic product (GDP) for the region as a whole.[14] Yet, it is also important to recognize that agricultural innovations must be continually renewed. The stability of yield gains depends heavily on the resistance of new varieties to various pest and disease risks. As those threats evolve, agricultural technology must strive to stay one step ahead. In summarizing this problem, the World Bank's *World Development Report 2008* makes reference to the Red Queen in Lewis Carroll's *Through the Looking Glass*, who complained, "Now here, you see, it takes all the

[14]World Bank, *World Development Report, 2008.*

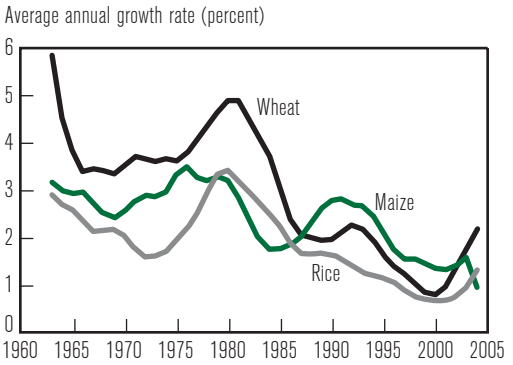


FIGURE 17–6 Growth Rates of Cereals Yields

Source: Figure 2.12 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

running you can do to keep in the same place."[15] That report then notes that one-third to one-half of current investments and crop breeding may be aimed simply at maintaining past gains.

This discussion of agricultural productivity trends highlights the diversity of experience across the developing world, yet a consistent theme is the relative lack of progress in sub-Saharan Africa. Figure 17–7 compares average maize yields with potential maize yields (the so-called **yield gap**) for selected countries in sub-Saharan Africa. In each case, average on-farm achievement falls far short of the technical ceiling implied by the maize yields obtained on demonstration plots, reflecting the best possible conditions

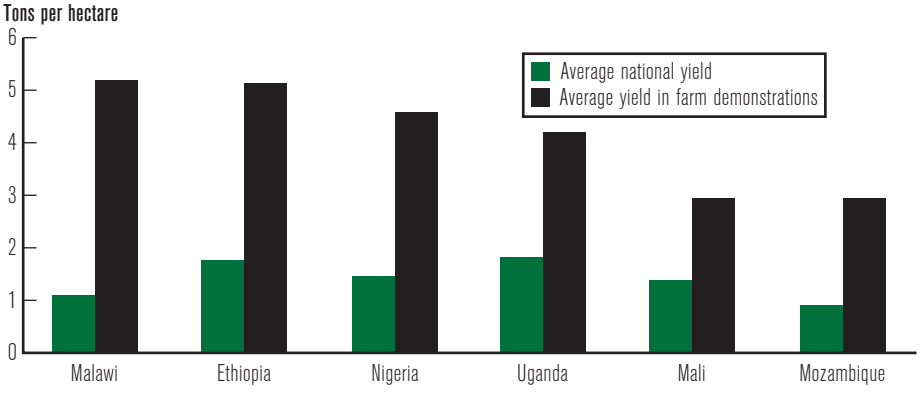


FIGURE 17–7 The Yield Gap for Maize in Selected African Countries.

Source: Figure 2.13 from The International Bank for Reconstruction and Development. The World Bank: *Agriculture for Development,* 2008. Reprinted with permission.

[15]World Bank, *World Development Report, 2008*, p. 161.

and optimal inputs available in each country. Yield gaps in Asian agriculture tend to be much smaller, consistent with the more rapid rate of yield growth seen in Figure 17–5. The problems in sub-Saharan Africa include the greater complexity of its agricultural systems and diversity of agroclimatic conditions (both of which have increased the challenges of creating new technologies for use there), in addition to its relative lack of market infrastructure and lower levels of education. Box 17–1 introduces an important theory about how the kinds of agricultural technologies that are invented depend on economic circumstances, the theory of induced innovation.

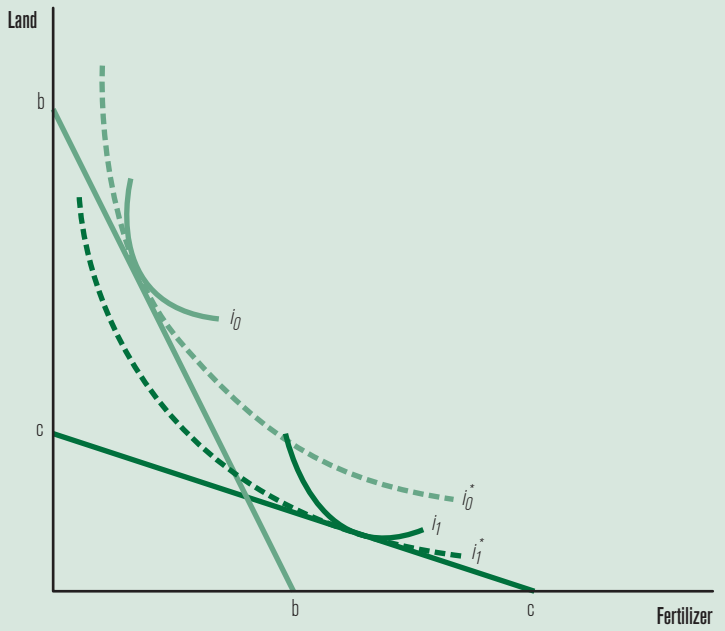
BOX 17–1 A MODEL OF INDUCED TECHNICAL CHANGE IN AGRICULTURE

The green revolution technologies that were so effective in Asia tended to emphasize biological technology. Improved seed varieties that were highly responsive to chemical fertilizers formed the core of the innovations. Biological approaches primarily work to increase output per unit of land and we can think of them as **land-saving technologies**. In contrast, mechanical innovations, such as better tractors and improved harvesting machinery, enable the cultivation of a larger area per worker. We can think of mechanical approaches as **labor-saving technologies**. Is it merely coincidence that land-saving technologies were developed for application in Asia, where labor is abundant and land relatively scarce? The broad idea that links factor endowments to the evolution of technical change is the **theory of induced innovation**.

The notion that new agricultural technologies can economize on either land or labor illustrates that multiple paths of technology development are possible. Advances in mechanical technology are motivated by the desire to reduce labor costs; advances in biological technology are motivated by the desire to increase crop output per unit of land (or animal output per unit of feed). Mechanical technology may be more appropriate in settings in which the supply of labor is constrained; biological technology might be more appropriate in settings in which the supply of land is constrained. The theory of induced innovation, popularized by economists Yujiro Hayami and Vernon Ruttan, addresses the question of which among these alternative paths of technology development is *actually* followed. Its answer is that technology evolves in response to changes in relative factor prices, economizing on the relatively scarce (and hence more costly) factors of production.[a]

[a]This theory thus portrays technical change as an endogenous process, in which realized technologies are determined by economic circumstances.

The figure below illustrates the process of induced innovation for the case of biological technology. The model builds on the concept of a **production isoquant**, which describes the various combinations of inputs (in this case, land and fertilizer) that can be combined to produce a given quantity of output.[b] The induced innovation model posits that at any given time, a large family of potential technologies could be invented. Each of these technologies can be represented by its own isoquant. We will consider the evolution of technology between two points in time (the initial point being designated "time zero," as indicated by the subscript 0, and the other point in time similarly designated as "time 1", as indicated by the subscript 1). In the figure, the isoquant i_0 might represent a particular crop variety, which could be grown in a given quantity by combining fertilizer and land in any of the combinations that lie along i_0. The more elastic (*dashed*) isoquant, i_0^*, is termed the **innovation possibility curve**.[c] In this model, the innovation possibility curve traces the envelope (or boundary) of the entire family of potential crop technologies that might be invented at that



[b]We revisit this construct and provide a more detailed exposition in the discussion of producer price incentives later in this chapter.

[c]The elasticity of an isoquant refers to the ease with which producers can substitute one input for another in response to changes in relative input prices. Graphically, an inelastic isoquant will have more of an "L" shape (indicating little scope for substituting one input for another), whereas a more elastic isoquant will curve more gradually.

time zero. The line *bb* is the **isocost line**. Every point on *bb* reflects a combination of fertilizer and land use with the same total cost. Its slope reflects the relative price of fertilizer to land. The specific crop technology that gets invented in the first period (i_0) is the one that minimizes input costs with relative prices of fertilizer and land indicated by the slope of line *bb*.

Technological progress is represented by a shift over time of the innovation possibilities curve toward the origin of the graph in the figure. This reflects the idea that technological progress permits the production of a given quantity of output using a smaller quantity of inputs. Thus by time 1 in this graph, the innovation possibilities curve has shifted inward from i_0^* to i_1^*. Suppose that between time 0 and time 1 fertilizer production increases, thus reducing the relative price of fertilizer to land. This is reflected in the flatter isocost line *cc*. In response to this change in relative factor prices, the induced innovation model suggests that a new technology (depicted by isoquant i_1) is developed, a new crop variety that is more responsive to fertilizer (perhaps semidwarf wheat). This new technology will lie along the envelope of technologies that could potentially be developed at time 1, summarized by the new innovation possibilities curve, i_1^*. This particular crop variety is invented because it minimizes production costs given the new set of relative factor prices. The key intuition is that the technology *actually* invented from among potential innovations is the one that most efficiently economizes on the factor of production that has become more costly.

Source: This box draws on Yujiro Hayami and Vernon Ruttan, *Agricultural Development: An International Perspective* (Baltimore, MD: Johns Hopkins University Press, 1985).

RAISING THE ECONOMIC CEILING

The height of the economic ceiling in the Mosher framework is determined by a combination of economic incentives, farmers' access to markets, and institutions such as the rules governing land tenure. We begin by considering the impact of prices on farm-level decision making and resource allocation. The goals of agricultural development strategies may include increasing the sector's output or promoting rural welfare. In either case, a price environment that provides positive incentives to producers and makes agriculture a profitable enterprise typically calls for relatively high food prices. Such strategies are inevitably constrained by the harmful implications of high food prices for consumers, many of whom may be poor and allocate large budget shares to food, the poorest of whom may be highly vulnerable to malnutrition.

The conflicting interests of producers and consumers in regard to food prices are a central challenge for policy makers, one that is sufficiently pervasive to be termed by economists C. Peter Timmer, Walter Falcon, and Scott Pearson "the food policy dilemma."

FOOD PRODUCTION ANALYSIS[16]

Designing policies to promote agricultural development first requires an understanding of how farmers respond to price incentives. This section addresses three dimensions of decision making by farmers: what mix of crops to produce, what combination of inputs to use in production, and what quantity to produce. Poor farmers, they have to make all these decisions simultaneously. We have the luxury to examine them in turn, and then see how they combine to describe the supply curve for food. We assume in each case that farmers allocate their resources with an eye toward maximizing profits, which implies that their choices equalize marginal costs and marginal benefits.

WHAT TO PRODUCE? THE PRODUCT–PRODUCT DECISION

Farmers in developing countries rarely produce only one crop; indeed in sub-Saharan Africa it is common to see four or five different crops simultaneously planted in the same field (an approach known as **intercropping**). Whether multiple crops are grown in the same field or in different fields, farmers must decide on an optimal crop mix. Economists model this decision by considering how the opportunities made available by nature interact with market incentives. Figure 17–8 illustrates this interaction between markets and nature with a **production possibility frontier** and an **isorevenue line**. In practice farmers must determine their crop mix by choosing from among a large number of commodities. We simplify this problem by reducing it to a choice between only two commodities, for instance corn and beans. In this case, the production possibility curve, *IKLE*, describes all of the different possible combinations of corn and beans that a farmer might grow in a given field in a single season. If the farmer were to devote all of his or her resources to the production of beans alone, output would be *OE*. Similarly, if the farmer devoted all resources to the production of corn, output would be *OI*. Typically however farmers will grow some of each crop. The slope of the production possibility frontier, known as the **marginal rate of technical substitution**, thus changes at every point along the frontier. At any given point that slope equals $\Delta C/\Delta B$, where the Greek letter Δ (capital delta) indicates changes in the quantity produced of corn (C) and beans (B). The production possibilities frontier bows outward to reflect the increasing marginal cost of produc-

[16]This section draws on Timmer et al., *Food Policy Analysis*, chap. 3.

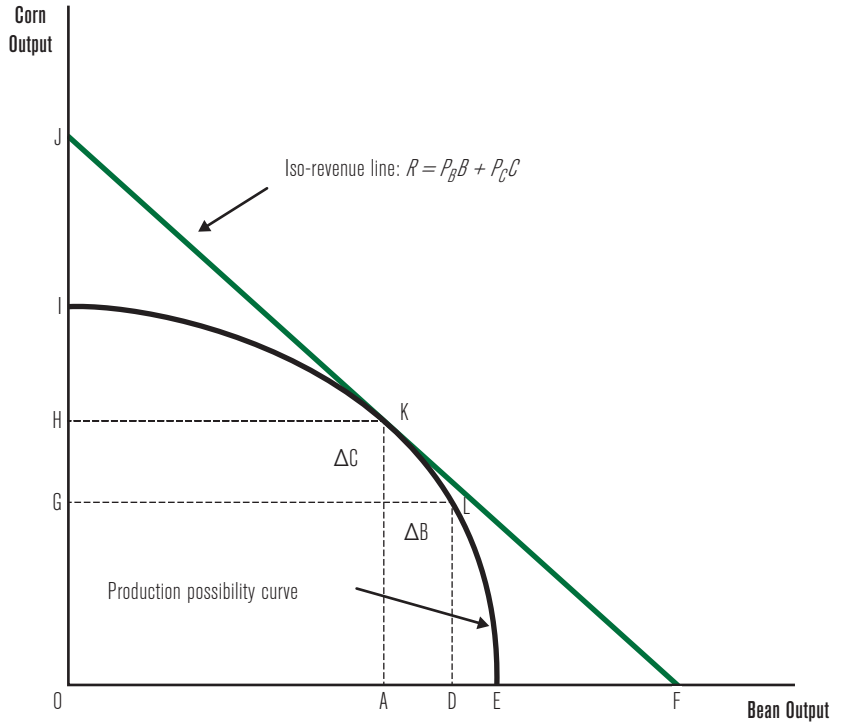


FIGURE 17–8 **The Product–Product Decision**

ing either commodity. This implies that a farmer who was initially growing a crop mix that included lots of beans but little corn might be able to increase corn output significantly by sacrificing a relative small quantity of bean production.

The production possibility frontier indicates all of the **technically efficient** combinations of corn and bean output. A crop mix indicated by a choice inside the production possibility frontier would be inefficient and would indicate a yield gap, as discussed earlier. A technically efficient crop mix is necessary for profit maximization (or **economic efficiency**), but it is not sufficient. The profit maximizing combination of corn and bean output can be determined only with reference to how markets value each of those crops. The total revenue generated by any particular combination of corn and bean output depends on their respective prices, P_C and P_B. The revenue generated by producing either commodity equals the price of that commodity times the quantity produced, and total revenue (R) is the sum over all of the commodities: $R = P_B B + P_C C$. For given quantities of corn and beans, this equation describes a constant level of total revenue and is known as an isorevenue line. As represented in Figure 17–8, the isorevenue line has a slope equal to $-P_B/P_C$.[17] The distance of the

[17]We can see that the slope of the isorevenue line is described by these relative prices by solving the revenue function for C.

isorevenue line from the origin in the graph indicates the level of revenue along any given isorevenue line, with higher lines indicating higher revenue. A profit-maximizing producer will choose a crop mix that places him or her on the highest possible isorevenue line, where *possible* is defined by the production possibility frontier. This is indicated in Figure 17–8 by the combination at point *K*, where the farmer produces *OA* quantity of beans and *OH* quantity of corn. The highest possible isorevenue line that the farmer can attain given the limitations imposed by nature and technology is the one that is just tangent to the production possibility frontier. This means that at point *K* the slope of the isorevenue line equals the slope of the production possibility frontier, that is $\Delta C/\Delta B = -P_B/P_C$. This is the only crop mix that is economically efficient among the set of technically efficient combinations defined by the production possibility frontier.[18]

We can see that the crop mix indicated by point *K* satisfies our requirements for profit maximization, that marginal revenue equals marginal cost by considering the implications of moving along the production possibility frontier from a point like *L* to *K*. This move would require forgoing bean production in the quantity ΔB, incurring marginal cost ΔB^*P_B, to produce additional corn of the amount ΔC, generating marginal revenue ΔC^*P_C. Setting marginal cost equal to marginal revenue would give us the expression $\Delta B^*P_B = \Delta C^*P_C$. Rearranging this expression, dividing both sides first by ΔB and then by P_C, gives us $\Delta C/\Delta B = -P_B/P_C$. Thus the tangency between the isorevenue line and the production possibility frontier meets our criterion for profit maximization.

Beans are an important source of protein. Suppose policy makers sought to increase the availability of protein for their population. This framework shows that they could do so by increasing the price of beans relative to the price of corn. This would increase the slope of the isorevenue line, shifting the tangency with the production possibility frontier to a point such as *L*, where bean output would be *OD* rather than *OA*. Similarly, many countries have recently sought to diversify their agricultural exports to include nontraditional products such as fresh vegetables. This product–product framework demonstrates that governments might promote such diversification by intervening to raise the relative prices of those nontraditional export goods. This framework should also help policy makers recognize that such a change of incentives designed to increase output of one commodity generally comes at the cost of reduced output of something else. In some cases, reducing output of something else might be the goal. Efforts to reduce poppy cultivation in Afghanistan, for instance, need to consider the relative prices local farmers face for poppy versus wheat, which is their primary alternative.

[18]In terms of the Mosher framework, we could think of point *K* as indicating a point on the achievement distribution that touches both the economic and the technical ceilings.

HOW TO PRODUCE IT? THE FACTOR–FACTOR DECISION

In addition to deciding what mix of crops to grow, farmers must also determine the least costly combination of inputs needed to produce a given quantity of output. This too is a multidimensional problem in that farmers face choices over multiple inputs, including land, labor, capital, fertilizers, water, pesticides, and insecticides. Again we simplify this problem by reducing it to two dimensions, capital and labor. Figure 17–9 illustrates farmers' factor–factor decision problem. As before, the optimal choice depends on the interaction between the constraints imposed by nature and those imposed by the market. The curve *MDEN* in Figure 17–9, known as an isoquant, describes the different combinations of capital and labor that farmers may combine to produce a given quantity of output of a particular crop (in this illustration, 100 kilograms of rice). The slope of an isoquant changes at every point on the curve, and is given by $\Delta L/\Delta K$, where L is the quantity of labor and K is the quantity of capital employed in production. In practice, farmers may face a range of discrete choices of technique in producing that 100 kilograms of rice. Point *P* on the isoquant employs lots of capital and little labor and may indicate a production system built around combine harvesters or other large farm machines. In contrast, point *D* on the isoquant employs lots of labor and little capital and may thus rely on hand labor using small tools.

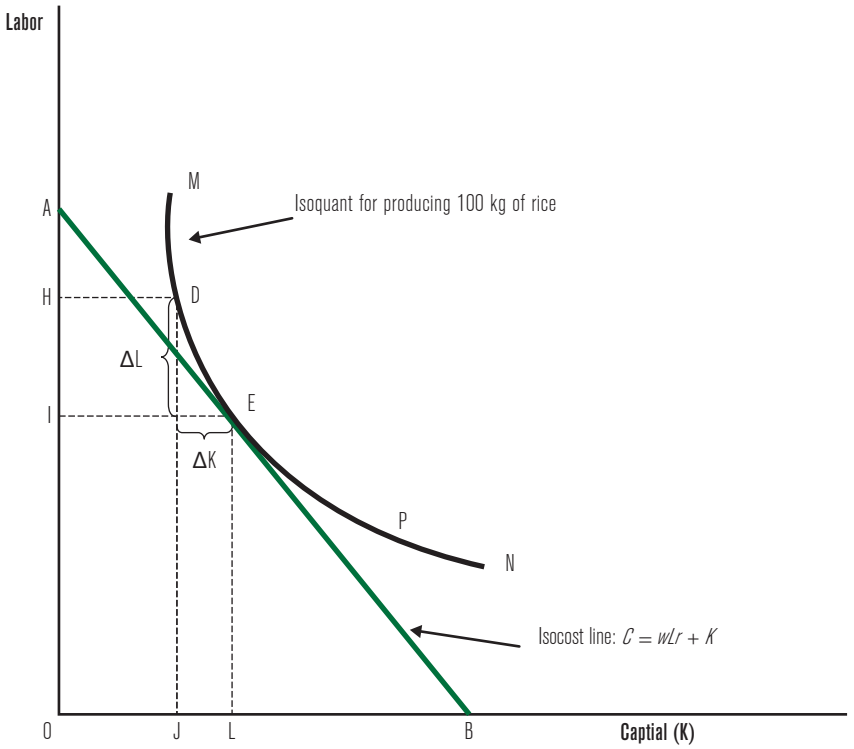


FIGURE 17–9 **The Factor–Factor Decision**

Every point on a given isoquant indicates a technically efficient combination of capital and labor for producing that quantity of output. Any combination of capital and labor that lies above the given isoquant would be inefficient because the same quantity of output could be produced using less of one or both inputs. Only one point among this set of technically efficient combinations of inputs is economically efficient, defined here as the minimum cost combination of inputs capable of producing a given quantity of output. Finding this optimal combination requires consideration of how the market values capital and labor. If factor markets are competitive, labor is paid a wage that reflects its marginal product. Similarly, producers must also pay for the capital (machinery) they employ. If there is a market for machinery, then the cost of renting a given machine is called the **rental rate of capital** and is analogous to the wage paid to labor. If the hourly price of labor is its wage (w) and the hourly price of capital is the rental rate paid for its use (r), then the cost of hiring a given number of workers is w^*L and the cost of hiring a given quantity of capital services is r^*K, then the total cost (C) of producing 100 kilograms of rice is given by $C = wL + rK$. This cost line is called an **isocost** line because the total cost of production is the same at every point on the line. The isocost line in Figure 17–9 has a slope equal to $-r/w$, the relative factor prices.

A cost-minimizing producer will want to choose the least cost combination of capital and labor that still enables the production of a given quantity of output. In Figure 17–9, this choice corresponds to choosing the lowest isocost line that permits production of 100 kilograms of rice. This is the isocost line that is just tangent to the isoquant, point E in the graph. This unique tangency means that the slope of the isoquant is equal to the slope of the isocost line, or $\Delta L/\Delta K = -r/w$. Does this point satisfy our criterion for profit maximization that marginal cost equal marginal revenue? Imagine moving from the input mix indicated by point D to point E. Doing so would imply substituting capital in the amount ΔK for labor in the amount $-\Delta L$. The marginal cost of this substitution is $-\Delta K^*r$ and the marginal benefit (savings) is ΔL^*w. Equating marginal cost and marginal benefit in this case implies that $-\Delta K^*r = \Delta L^*w$. Rearranging these terms, as above, gives us $\Delta L/\Delta K = -r/w$, indicating that the tangency of the isoquant and isocost line describes the profit-maximizing (cost-minimizing) mix of capital and labor to produce 100 kilograms of rice.

This framework also gives rise to a number of interesting policy applications. Is farm mechanization always appropriate? This framework suggests that if labor in a given country is abundant (and, by extension, inexpensive) and capital is scarce (and expensive), then a capital-intensive production technique may not be economically efficient. If policy makers want to increase rural employment, then policies such as overvalued exchange rates (that make imported capital artificially cheap) or ceilings on interest rates (potentially reducing the rental rate on capital below its equilibrium value) would be a bad idea. Such interventions would reduce the slope of the isocost line, creating an incentive for producers to shift their choice of technique in Figure 17–9 from point E toward a point like P. This would motivate an inefficient

allocation of resources (adding to unemployment in this case) if these altered factor prices did not reflect society's true opportunity cost of labor and capital.

HOW MUCH TO PRODUCE? THE FACTOR–PRODUCT DECISION

Having decided what to produce and how to produce it, farmers must still decide how much to produce of any given crop. This relates directly to the decision of how much of any given input to use in production. The physical relationship that maps the quantity of inputs used to the quantity of output produced (once again, nature's constraints) is represented by the production function. In presenting the Solow growth model in Chapter 4, we introduced a production function that represented the entire economy's output. The production function in Figure 17–10 has all of the same technical characteristics as the previous ones, but now describes output for a particular crop. We can consider a two-dimensional relationship between output and a single input if we are clear that the resulting figure implicitly holds all other inputs at a constant level. A production function such as curve *EJKP* in Figure 17–10 thus maps the quantity of fertilizer used to rice yield per hectare (for given fixed quantities of all other inputs). The curve *EJKP* reflects **diminishing marginal returns** to fertilizer use. This implies that an additional unit of fertilizer applied when fertilizer use is low, such as ΔF in Figure 17–10, may bring forth a relatively large output response in rice, ΔR. Yet, the increase in rice resulting from additional units of fertilizer grows

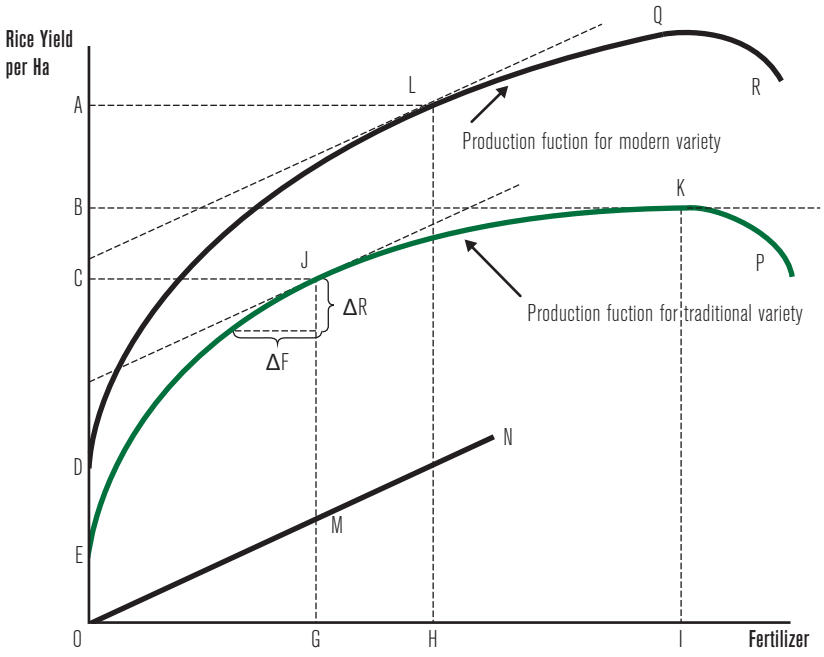


FIGURE 17–10 **Agricultural Production Functions and the Factor–Product Decision**

smaller with each incremental unit of fertilizer applied and ultimately may reduce rice output (as happens beyond point *K* in the graph) if too much fertilizer is applied. These diminishing returns to fertilizer are reflected in the decreasing slope of the production function as fertilizer use increases. The slope of the production function at any point, $\Delta R/\Delta F$, is the **marginal product** of fertilizer.

The maximum rice yield in Figure 17–10 is attained at point *K*, where fertilizer use is *OI*. Yet, the presence of diminishing returns to fertilizer makes it unlikely that *OI* would be the profit-maximizing quantity of fertilizer. As in the previous cases, we must consider the interaction between what nature makes possible (for a given technology) and the economic incentives signaled by the market. In this case, we must consider the market prices for rice and fertilizer. The ratio of these prices is the rate at which farmers can exchange rice for fertilizer. In Figure 17–10, the slope of line *ON* is given by the relative prices of fertilizer and rice, P_F/P_R (increases in the price of fertilizer make this line steeper, whereas increases in the price of rice make it flatter). How much fertilizer would a profit-maximizing farmer apply (and hence how much rice would he or she produce)? So long as the value of the additional rice produced by applying additional fertilizer is greater than the cost of that additional fertilizer, it pays to use more fertilizer.

In Figure 17–10, line *ON* reflects the cost of production while the production function *EJKP* reflects revenue. Profits are maximized at the point at which the difference in height between the production function and the cost function are greatest, which we find by shifting the line *ON* upward (in a parallel manner) until it is just tangent to the production function (which happens at point *J*). A profit-maximizing farmer will thus apply quantity *OG* of fertilizer and produce quantity *OC* of rice. It is only at point *J* that the slope of the production function (the marginal product of fertilizer) exactly equals the relative price of fertilizer to rice, or $\Delta R/\Delta F = P_F/P_R$. Applying the criterion that marginal cost equal marginal revenue, we must compare the value of the additional rice produced by an additional unit of fertilizer with the cost of that additional unit of fertilizer. With diminishing returns, the value created by each additional unit of fertilizer is less than that created by the previous unit. The marginal revenue created by increasing rice production is $\Delta R \times P_R$, and the marginal cost of using more fertilizer is $\Delta F \times P_F$. Setting these terms equal to one another and rearranging as in the previous examples, we find that the profit-maximizing combination of fertilizer use and rice output occurs precisely where $\Delta R/\Delta F = P_F/P_R$, our expression for the tangency at point *J*. In practice, most farmers tend to hedge against weather and price risk by applying less than the optimal quantity of fertilizer (even if they could afford the optimal quantity).

In Chapter 3 we presented a sources of growth analysis and represented growth in total factor productivity (TFP) as a vertical shift of the production function, leading more output to be produced from a given quantity of inputs. The same concept applies to the agricultural production function depicted in Figure 17–10. TFP growth in agriculture would be reflected in a vertical shift of the production function in Figure 17–10, to curve *DLQR*. This increase in agricultural TFP reflects that a given quan-

tity of fertilizer application results in a higher level of rice output per hectare than previously. Indeed, if rice is responsive to fertilizer, the optimal allocation of rice may increase with productivity growth, from *OG* to *OH*, even though the relative price of fertilizer to rice is unchanged.

As in the previous farm-level decision models, important policy implications arise. Governments can intervene to encourage increased food production by reducing the relative price of fertilizer to rice (or the relative price of inputs to outputs more generally). Doing so flattens the relative price line, leading to a tangency with the production function with increased fertilizer being used to produce increased rice. The choice regarding which policies to deploy is complex. Governments can lower this price ratio either by increasing the price of rice or by reducing the cost of fertilizer (or both). Either choice is likely to be both expensive and controversial. In recent years several countries have revived the strategy of subsidizing fertilizer, a previously common approach that countries were encouraged by such organizations as the World Bank to abandon as being too costly, economically inefficient, and ultimately unsustainable. (See Box 17–2 for a discussion of Malawi's renewed fertilizer subsidy scheme.) Conversely, any strategy that involves increasing food prices to incentivize farmers immediately brings us back to the food policy dilemma when we consider the impact of higher food prices on consumers, especially poor ones.

BOX 17–2 FERTILIZER SUBSIDIES IN MALAWI

"Ending Famine, Simply by Ignoring the Exports," so claimed a front page headline in the *New York Times* in describing a controversial fertilizer subsidy program in the southern African country of Malawi.[a] Malawi had long suffered from high rates of chronic hunger, and the country's average maize yields were just over one-tenth the average level in the United States. In 2007, however, Malawi attained a record-breaking maize harvest. According to the *New York Times*, "Farmers explain Malawi's extraordinary turnaround—one with broad implications for hunger-fighting methods across Africa—with one word: fertilizer." During 2004, Malawi experienced both a severe hunger crisis and a contentious presidential election in which the main parties competed in their claims about how much each would do to support farmers. A parliamentary committee recommended a nationwide program of fertilizer subsidies, which President Bingu wa Mutharika announced in June 2005.

Two initial points of contention were whether the fertilizer subsidy program would be targeted (and if so, toward which income groups) and whether it would

[a] *New York Times*, December 2, 2007.

include tobacco, one of Malawi's chief exports, along with maize, its primary consumption crop. Initially, the government decided to include only maize producers, who were unable otherwise to purchase fertilizer, so long as they were productive farmers. The program was not targeted at the smallest, least-able producers. A universal (untargeted) subsidy program was rejected by the president as being too expensive, but he later bowed to political pressure and included both maize and tobacco producers in a universal subsidy program.

The plan was to include both fertilizer and improved maize seed. In its first year, the program excluded private-sector fertilizer suppliers, but several large suppliers subsequently were included. The program was built around a voucher scheme, in which qualifying farmers would receive coupons for fertilizer, entitling them to purchase a 50-kilogram bag of fertilizer for US$6.50, or just over one-quarter of the full cost (with the government paying the remainder). During the 2005–06 growing season, the program distributed about 175,000 metric tons of fertilizer and 4,500 metric tons of improved maize seed, costing $91 million.

Malawi's record maize harvest in 2007 resulted, not only from this seed and fertilizer distribution program but also from quite favorable weather. The *New York Times* reported, "In the hamlet of Mthungu, Enelesi Chakhaza, an elderly widow whose husband died of hunger five years ago, boasted that she got two ox-cart-loads of corn this year from her small plot instead of half a cart." Malawi's maize production in 2004–05 had fallen to 1.23 million metric tons from 1.61 million the previous year. In 2005–06, the first year of the subsidy program, output increased to 2.58 million metric tons, and to 3.44 million in 2006–2007.

Adoption of the fertilizer subsidy program created a dilemma for the government of Malawi, which had counted on restoring donor confidence as the country recovered from a severe economic crisis. The donor community in general looked unfavorably on large subsidy schemes as being costly, unsustainable, and economically inefficient. The strongest donor opposition came from the International Monetary Fund (IMF) and U.S. Agency for International Development (USAID). Their concern, in addition to the economic inefficiency of subsidies, was that the scheme threatened to undermine Malawi's nascent private fertilizer marketing sector, which had emerged after the end of the government's fertilizer monopoly in the early 1990s. Other donors, such as the United Kingdom, the World Bank, and the European Union were skeptical but willing to work with the government to minimize the potential negative effects of the program. In contrast, the United Nations and many nongovernmental organizations (NGOs) were openly supportive of Malawi's fertilizer subsidy program.

Precisely how much of the increased maize harvest of 2007 resulted from the fertilizer program and how much from that year's favorable rainfall remains unclear. One of the few formal attempts to estimate the subsidy's contribution

to increased maize production concluded that Malawi's maize harvest was 15 to 22 percent greater as a result of the program.

Donor attitudes toward input subsidies may have softened in response to this experience. Talk now is of "smart subsidies." As economist Derek Byerlee, a co-director of the team that produced the World Bank's *World Development Report 2008: Agriculture for Development*, notes, "The World Development Report takes a pragmatic position, favoring 'smart subsidies' in select circumstances. Smart subsidies should be transparent, well targeted to the poor . . . , should help jump-start agricultural input markets and should be part of a comprehensive strategy to improve agricultural productivity." Commenting specifically on the Malawian experience, Byerlee continues, "The Bank's position is that the program should be targeted to achieve as large a set of development gains as possible. . . . Smart subsidies are important but need to be weighted against other priorities in national budgets. For example, Malawi only spends 3% of its agricultural budget on R&D—an important source of future productivity increases."[b]

[b]Interview with the World Bank Meetings Center, December 18, 2007, available at http://blogs.worldbank.org/meetings/discuss/world-development-report-2008, accessed February 2012.

Source: This box draws largely on two sources: Blessings Chinsinga, "Reclaiming Policy Space: Lesons from Malawi's 2005/2006 Fertilizer Subsidy Programme," Futures Agricultures and University of Malawi, July 2007; I. Minde, T. S. Jayne, E. Crawford, J. Ariga, and J. Govereh, "Promoting Fertilizer Use in Africa: Current Issues and Empirical Evidence from Malawi, Zambia, and Kenya," International Water Management Institute, Pretoria, South Africa and U.S. Agency for International Development, November 2008.

The expectation that higher food prices motivate increased food production ties together each of the dimensions of farm-level decision making that we have discussed. Economists summarize these multiple dimensions by considering the relationship between changes in price and changes in supply for a given crop or for agriculture in general, a relationship known as farmers' **supply response**. Figure 17–11 illustrates the theory. Figure 17–11a summarizes the relationship between input and output in terms of total revenue and total cost of production. For a given price of rice, total revenue increases by the same amount for each additional unit of output. The total revenue function (0*B*) in Panel a is thus linear with a slope equal to the unit price of rice, P_0. The total cost of producing rice also increases with the quantity produced. However the diminishing-returns characteristic of the production function implies that the total cost function (0*A*) curves upward as illustrated. Because profit is the difference between total revenue and total cost, profit is maximized in Figure 17–11a at the level of output where the vertical difference between the height of the total revenue line and the total cost curve is greatest. As in the example of the production function, we can find this point geometrically by

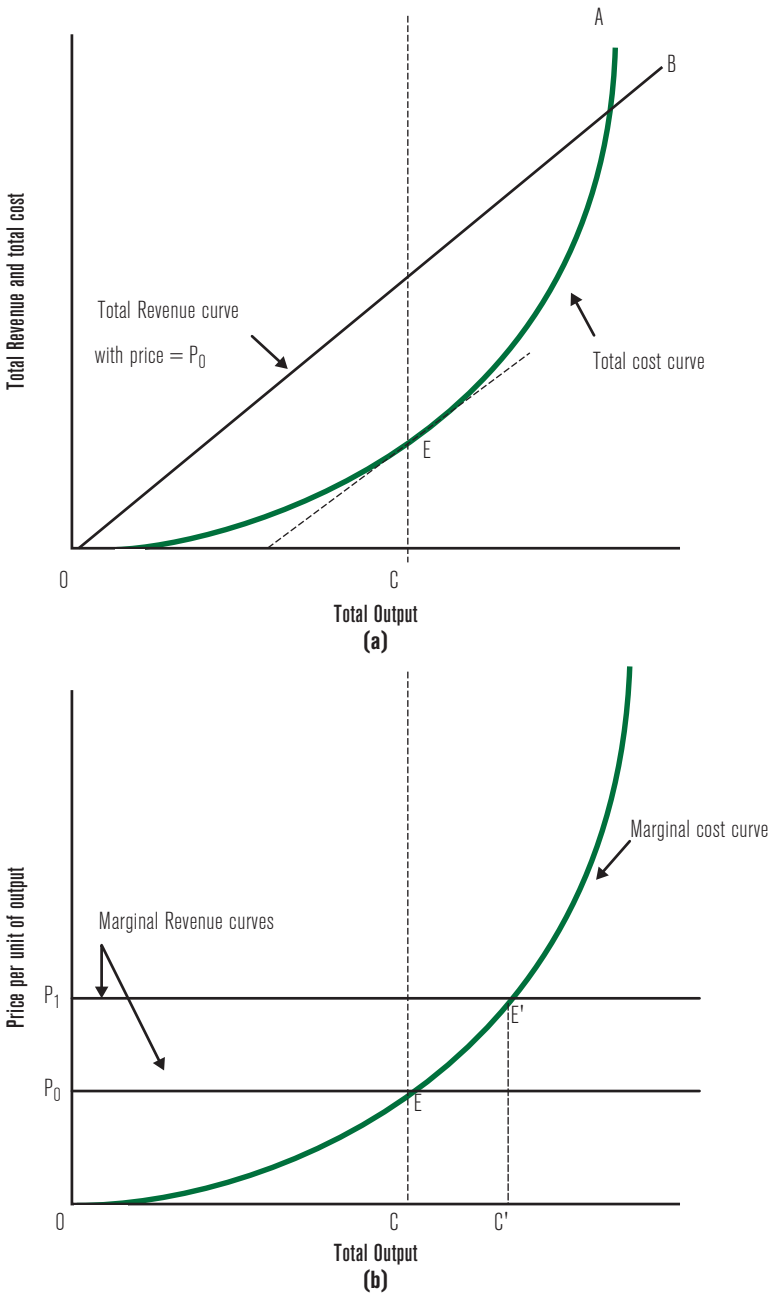


FIGURE 17–11 The Relationship between Output and Price

locating the tangency point between the total cost curve and a line parallel to the total revenue function. This tangency is at point *E*, implying that 0*C* is the profit maximizing quantity of rice.[19]

[19]We can think of a subsistence farmer tending to his or her own small plot with little more than a few hand tools as having no fixed costs, in which case the *y*-intercept of the total cost curve in Figure 17–11 would be zero (as drawn). A larger producer, for instance with a significant capital stock, may have fixed costs. In that case, the total cost curve would have a positive intercept, but the equilibrium condition would still be $MR = MC$.

Figure 17–11b demonstrates that rice production in the quantity 0*C* satisfies our condition for profit maximization. Panel b simply summarizes the relationships in panel a by recording the slopes of the total revenue and total cost functions at each level of output. Bear in mind that the slope at any point of a "total" function defines the analogous "marginal" function. Because the total revenue function is linear, its slope is constant so the marginal revenue function in Figure 17–11b is simply a horizontal line the height of which measures the initial price of rice, P_0. Because the total cost curve bows upward, its slope increases as a function of total output. The marginal cost curve in panel b records the slope of the total cost curve at each level output and thus also slopes upward. Figure 17–11b directly illustrates that total output equal to 0*C* is the profit-maximizing level because that is where marginal revenue equals marginal cost. If the price of rice were to increase from P_0 to P_1, the marginal revenue line would shift up as illustrated and intersect the marginal cost curve at E' instead of point *E*. At this higher price, the profit-maximizing quantity of output would increase to $0C'$. The marginal cost curve thus defines the supply curve.

Based on these relationships, we can summarize farmers' supply response to increased prices by the price elasticity of supply for rice:

$$\varepsilon^s = \frac{\%\Delta R}{\%\Delta P_R} \qquad [17\text{–}1]$$

where *R* is the quantity of rice output, and P_R is the price of rice. Estimates of the magnitude of this elasticity, calculated for agriculture more generally, vary from near zero (which would imply that producers do not increase supply in response to higher prices) to about 1.5 (which would imply that a 10 percent increase in price brings forth a 15 percent increase in supply). One reason supply elasticity might be quite low could be the existence of **nonprice constraints** to production. Farmers may lack access to the inputs necessary to increase output, either because markets are inadequate or inaccessible, input costs are too high, or credit is unavailable. Production may also be constrained by poor institutions, such as lack of secure land tenure.

One potential point of confusion in the estimation of supply elasticity lies in the question of whether this elasticity refers to a single commodity or to agriculture in general. Although the price elasticity of supply for a single commodity in a given country may be relatively high, the elasticity of supply for agriculture in general in that same country may be relatively low. Why? Recall the production possibility frontier of Figure 17–8, in which increased production of corn comes at the direct expense of reduced production of beans. Generalizing from this example, opportunities for farmers to substitute across commodities in response to changes in relative prices make such a result likely. It is also important to consider the time frame of the analysis. In the short run, many of the conditions of production are fixed; yet, in the long run conditions can change, perhaps because nonprice constraints are relaxed. As a rule of thumb, the long-run supply elasticity for given crops is approximately twice the magnitude of the short-run elasticity. From a theoretical perspective, we can think of short-run supply response as reflecting a movement along a given (static) supply curve as quantity produced responds to a change in price. Yet, it is important to note that in the

long run, technical change can also contribute to farmers' supply response. In this context, we can think of technical change as reducing the marginal cost of production. This shifts the marginal cost curve downward (or, equivalently, shifts the supply curve to the right), increasing the supply of food at every price. In addition, in the long run other nonprice constraints to production response may be overcome. Improved infrastructure might provide farmers greater market access for both inputs and outputs, credit supply might increase, and farmers may learn how to improve their farming practices.

Estimates of the aggregate supply elasticity for African agriculture have tended to be low, on the order of 0.2 in the short run and 0.4 in the long run. Yet some studies of specific crops have claimed to find significantly greater elasticities. One study of Trans Nzoia, a corn-producing district in Kenya, estimated the short-run elasticity of supply of corn to be 0.53 and the long-run elasticity to be 0.76. Thus a 10 percent increase in the price of corn would result in a 5.3 percent short-run increase in corn production and a 7.6 percent increase in output in the long run.[20] Another study of food production in Tanzania found that a 10 percent increase in the price of food crops (relative to export crops) would result in a 3.9 percent short-run increase in per capita food production and a 9.2 percent long-run increase.[21]

This discussion of how agricultural producers respond to price incentives provides a foundation for understanding how policy makers might intervene to change farmers' production decisions. Whether the goal is to reduce a country's dependence on imported cereals, to increase the local supply of vegetables, to increase its supply of coffee for export, or to employ more rural labor, price policies provide a powerful, if blunt, tool for policy makers. Box 17–2 discusses fertilizer subsidies in Malawi; yet this is only one of many possible examples of how governments have tried to influence farmers' behavior. It bears repeating, however, that producer incentives are only one component of the broader realm of food policy. Food policy analysis would be relatively simple if the only goal were to increase production and maximize producers' incomes. The challenge in practice (the food policy dilemma) is how to make farming profitable while at the same time protecting poor consumers from the threat of high food prices. The prices received by farmers are lower than the prices paid by consumers. The margin between the **farm-gate price** and the **retail** price is called the **marketing margin**. This is the cost required to store, transport, and process basic farm output so that it can be purchased and used by consumers who may be far from the point of origin or who may be purchasing the food months after harvest. To the extent that these marketing margins are increased by poor infrastructure or limited access to markets, the food policy dilemma is aggravated. It is important to consider market access in conjunction with the analysis of producer incentives.

[20]Lawrence O. Moses, Kees Burger, and Arie Kuvyenhoven, "Aggregate supply Response to Price Incentives: The Case of Smallholder Maize Production in Kenya," *African Crop Science Conference Proceedings*, 8 (2007), 1271–75.

[21]Andrew McKay, Oliver Morrissey, and Charlotte Vaillant, "Aggregate Export and Food Crop Supply Response in Tanzania," DFID-TERP: CREDIT Discussion Paper 4 (No. 98/4), Center for Research in Economic Development and International Trade, University of Nottingham, Nottingham, England, 1998.

MARKET ACCESS

Poor infrastructure in much of the developing world limits farmers' access to markets and their ability both to purchase inputs and to sell the crops they produce. Farmers may see little benefit in accepting the risk of adopting modern varieties and incurring debt to buy inputs in return for the promise of increased production when markets for their produce are remote. Selling output in distant markets, particularly when transport is scarce, can be costly for farmers. Many may be limited to moving small quantities, perhaps over long distances. Production incentives are reduced when farmers face high marketing costs. The World Bank reports that 16 percent of the rural population of developing countries (about 440 million people) lives in remote areas that are at least five hours from a market town of 5,000 or more.[22] This problem varies widely across regions: in sub-Saharan Africa, the Middle East, and North Africa, nearly one-third of the population lives in these remote areas, while that is the case for 17 percent of the population of East Asia and only 5 percent of the population of South Asia. In many cases even high-potential agricultural areas lack the infrastructure necessary to connect them to markets. In Ethiopia, the World Bank reports that 68 percent of the rural population lives in areas with ample rainfall, but the average farm household lives 10 kilometers from the nearest road and 18 kilometers from the nearest public transport. These variations in quality of rural infrastructure relate to differences in both income levels and population density (as investments in public goods such as roads are more costly in low population density areas). In addition to poor infrastructure, poor institutions can also limit the agricultural supply response and more generally impede agricultural development. Box 17–3 describes the growing role of cell phones in agricultural marketing in developing countries.

BOX 17–3 CELL PHONES AND AGRICULTURAL DEVELOPMENT

Cell phones are transforming rural markets in developing countries, particularly in sub-Saharan Africa where traditional land lines were largely unknown in rural areas. By 2009, there were 10 times as many cell phones as land lines in sub-Saharan Africa, and the number continues to grow rapidly. In 1999, only 10 percent of Africans were potentially covered by mobile phone technology; yet, by 2008, this number had increased to 60 percent (though coverage remains much greater in some countries than others). This rapid growth prompted Rwandan president Paul Kagame to declare that "what was once an

[22]World Bank, *World Development Report, 2008*, p. 54.

object of luxury and privilege, the mobile phone, has become a basic necessity in Africa."[a] Efficient markets require easy access to information by market participants. Knowledge of market prices, wages, weather patterns, and other relevant events are critical pieces of information for farmers and traders. The rapid penetration of cell phones has greatly reduced the cost of gaining such information.

Traditionally, traders in rural African markets could learn prices in distant markets only by actually traveling to those markets. The cost and time required imposed a severe limitation on traders' knowledge of the markets in which they operated. For the traders with cell phones, information that previously required hours of travel to obtain can be learned in a short phone call or text message. Economist Jenny Aker finds that the use of cell phones by grain traders in rural Niger reduces their search costs by 50 percent.[b] She interviewed grain traders in Niger, one of whom said that with a cell phone, "in record time, I have all sorts of information from markets near and far." A second trader reported to Aker that, "[Now] I know the price for two dollars, rather than traveling [to the market], which costs $20." As a result, prices across distant markets tended to converge and consumer grain prices fell by 3.5 percent. Both traders and consumers were better off. Similarly, economist Robert Jensen finds that the availability of mobile phones in India creates substantial benefits for fish traders and consumers. There, too, prices across markets tended to converge (a sign of efficient price formation) when traders had access to mobile phones; fishermen's profits increased by 8 percent, and consumers of fish benefited from a 4 percent decline in retail prices.[c] The benefits of such rapid information sharing are particularly clear in fish markets, where the product is highly perishable. Economists Megumi Muto and Takashi Yamano note that access to mobile phones increased the likelihood that banana producers in Uganda would participate in markets by 10 percent.[d] Bananas, like fish, are highly perishable. For the marketing of such commodities, the benefits of speedy information flows are particularly clear. The *digital provide*, as Jensen called it, will surely continue its rapid expansion and transformation of rural markets around the world.

[a]Address at the Connect Africa Summit, Kigali, Rwanda, October 29, 2007.

[b]J. Aker, "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger," *American Economic Journal: Applied Economics* 2, no. 3 (2010), 46–59.

[c]R. Jensen, "The Digital Provide: Information (Technology), Market Performance and Welfare in the South Indian Fisheries Sector," *Quarterly Journal of Economics* 122, no. 3 (2007), 879–924.

[d]M. Muto and T. Yamano, "The Impact of Mobile Phone Coverage Expansion on Market Participation: Panel Data Evidence from Uganda," *World Development* 37, no. 12 (2009), 1887–96.

Source: This box draws on Jenny C. Aker and Isaac M. Mbiti, "Mobile Phones and Economic Development in Africa," *Journal of Economic Perspectives*, 24, no. 3 (summer 2010): 207–32.

INSTITUTIONS FOR AGRICULTURAL DEVELOPMENT

Economic historian and Nobel laureate Douglass North famously defined institutions as "the humanly devised constraints that structure human interaction. They are made up of formal constraints (rules, laws, constitutions), informal constraints (norms of behavior, conventions, and self-imposed codes of conduct), and their enforcement characteristics. Together they define the incentive structure of society and specifically economies."[23] Institutions thus define the rules of the game by which individuals in society interact with one another. In the context of agricultural development, institutions governing the ownership and transfer of land are of particular importance.

In most countries, land ownership is a key determinant of power and social status, making land allocation a highly political concern. Land (like income, wealth, and political power) is unevenly distributed. For selected countries, Table 17–2 provides data on the size of an average land holding as well as an indicator of the equity with which land is distributed within the country. The equity of land distribution is summarized in Table 17–2 by a Gini coefficient for land ownership. Recall from Chapter 6 that a Gini coefficient summarizes relative distributional equity with a number ranging between 0 and 1. The closer the Gini coefficient is to 1, the more unequal is the distribution. Several points stand out in these data. There is great variation in the size of average land holdings and the equity of land ownership across regions and countries, reflecting large differences in population density, culture, and history. Average farm sizes in Africa and Asia tend to be quite small relative to some of the larger countries in South America. In Ethiopia, the average farm size is 0.8 hectare, as compared with 469 hectares in Argentina and 740 hectares in the United States.

The Gini coefficients for land ownership in Africa tend to be on the low side, such as 0.37 in the Democratic Republic of the Congo and 0.38 in Namibia. These low Gini coefficients reflect highly equitable distributions of land ownership. Conversely, the distribution of land in South American countries is highly skewed. Gini coefficients on the order of 0.85 for Brazil and 0.93 in Paraguay indicate that a large portion of land is owned by a small number of large landowners, while the large majority of rural dwellers in those countries must divide up whatever land is left over. The highly unequal distribution of land in Latin America has such deep historical roots that special terms have evolved to describe it: **latifundia** (the term for the large estates that dominate the distribution) and **minifundia** (the small family farms on which peasant farmers struggle to feed their families).

Institutions governing land rights play a large role in shaping incentives for agricultural producers, both in terms of their level of effort and their willingness to invest in maintaining the quality of their land. Insecure land rights can undermine

[23]Douglass North, "Economic Performance through Time," in Torsten Persson, ed., *Nobel Lectures, Economics 1991–1995* (Singapore: World Scientific, 1997).

TABLE 17-2 Farm Size and Equity of Land Holdings in Selected Countries

REGION/COUNTRY	CENSUS YEAR	AVERAGE FARM SIZE (HECTARES)	GINI COEFFICIENT FOR LAND HOLDINGS
Africa			
Burkina Faso	1993	3.92	0.42
Congo	1990	0.53	0.37
Egypt	1990	0.95	0.65
Ethiopia	1989–92	0.80	0.47
Guinea	1995	2.03	0.48
Malawi	1993	0.75	0.52
Namibia	1995	2.64	0.38
Uganda	1991	4.70	0.59
North and Central America			
Bahamas	1994	11.5	0.87
Barbados	1989	190.0	0.94
Canada	1991	349.1	0.64
Honduras	1993	11.17	0.66
Panama	1990	110.0	0.87
Puerto Rico	1987	100.0	0.77
United States	1987	740.0	0.74
South America			
Argentina	1988	469.0	0.83
Brazil	1985	64.64	0.85
Colombia	1988	120.0	0.79
Paraguay	1991	77.53	0.93
Peru	1994	20.15	0.86
Asia			
India	1991	1.55	0.58
Indonesia	1993	0.87	0.46
Japan	1995	1.20	0.59
South Korea	1990	1.05	0.34
Nepal	1992	0.95	0.45
Pakistan	1990	3.78	0.57
Philippines	1991	2.16	0.55
Thailand	1993	3.36	0.47
Turkey	1991	5.76	0.61
Vietnam	1994	0.52	0.53

Source: FAO, *World Census of Agriculture* (Rome: Food and Agriculture Organization of the United Nations, 2000).

producers' incentives and limit productivity. Secure land rights can enhance agricultural productivity by creating access to credit markets through which farmers can purchase improved inputs. The rules governing land sales can influence agricultural productivity by facilitating the transfer of ownership rights to the most efficient producers.

Land rights take a variety of forms in developing countries, ranging from purely individual rights to purely collective forms of land ownership, with numerous forms

in between these polar extremes. Thus at one end of the spectrum, owners have full control over who uses their land and what they produce; at the other extreme, all members of a community are entitled to use the land and benefit from the output. The most common intermediate forms of land rights are **fixed-rent tenancy** and **share cropping**. Share cropping is the dominant contractual arrangement in South Asia, while other regions of the developing world more typically rely on fixed-rent tenancy. Under tenancy, a family might rent land from a large landlord and pay a fixed cash rent. Under share cropping, the farm family rents a plot of land from the landlord, with whom the family shares a fixed portion of their output. A key difference between these systems lies in how the risks inherent in farming are divided between the landlord and the tenant. Under a fixed rent scheme, the tenant farmers bear all of the risk; their rent due on the land does not depend on the level of their output. In contrast, under share tenancy, the landlord and the tenants share the risk more evenly because the in-kind payments to the landlord depend on the tenant's level of output. Incentives for producers may also differ across types of tenancy. On large commercial farms worked by hired wage labor, owners face a complicated problem of monitoring the level of effort of their workers. In contrast, on purely communal farms, on which everyone shares the communal output, workers' incentives might be limited because an individual's rewards may not reflect that individual's efforts to maximize collective output. In terms of producer incentives, a key distinction across types of tenancy arrangements is whether the producers are **residual claimants** to the output of the farm—that is, whether their individual benefits increase with their level of effort.

Another critical distinction between different land institutions is the security and stability of producers' rights to farm particular parcels of land. A rental contract written for one or two years is certainly less secure than clear title to a given parcel. Economists Klaus Deininger and Gershon Feder summarize the costs and benefits of secure individual property rights.[24] The main benefits include (1) improved incentives to conserve and invest in the land itself; (2) the ability to transfer land ownership, possibly to those able to make the best use of the land (for instance by taking advantage of economies of scale); and (3) the ability to use land as collateral, providing farmers access to credit markets (and the potential to adopt modern varieties and inputs). The main costs are the administrative costs of actually defining land boundaries and enforcing land rights and the risk that the poorest farmers, who often depend on access to **common property resources** (such as open pastures for grazing their livestock), could be deprived of their sources of livelihood.

There is substantial evidence to support the idea that secure land rights increase both agricultural productivity and investments in land conservation (which enhance future productivity). In China, the transition from collective to private cultivation has been associated with significant increases in productivity, as individual farmers

[24]Klaus Deininger and Gershon Feder, "*Land institutions and land markets*," in B. L. Gardner and G. C. Rausser, eds., *Handbook of Agricultural Economics*, 1st ed. (Amsterdam: Elsevier, 2001) vol. 1, pp. 288–331.

stand to retain more of the benefits of increased effort. Longer-term investments in land conservation are also evident. In Ghana, studies have found farmers with greater tenure security are more likely to plant trees and to invest in drainage and irrigation. In Niger, farmers who owned their land applied larger quantities of manure to maintain their soil fertility than did tenant farmers. In addition, there is evidence that secure tenure rights not only induce investment in land conservation but also facilitate access to the formal credit markets necessary to finance those investments. In Thailand, land ownership titles not only increased the value of land and boosted productivity but also increased the supply of credit. Another study in Paraguay found that the benefits of land titling were roughly equivalent to 10 percent of farm income. Yet, that study also found that those benefits were strongly concentrated among the larger land owners, whereas producers with less than 20 hectares received no benefits of increased credit supply.[25]

Improving systems of **land titling** is one approach to securing tenure rights. Land titling involves providing legal documentation of land ownership. Plans to improve tenure security through titling must also consider the costs of implementation. These costs are largely administrative and include measurement and demarcation of areas, adjudication of conflicting land claims, and the cost of documentation. Land titling can also facilitate the operation of land markets, which are often underdeveloped in poor countries. Land markets, however, may not facilitate access to land by the poor if the poor lack access to credit markets. When land titling alone is insufficient to achieve social goals in regard to land rights, various forms of land reform have been attempted.

LAND REFORM

Efforts to reform land rights can take a variety of forms, including the following:

- *Reform of rent contracts* works to increase the security of tenure of the tenant farmer. Laws requiring long-term contracts that restrict the landlord's right to remove the tenant strengthen the tenant's property rights at the expense of those of the landlord without necessarily transferring income from owner to tenant.
- *Rent reduction* typically involves placing a ceiling on the percentage share of the crop that a landlord can demand as rent. If the percentage share is substantially below what prevailed in the past, the impact both on tenant welfare and the tenant family's surplus available for investment can be substantial.
- *Land to the tiller* (the former tenant) *with compensation* to landlord for loss of land is a measure that can take many different forms. Government

[25]These studies are summarized in Deininger and Feder, "Land Institutions and Land Markets."

might pass a law stating a ceiling on the number of acres an individual can own and so force individuals to sell all land over that limit. Or the reform law can state that only those who actually till land can own it, and all other land must be sold. A key issue in this kind of reform is whether the former landlord receives full or only partial compensation for the land that must be sold.

- *Land to the tiller without compensation* is the most radical transformation of land rights. All land not cultivated by its owner is confiscated, and the former landlord receives nothing in return. This type of reform is most commonly found in the aftermath of a major political upheaval, such as a revolution or losing a war.

The implementation of these types of land reforms becomes more complex as we move down the list, and the potential for harmful unintended consequences exists in each case. Placing ceilings on land rents is one of the least intrusive of these potential reforms. Yet rent ceilings may create incentives for landowners to evict their tenants. This was a common result in Latin America as well as in India, where similar reforms led to the loss of 30 percent of the total area operated by the poor. Conversely, Deininger and Feder report that tenancy reform in West Bengal was associated with productivity gains of 40 percent.[26] Efforts to limit the size of land holdings have generally failed. The major problem in these cases is that ceilings can be evaded, for instance, by subdividing large landholdings among family members.

Land to the tiller programs are examples of redistributive land reform. These programs involve transferring ownership of land and thus face increasing political obstacles. Experience has been mixed, with greater success under certain preexisting circumstance. In settings in which tenants already cultivate given parcels of land that are pieces of larger landlord estates, transferring ownership of the land is administratively simple and can increase productivity by improving producers' incentives. Successful examples of this type of reform have occurred in Bolivia, Ethiopia, India, Iran, Japan, Korea, and Taiwan. Yet, experience with land reform in **hacienda** systems, by which tenants typically work most of the time on a landlord's plot of land but have their own subsistence plots, has been less positive. A common experience in Latin America when faced with the prospect of redistributive land reform was for landlords to evict their tenants or to convert them into wage laborers. In addition, land compensation schemes can be extremely costly and the task of estimating a fair compensation price for confiscated land is complicated by the poor quality of land markets in many countries. When land redistribution creates new landowners who were not previously cultivating that land, the need for complementary types of support (such as infrastructure and farmer training programs) might add substantially to the costs.[27]

[26]Deininger and Feder, "Land Institutions and Land Markets."

[27]Deininger and Feder, review these issues in greater detail; see "Land Institutions and Land Markets."

If land redistribution with compensation is complicated, redistribution *without* compensation is virtually impossible in the absence of a social revolution, foreign invasion, or both. The successful examples of redistributive land reform in Japan, the Republic of Korea, and Taiwan were possible only as the result of defeat in war or foreign occupation. In China and the Soviet Union, foreign invasions preceded social revolutions that destroyed the old agrarian orders and paved the way for large-scale land reform. In other cases, revolution alone created the conditions necessary for redistributive land reform. Examples include Cuba, Egypt, Ethiopia, Nicaragua, and Vietnam.

Redistributive land reform continues to be a key and controversial issue in many countries. This is particularly true in countries with highly unequal distributions of land ownership, often the result of historical legacy or ethnic discrimination. The southern African states of Zimbabwe and South Africa, for example, were settler colonies in which white minorities owned much of the prime agricultural land. Redistributive land reform has been a dominant social issue in these countries. Forced expropriation of white-owned farms in Zimbabwe began in 1999 with the support of President Robert Mugabe. Land reform in Zimbabwe has been widely criticized, not only for its violence but also because political favoritism in distributing newly seized land has contributed the failure of Zimbabwe's program to benefit most of the rural poor. In South Africa, the collapse of the Apartheid regime and the transition to democracy in 1994 brought with it high expectations among the indigenous majority for restitution in the form of land. These expectations were fueled by events in Zimbabwe. Yet, in contrast to that example, the government of South Africa has taken a more cautious approach to land reform. A desire to avoid the economic chaos that befell Zimbabwe helps explain this caution. After its election in 1994, the government of the African National Congress announced plans to redistribute 30 percent of commercial farmland to blacks by 2014; yet as of 2010, progress was limited to only 6 percent, and tensions continued to surround this issue. Redistributive land reform is also a key issue in many postconflict settings. Mozambique and Liberia, after long periods of civil war, face serious challenges of resettling large displaced populations on land since occupied by others. Land reform has also been a central issue in South America, where indigenous Indian minorities have long suffered discrimination. In Bolivia, for example, the government planned to redistribute 400,000 acres of land in 2011, a substantial increase over the 2010 level.

THE WORLD FOOD CRISIS OF 2005–08

During the years 2005–08, the world suffered its most severe food crisis since the early 1970s. The key indicator of this crisis was the rapid increase in the prices of a range of basic foods. As Figure 17–12 shows, the prices of rice, maize, and wheat increased

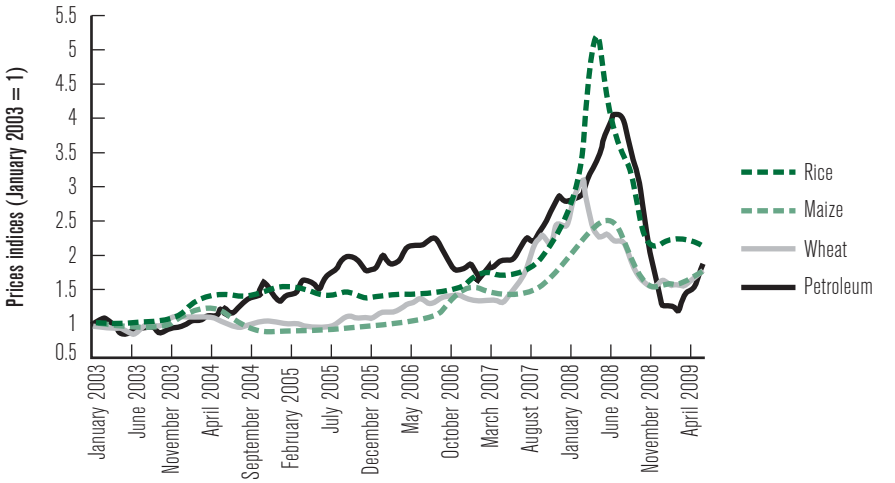


FIGURE 17–12 Trends in the Nominal Prices of Cereals and Oil, January 2003 to May 2009

Source: Headey, D., S. Malaiyandi, and S. Fan. 2009. "Navigating the Perfect Storm: Reflections on the Food, Energy, and Financial Crises." IFPRI Discussion Paper 00889, Figure 1. Washington, DC: International Food Policy Research Institute.

dramatically between early 2007 and mid-2008, following more gradual increases beginning in 2004. The price of maize showed the least increase, yet still reached a level by June 2008 that was two and a half times greater than it had been just two years earlier. The run up in wheat prices was even more severe, but both were outpaced by the skyrocketing price of rice during 2007–08. From October 2007 to April 2008, the world price of rice tripled (increasing from $335 to over $1,000 per ton).

CAUSES OF THE CRISIS

Explanations for the crisis include a wide range of possibilities and are the subject of continuing debate. The list of potential culprits includes both supply shocks and demand shocks: growth in demand from China and India, speculation on financial markets, hoarding and export restrictions, weather shocks, decreased productivity, low interest rates, a depreciation of the U.S. dollar, rising oil prices, declining food stockpiles, and demand for biofuels.[28] There is no definitive consensus regarding the relative contributions of these potential causes. Notably, some of these potential causes are common to all commodities (such as the effect of oil price increases and the depreciation of the U.S. dollar), whereas other causes are commodity-specific (such as supply shocks and the effect of biofuels demand).

[28]See D. Headey and S. Fan, "Anatomy of a Crisis: The Causes and Consequences of Surging Food Prices," *Agricultural Economics* 39 (suppl., 2008), 375–91.

Among these causes of the world food crisis, the most hotly debated has been the effect of demand for biofuels, which has increased rapidly since 2003 (in part due to subsidies by the U.S. government and related demand-stimulating policies). This development was particularly important for corn prices. By 2007, biofuels consumed 25 percent of the U.S. corn crop. Economists Derek Headey and Shenggen Fan conclude that among the widely varying estimates, the best estimate is that 60 to 70 percent of the increase in corn prices was explained by the diversion of corn for biofuels production. Other studies attribute a much smaller role to biofuels in explaining the increase in corn prices. That debate illustrates two important considerations that complicate analysis of the food crisis. One is that although the prices of a wide range of commodities increased, the causes of these increases differed across specific commodities. (Rice, for example, has not been widely used in biofuels production.) A second point that arises in considering the impact of biofuels demand is the need to consider interactions between crops. In the case of biofuels, the increased demand for corn is thought to have explained up to 40 percent of the increase in soybean prices. What's the connection? The 23 percent increase in acreage in the United States devoted to corn came at the expense of a 16 percent reduction in soybean acreage in 2007. In addition, to the extent that consumers substitute between staple gains, increased corn prices have contributed to increased demand for wheat and rice, adding to price increases for those commodities.

The explanation for the tripling of world rice prices relates more to the combustible mixture of distortionary trade interventions and panic buying. Historically, world rice markets are thin compared with those for other grains. That is, only 5 to 7 percent of global rice production is traded internationally (compared with 20 percent of wheat). This makes world rice prices relatively volatile. In addition, global rice exports are dominated by a small number of countries: Thailand, India, Vietnam, and Pakistan alone account for about 70 percent of global rice exports. So conditions in the world rice market were ripe for an implosion in November 2007, when India reacted to rising rice prices by banning exports. India's ban on rice exports set off a chain reaction in world rice markets. Importing nations, the Philippines in particular, accelerated their import orders, fearing further price increases. These orders, themselves, helped drive up prices, in response to which Vietnam, China, Cambodia, and Egypt also intervened to restrict rice exports. Figure 17–13 places these and related events on a chart of world rice prices.[29]

Commodity-specific factors also affected world wheat prices. In this case, the main culprit was a weather-induced supply shock. Australia, one of the world's leading exporters of wheat, suffered its worst drought in a thousand years in 2006, cut-

[29]D. Headey, S. Malaiyandi, and S. Fan, "Navigating the Perfect Storm: Reflections on the Food, Energy, and Financial Crises," Discussion Paper No. 00889, International food Policy Research Institute, Washington, DC, August 2009. For a comprehensive review of the rice crisis, see D. Dawe, ed., *The Rice Crisis: Markets, Policies and Food Security* (London: Earthscan, 2010).

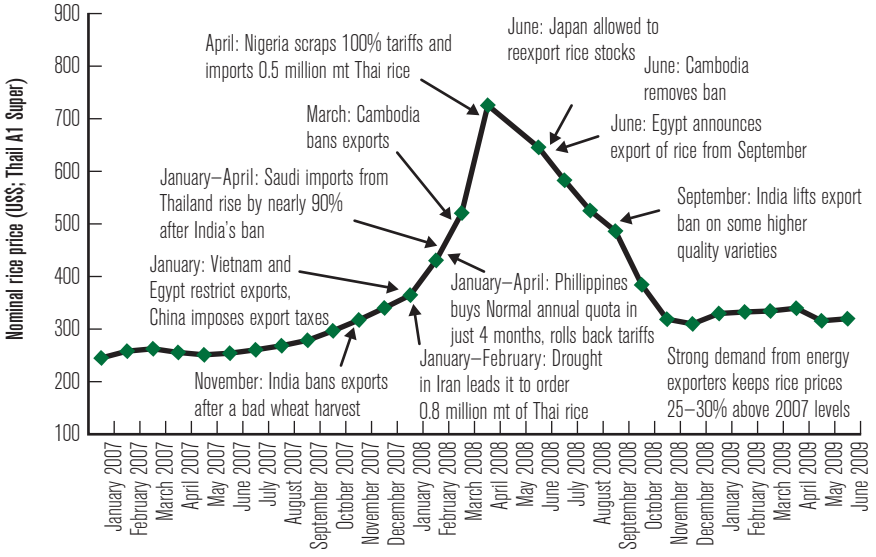


FIGURE 17–13 Events in World Rice Market

mt, metric tons.

Source: Headey, D., S. Malaiyandi, and S. Fan. 2009. "Navigating the Perfect Storm: Reflections on the Food, Energy, and Financial Crises." IFPRI Discussion Paper 00889, Figure 2. Washington, DC: International Food Policy Research Institute.

ting in half that country's wheat exports. That year's harvest in other major producing countries, including the United States, Russia, and Ukraine, were also smaller than usual.

A more general factor was the price of oil. The price of oil surged just in advance of the prices of wheat, maize, and rice. Oil prices are relevant for several reasons. For instance, chemical fertilizers are largely petroleum-based products, and most farm machinery runs on gasoline or diesel fuel. In addition, higher oil prices also increased the cost of transporting agricultural output, both within and between countries. Here too we see an interaction effect with other potential causes of the crisis: biofuels production becomes more profitable when oil prices are high. In addition, the ability of markets to respond to price increases in a range of commodities by releasing stockpiles of those grains into the market was severely constrained by the historically low levels of grain reserves at that time. Ironically, one cause of the low levels of grain reserves was the secular decline in prices before the crisis.

CONSEQUENCES OF THE CRISIS

The consequences of the world food crisis of 2005–08 are even more difficult to sort out. While our ultimate concern may be the welfare of poor households, the causal chain running from global cereals price shocks to household welfare is long and

complex. Headey and Fan distinguish between macroeconomic and microeconomic links in that chain. The macroeconomic linkages entail such considerations as the size of a country's food import bill, changes in exchange rates between the local currency of an importing country and the U.S. dollar (the currency in which international grain prices are quoted), and countries' stocks of foreign exchange reserves (the funds used for imports). These considerations generally indicate the extent to which governments of food importing countries may be able to buffer their domestic consumers (poor consumers especially) from rising prices of food on international markets. For instance, the price in local currency (say, in pesos) of an imported commodity such as wheat is equal to the world price in dollars times the exchange rate (in pesos per dollar) for the local currency. So if the world price of wheat in U.S. dollars increases at the same time that the peso appreciates against the dollar (that is, if the peso–dollar exchange rate changes such that a peso purchases more dollars, resulting in a "lower" exchange rate indicating fewer pesos per dollar), then some of the increase in world wheat prices will be absorbed by this exchange rate movement and not passed along to domestic (local currency-denominated) food prices. Over the period 2003–07 countries' experiences in this regard varied widely.

Yet the welfare impacts also depend on a set of microeconomic factors that come into play once world market events have been translated into domestic food prices. At that point, changes in the relative prices of foods may motivate consumers to substitute among commodities (in favor of those that have become relatively cheaper). Opportunities for substitution away from suddenly more expensive commodities can help dampen the potential adverse effects on consumers. But these opportunities also depend on consumer preferences and the patterns of food consumption for households at different levels of income. In particular, the effect of rising food prices on household welfare depends on whether the household is a net buyer or net seller of food (as discussed earlier).

Estimates of the welfare impact of the world food crisis have also varied widely. Studies have tended to conclude that poverty (even rural poverty) tends to increase with rising food prices. This is not surprising, given that nearly all urban dwellers and a substantial proportion of rural dwellers (especially the poor) are net buyers of food (though net sellers, many of whom are still relatively poor, have benefited). While countries varied widely in the impact of the food crisis on poverty, effects also varied widely within individual countries. Economists Maros Ivanic and Will Martin concluded that on average (over the nine countries they studied) poverty rates increased by 4.5 percentage points in the short run.[30] Projecting this result to all low-income countries, they suggest that the food crisis could have increased the global poverty head count by 105 million. Their estimate is thus substantially higher than that of the Food and Agricultural Organization, which found that the food crisis increased the

[30]M. Ivanic and W. Martin, "Implications of Higher Global Food Prices for Poverty in Low-Income Countries," *Agricultural Economics* 39 (suppl. 2008), 405–16.

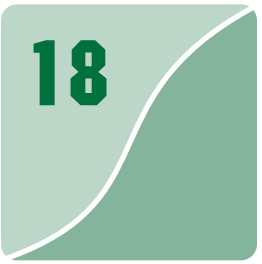
poverty head count by 40 million. These examples demonstrate both the variation in estimates of the magnitude of the impact and the broad consensus that the world food crisis hurt the poor.

SUMMARY

- A variety of characteristics distinguish agriculture from other sectors in a typical developing economy. These unique features of agriculture include its large share of GDP and employment, several technical characteristics of the agricultural production function, that much of the sector's output is directly consumed by its producers, and agriculture's role as a resource reservoir.
- Farms and farmers in developing countries differ widely from one another, both within and between countries. These differences include wide variations in farm size, technology, environmental conditions, crop mix, and degree of integration in markets. Nonetheless, it is possible to characterize traditional agriculture at a broad level of generalization as consisting of small-scale family farms producing food primarily for their own consumption, using traditional crop varieties and techniques and few purchased inputs, with low levels of productivity.
- Agricultural systems can take a variety of forms, with the major types being systems of shifting cultivation, pastoral nomadism, and settled agriculture. Most of the progress made in promoting agricultural development in recent decades has concentrated on systems of settled agriculture.
- Growth in agricultural productivity (commonly defined as the level of output per unit of land) is a central component of agricultural development. The growth rate of yields for cereal crops has been consistently high in East Asia, but quite low in sub-Saharan Africa (with the performance of other developing regions ranging between these extremes).
- Productivity growth in agriculture may be constrained by a variety of factors, including a lack of improved technology, unfavorable economic incentives, and poor-quality institutions.
- The green revolution produced new agricultural technologies and contributed significantly to agricultural productivity growth in Asia and Latin America. The benefits to farmers in sub-Saharan Africa have been more limited.
- Government policies that influence food prices play a critical role in shaping the incentives for farmers as they decide what to produce, how to produce it, and how much to produce. Meeting the goal of increasing agricultural output and productivity requires positive incentives for

farmers, typically in the form of higher food prices. Yet government efforts toward that end may be severely constrained by the negative impact of high food prices on poor consumers, many of whom may be at severe risk of undernutrition and hunger.

- Institutions governing land rights are of particular importance in shaping incentives and opportunities for farmers. Secure tenure arrangements encourage both greater levels of effort in current farm production and greater levels of investment in conserving soil resources for future production. Efforts to improve tenure security through land reform are often appealing in theory, but difficult to implement.



Trade and Development

In April 2009, Trinidad and Tobago hosted the fifth Summit of the Americas. The summit was viewed as a huge success, devoid of confrontation and heralded as a new beginning for relations with the region. It stood in sharp contrast to the summit held four years earlier in Mar del Plata, Argentina, when thirty-four presidents from the Western Hemisphere assembled to discuss the Free Trade Area of the Americas, an idea proposed in 1994 by Bill Clinton and supported by his successors. Mexican President Vicente Fox also backed the proposal, as did other Latin American leaders. But there were voices of dissent. Populist presidents from Argentina and Brazil had their doubts. Most critical was Hugo Chávez, the leftist leader of Venezuela, who rallied a crowd of an estimated 25,000 protesters in a stadium near the closed-door meetings proclaiming, "Each one of us brought a shovel, a gravedigger's shovel, because here in Mar del Plata is the tomb of the Free Trade Area of the Americas."

Rioting broke out after the rally, something that has happened repeatedly during official meetings about international trade. The meetings of the World Trade Organization (WTO) in Seattle in 1999 were disrupted by street protesters who felt free trade hurt workers in both developing and industrialized nations. In Cancun, Mexico, in 2003, there also were protests during the WTO meetings, both among official participants in the trade talks and by demonstrators outside. One of those demonstrators was Lee Kyang Hae, who headed South Korea's Federation of Farmers and Fishermen. Lee stabbed himself in the chest, and later died, in protest against the WTO and the more open-trade policies it advocated. Lee specifically opposed further trade liberalization of agriculture, which he believed would ruin the livelihood of many

Korean farmers. As these events demonstrate, globalization and free trade generate some of the world's most heated controversies over economic policy.

But, despite these controversies and concerns, trade among nations has been growing rapidly, and most developing countries are actively trying to expand their trade. Global trade has increased more than fivefold in the last three decades, with trade in developing countries increasing by even more. Developing nations import many more goods and services than they once did, ranging from food to pharmaceuticals to sophisticated machinery. They also export much more, not only agricultural products, oil, and other raw materials but also automobile parts, clothing, semiconductors, shoes, and toys. Revolutions in transportation (including containerized shipping), telecommunications, and production methods have facilitated this process. Supply chains have been broken up, with components made in one country and finished good assembled in another. This is true for complex products, such as automobiles, so there no longer is such a thing as "a national car." Even simple products undergo global manufacturing; a Barbie doll made in China may be assembled from plastic pellets from Taiwan, given nylon hair from Japan, and packed in a cardboard box from the United States.

Trade provides low- and middle-income nations with significant opportunities to improve welfare and accelerate growth and development. With more open trade, families and businesses have more and better choices in price, quality, and array of products than if they buy only from domestic firms. Producers have much larger markets to which they can sell. If successful, exporting firms can generate rapid job growth for large numbers of low-skilled workers, which can have a strong impact on poverty reduction. Trade also invites investment and creates the possibility for the transfer of new technologies from rich to poor countries, which can raise productivity and incomes.

But the news is not all good. Although more open trade creates many winners, it also creates losers. Firms and farmers producing at high prices for the domestic market can be forced out of business, with job losses and other disruptions. Exporters operating on world markets face the risk of rapidly falling prices or other vagaries of the world market that are out of their control. Poor weather in Nicaragua may hurt its coffee crop, sending up world prices and benefiting coffee farmers in Uganda, but a bumper crop in Vietnam can lead to a drop in prices that can leave the Ugandan farmer with much lower income. The 2008 financial crisis that precipitated a steep recession in the United States and European Union hurt exporters most everywhere in the world.

The evidence suggests that, on balance, more open trade is beneficial for developing countries and leads to more rapid growth and poverty reduction, particularly when exports are focused on labor-intensive products including agriculture and basic manufacturing. But there is plenty of controversy. Some economists argue that the contribution of trade, while positive, is smaller than many suggest and other factors are much more important in the development process. There is much debate

about the key policies necessary to stimulate trade and the balance between traditional market-oriented policies and government interventions to help spur exports. Some believe that more open trade creates sweatshops and leads to a race to the bottom in wages, labor standards, and environmental outcomes. And some argue that multilateral trade negotiations are biased against poor countries, allowing rich countries to continue to protect their own agricultural and textile producers, as well as intellectual property, while forcing developing countries to open their economies.

In this chapter, we begin to examine these debates by considering recent trends and patterns in trade flows across countries. We then develop the core concept of comparative advantage and its powerful implications for international trade. Building on this essential theory, we turn to those economies whose comparative advantage lies in primary products, including food and other agricultural products; oil and natural gas; timber, minerals and other raw materials. The advantages and disadvantages of trade in such goods are then evaluated.

TRADE TRENDS AND PATTERNS

World trade has expanded dramatically over the past 50 years. Figure 18–1 shows the level of exports, expressed in current U.S. dollars, by region, from 1974 to 2007.[1] The trend in imports by region would look similar even though some nations, like the United States, import more than they export and others, like China, export more than they import. The persistence of trade deficits and surpluses do little to alter the aggregate pattern of regional trade flows.

Exports (and imports) from high-income nations continue to dominate global exchange. This may surprise you because it often seems that everything is made in China. The label, "Made in China," appears in much of our clothing, shoes, toys, and hundreds of other consumer goods. But the bulk of world trade, whether in airplanes, automobiles, computer software, machinery, pharmaceuticals, wheat, and other goods still originates and is exchanged between high-income economies. Beginning in 2009, China replaced Germany as the world's largest exporter of goods. But even as the world's largest exporting nation, China accounted for only 10 percent of world exports. Canada, not China, remains the major trading partner of the United States.

Measured at market exchange rates, the high-income economies account for almost 80 percent of world gross domestic product (GDP). Even if these nations trade a smaller percentage of their GDP than China, as many do, they still dominate world

[1]Figure 18–1 refers to merchandise exports only—that is, exports of goods, including primary products (such as agricultural goods, oil, and minerals) and manufactures, measured in current U.S. dollars unadjusted for price inflation. Trade in services, such as tourism and international transport, insurance and financial services, and call-center activities and data processing, is not included. Merchandise trade in 2007 accounted for 80 percent of total world trade.

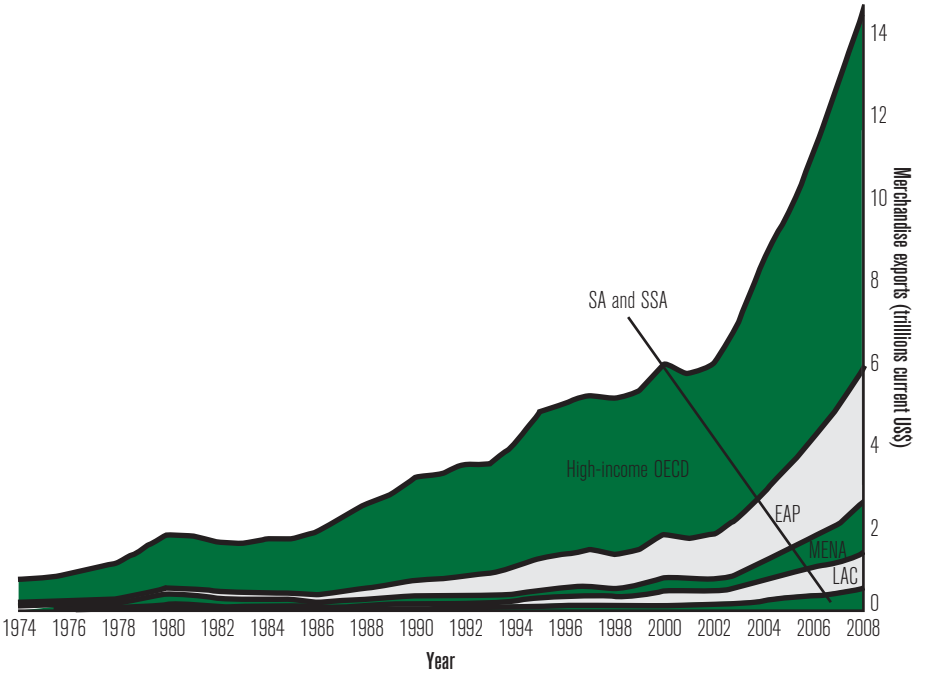


FIGURE 18–1 Growth in World Exports by Region, 1974–2007

EAP, East Asia and Pacific; LAC, Latin America and Caribbean; MENA, Middle East and North Africa; OECD, Organization for Economic Co-Operation and Development; SA, South Asia; SSA, sub-Saharan Africa. Exports from Europe and Central Asia are not included. Korea is included in EAP.
Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

trade. Remember, GDP is the sum of all expenditures on consumption, investment, and government spending plus net exports (exports minus imports). If the high-income economies produce 80 percent of world output, it is no surprise that they also produce almost 75 percent of world exports and imports.

But there has been significant growth in exports from low- and middle-income economies. This growth can be traced back to the late 1970s, after China's reversal in policy toward a more outward-looking trade strategy. Even earlier, Hong Kong, Korea, Singapore, and Taiwan had considerable success with trade serving as an engine of economic growth. Other nations, including China, began to follow a similar model. The growth in exports from East Asia is shown in Figure 18–1 and can be seen more dramatically in Figure 18–2, which shows the share of total trade in GDP, including imports and exports of both goods and services, by region and decade. East Asia, the region with the highest rate of growth in GDP per capita over the past 30 years is also the region with the steepest increase in the ratio of trade to GDP. This ratio averaged around 20 percent in the 1970s but exceeded 75 percent in the 2000s, considerably higher than any other region. There is little doubt that trade played a central role in East Asia's achievements in both economic growth and development.

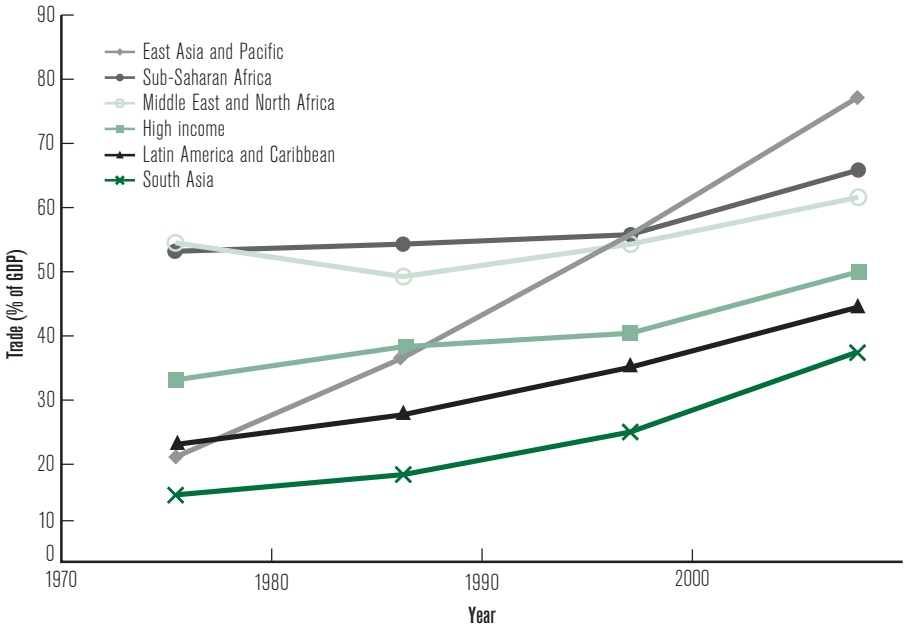


FIGURE 18–2 Importance of Trade to Economies, 1970–2007

The ratio of trade to gross domestic product (GDP) is the average value by decade by region.
Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

For developing countries as a whole, imports plus exports on goods and services combined are now equivalent to over 64 percent of their total output, indicating that large portions of these economies are influenced by global markets. But there are stark regional differences in these trends, and Figure 18–2 reveals some patterns that may be less familiar. Sub-Saharan Africa is often thought to be marginalized from the global economy. This is true in terms of the region's minuscule contribution to world exports and imports. Its merchandise exports constituted 3.6 percent of the world total in 1970, falling to just 1.8 percent by 2007. But viewed from the perspective of the region, trade has always accounted for a large share of Sub-Saharan Africa's GDP. In the 1970s the trade to GDP ratio, at 53 percent, was higher than in almost any other region at that time; in the 2000s, it averaged 66 percent, second only to East Asia. The problem for the region has not been the share of trade in output but the failure of trade and GDP to grow by very much overtime. South Asia, where India's economy dominates, is the region where trade continues to play the smallest role, although that is beginning to change with a greater emphasis on both exports and imports in India and elsewhere.

As economies grow and the share of trade increases, the products imported and exported tend to change as well. On the import side, as incomes grow, countries

typically import more sophisticated consumer goods and intermediate products as inputs to manufacturing. On the export side, as workers gain new skills and increase productivity, the composition of exports shifts away from primary products to manufactured products and even to some services, including call centers, data processing, and other examples of what is commonly referred to as outsourcing.

Figure 18–3 shows how the composition of merchandise exports has changed by region. The concern, once widely held, that developing nations might get stuck exporting primary products in return for imports of manufactured goods from developed countries, has limited validity today. Over the past three decades manufactures have replaced primary products as the main exports for three regions: East Asia, Latin America, and South Asia. These regions still export primary products, including natural gas from Indonesia, copper from Chile, and tea from Sri Lanka, but the export of manufactured goods now constitutes the majority.

Mauritius, an island economy off the east coast of Africa, has followed a similar pattern. In 1975, Mauritius's primary export was sugarcane, with manufactured goods accounting for only 12 percent of merchandise exports. By 2000, 81 percent of merchandise exports was from manufactures, as Mauritius became a major producer of clothing. Today, Mauritius is relying less on sugarcane and clothing and is developing exports in services, such as offshore banking, telecommunications, and

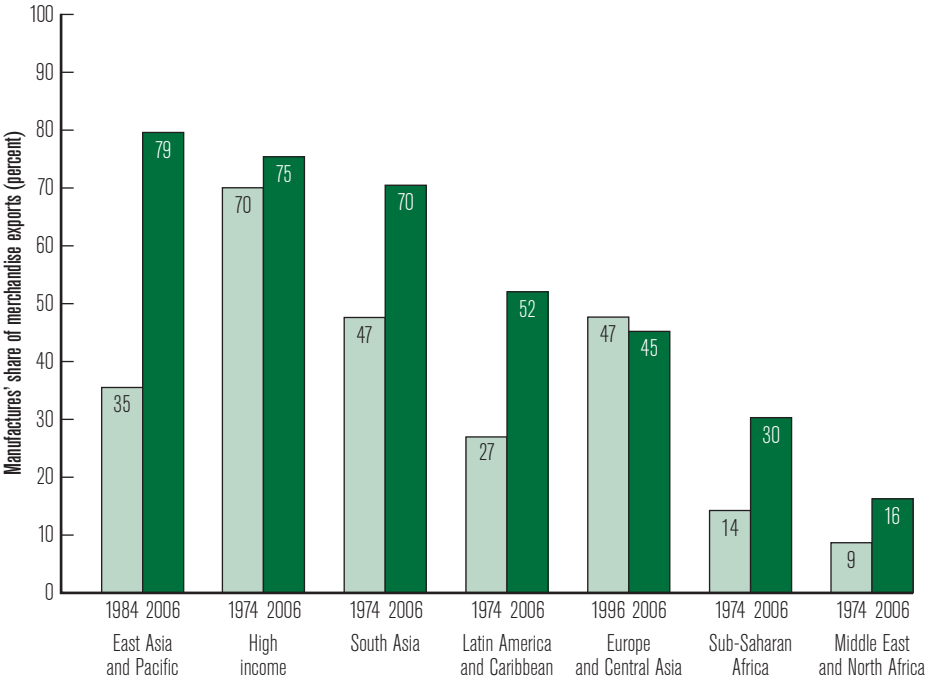


FIGURE 18–3 Importance of Manufactured Exports by Region

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

tourism. But on the African continent, primary products, including coffee, diamonds, groundnuts, oil, and minerals, continue to dominate merchandise trade. The impact of primary product exports on African and other economies is considered in detail later in this chapter.

For all developing regions, the majority of trade (around two thirds) continues to be with high-income economies. Since 1990, this share, not surprising, has increased significantly for the former Soviet Union and Soviet bloc. It has fallen in the Middle East as more oil exports go to other developing areas. The biggest change in Asia is that almost half of Asia's exports today are sent to other countries in Asia, including the larger high-income economies in the region (Japan, Korea, Singapore, and Taiwan). In 1980, within-Asia trade made up about one third of the region's total. Raw materials and intermediate products produced in one Asian country now are likely to be exported to a neighboring country for finishing and final export. China's dramatic rise is particularly significant. Its imports have grown almost as fast as its exports, and it is now one of the biggest markets in the world. A significant share of China's manufactured export products are assembled using components imported from other Asian economies.

WHO TRADES?

The extent to which a country trades with the rest of the world depends on many factors. Small countries, measured in this instance by population size, tend to trade more than larger ones, as they cannot efficiently produce the full range of consumer goods, intermediate products, and capital equipment demanded by consumers and businesses. Just as a small town cannot offer the same range of stores and services for shoppers as a large town, small economies cannot efficiently produce everything that households and firms would like to buy, so they tend to import more products to satisfy those demands. On the export side, as firms in small economies increase production, they are more likely to be constrained by the limited size of the market and unable to take advantage of economies of scale if they sell only locally. By exporting to global markets, they can sell larger amounts of more-specialized products. Export earnings then pay for the imports small nations demand. Producers in larger economies can sell a much larger share of their output locally without being constrained by market size. In Guyana, with a population of under 1 million, imports and exports together exceed 200 percent of GDP; in Mauritius they reach 133 percent. But in much larger Brazil and India, the ratios are 27 and 45 percent, respectively. Figure 18–4 confirms this relationship across all countries. It helps explain why sub-Saharan Africa, which has a large number of nations, half of them with populations below 10 million, has such a relatively high ratio of trade to GDP.

Trade patterns are also influenced by a country's geographical characteristics, including its access to shipping routes—especially whether it is landlocked—and its location relative to major markets. For the last several hundred years, the cheapest

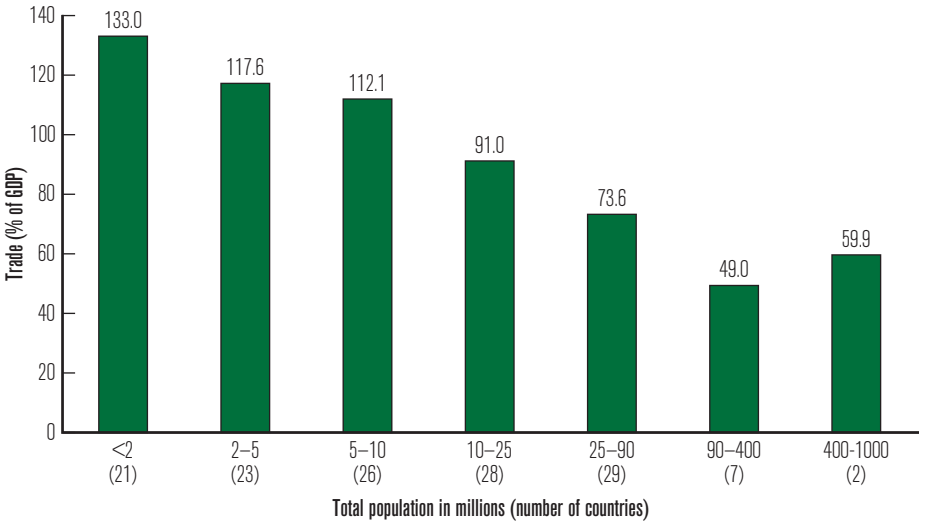


FIGURE 18–4 Importance of Trade to Less Populous Nations

Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

and most important form of transporting goods between countries was by sea, and to this day, countries with easier access to sea-based shipping tend to have larger exports and imports than do landlocked countries. Adam Smith recognized the advantages of access to the sea for trade and commerce more than 230 years ago in *The Wealth of Nations*:

> As by means of water-carriage a more extensive market is opened to every sort of industry than what land-carriage alone can afford it, so it is upon the sea-coast, and along the banks of navigable rivers, that industry of every kind naturally begins to sub-divide and improve itself, and it is frequently not till a long time after that those improvements extend themselves to the inland part of the country.[2]

As Smith predicted, most landlocked countries trade less. Figure 18–5 identifies 44 nations that do not have a coastline. As a group, their trade to GDP ratio is 25 percent as compared to 75 percent for all other economies.[3] Less trade among the landlocked countries is a result of much higher transport costs, as they must not only pay for sea-based shipping but overland costs to and from the nearest seaport. These costs can skyrocket when relationships between neighboring countries deteriorate. Periodic conflict between Ethiopia and Eritrea has forced firms in Ethiopia to seek

[2]Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (New York: Modern Library, 1976), bk. 1, chap. 3, para. 3.

[3]These trade to GDP ratios are *unweighted*—for example, in the case of the landlocked nations it is the simple average of the 44 countries. When presenting most averages, by region or income group, we usually provide *weighted* averages, implicitly giving more importance to the region or group's largest economies.

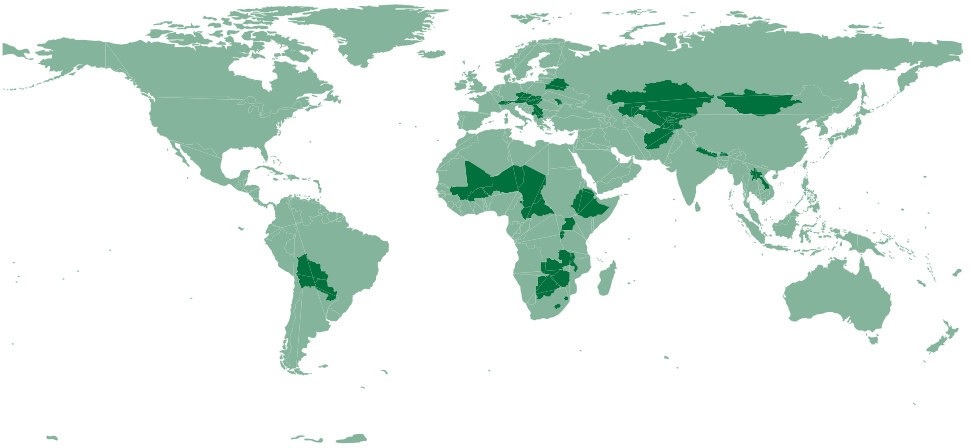


FIGURE 18–5 **Landlocked Nations of the World**

other more distant ports, for example, in Djibouti or Kenya. Landlocked Nepal almost totally depends on shipment through India. In Africa, 16 countries are landlocked, and their shipping costs have been two or three times higher than for their coastal neighbors.[4] As a result, everything that is imported is much more expensive. Similarly, exporters must pay more for shipping, raising the costs of export products and making firms less competitive on world markets.

As noted in Chapter 3, being landlocked not only reduces trade but is also associated with slower economic growth. An inability to trade more may account for this slower growth. But there are exceptions. Some landlocked high-income countries have done well, such as Austria and Switzerland, but they are located in the midst of major markets, are connected by high-quality road networks, and have had good relations with neighboring states, so shipping costs are not a major barrier. Landlocked countries that ship raw materials for which the margins between production costs and world prices are large are also in a better position to export, despite higher shipping costs. Landlocked Botswana has been very successful, as its diamonds bring in plenty of revenue to overcome higher shipping costs. Uganda's coffee, the nation's primary export, does not enjoy the same advantage. In addition to landlocked countries, countries isolated from major markets, such as Samoa and other small island countries of the Pacific Ocean, face similar issues with high shipping costs. As air shipment costs decline these disadvantages are reduced, but the vast majority of trade still takes place by sea.

In addition to size and geography, government strategies for trade and choices about trade policy determine trade outcomes. Broadly speaking, governments in developing countries have employed two different trade strategies, import substitution

[4]United Nations Conference on Trade and Development (UNCTAD), *Review of Maritime Transport 1995* (Geneva: United Nations Publications, 1995).

and outward-looking development. **Import substitution** is the production of goods and services that replace (or substitute for) imports. **Outward orientation** shifts the focus to producing for export for global markets. Before examining these two strategies in Chapter 19 it is necessary to understand some of the core ideas that shape economic reasoning about international trade.

COMPARATIVE ADVANTAGE

For more than 200 years, trade theorists have tried to explain why nations engage in international trade, what goods and services they trade, and how firms and consumers gain or lose. The workhorse models rely primarily on the **theory of comparative advantage**, which describes trade patterns under assumptions of *static conditions* that hold the factors of production in fixed supply and unable to cross borders. Comparative advantage has rich implications about the gains from trade. Among the most powerful results are the following:

- A country can increase its welfare by trading because the world market provides an opportunity to buy and sell goods at different prices than those that otherwise would be available to a nation that closed its economy to the rest of the world.
- The smaller the country, the greater is this potential gain from trade.
- A country often gains most by exporting commodities that it produces using its abundant factors of production most intensively, while importing goods whose production requires relatively more of scarcer factors of production.

The first implication is one of the most powerful ideas in all of economics. It is associated with the work of David Ricardo appearing in his *Principles of Political Economy and Taxation*, published in 1817. It is a testament to Ricardo's core insight that almost two centuries later comparative advantage still provides the foundation for thinking about trade between nations.

Any country can engage in and benefit from international trade, including the world's highest-cost and lowest-cost producers of any good. To see why, let us consider some examples. Mexico is the third largest trading partner of the United States. The United States is Mexico's largest trading partner. Many goods cross their borders. Among them, it is easy to understand why Mexico exports crude oil to the United States and why the United States exports aircraft to Mexico. The United States has depleted much of its oil reserves and imports oil from many countries, including Mexico. Mexico does not have an aircraft industry, nor is it likely to develop one anytime soon. That these nations would chose to trade, in essence, oil for airplanes is easy to understand. Some world trade is of this kind, by which nations trade goods that they essentially could not produce on their own. But most international trade does not have this characteristic, making the next example far more interesting.

TABLE 18–1 Production Costs and Comparative Advantage

LABOR-HOURS TO PRODUCE	MEXICO	UNITED STATES
Agricultural goods (1 ton of tomatoes)	50	40
Farm equipment (1 tractor)	300	200
Relative price (tons of tomatoes per tractor)	6	5

Mexico and the United States also trade farm equipment, say, tractors, and agricultural goods, say, tomatoes. These are goods, unlike airplanes and oil, that both nations are capable of producing. We can build a highly simplified model of trade in these goods between these two nations. The labor required to produce each product differs in the two countries, as shown in Table 18–1. By assumption, it takes fewer labor hours to produce either product in the United States. The United States is said to have an **absolute advantage** in both goods, the result, in this example, of a labor force that is more productive than Mexico's in producing both tractors and tomatoes. In our earlier case of airplanes and oil, the United States had an absolute advantage in airplanes and Mexico had one in oil. In the tractor and tomatoes example, Mexico has an **absolute disadvantage** in both goods. Yet even under the conditions described in Table 18–1, both nations are better off engaging in trade.

Without trade, Mexicans would have to produce six tons of tomatoes to buy one tractor in their home market because each tractor takes 300 labor-hours to produce compared to 50 labor-hours for one ton of tomatoes. In the United States, tractors sell for the equivalent of five tons of tomatoes because each tractor requires 200 labor-hours to produce compared to each ton of tomatoes, which requires 40 labor-hours.[5] Since each nation's **autarky prices** (domestic prices without trade) differ, there is the opportunity for mutually beneficial trade. Trade will occur when the price of tractors in terms of tomatoes lies in between the Mexican and United States autarky prices. Under free trade, the forces of supply and demand guarantee that prices will fall within this range.

Assume that both nations agree to trade one tractor for 5.5 tons of tomatoes. By selling to the United States, Mexico needs to give up only 5.5 rather than 6 tons of tomatoes to get a tractor. Thus Mexico is better off by switching its labor into producing more tomatoes and selling them to the United States. Instead of working 300 hours to produce one tractor, Mexican labor can work for 275 hours (50 hours/ton times 5.5 tons of tomatoes) and import one tractor. The 25 hours that are saved through trade can be used to produce more tomatoes and consume more of either good. The surprising result is that the United States also is better off if it buys tomatoes from

[5]This way of calculating relative prices, in Mexico tractors selling for six times the price of a ton of tomatoes and in the United States for five times, works in this oversimplified example because labor is the only input into production and production costs are the only determinant of relative prices.

Mexico and sells tractors in return, even though it can produce both tractors and tomatoes at home with fewer hours of labor. U.S. workers can spend 200 hours to produce either one tractor or 5 tons of tomatoes. If U.S. labor is shifted away from farming and into tractor production, one tractor exported to Mexico is equivalent to producing 5.5 tons of tomatoes, half a ton more than those 200 U.S. labor-hours could produce themselves. U.S. firms can produce enough tractors to satisfy domestic demand and export to Mexico, and U.S. consumers get to consume more tomatoes. Both Mexico and the United States are better off in the sense that for the same amount of effort, measured in labor-hours, they can consume more goods by engaging in trade.

Ricardo's insight was that trade would be mutually beneficial if each nation pursued the good in which it is *relatively more productive*—that is, the good in which it has a **comparative advantage**. In Mexico the opportunity cost of one tractor is six tons of tomatoes (or for one ton of tomatoes, one sixth of a tractor.) In the United States the opportunity cost of one tractor is five tons of tomatoes (or for one ton of tomatoes, one fifth of a tractor.) The United States gains by trading tractors and getting something more than five tons of tomatoes in return. Mexico gains by trading tomatoes and getting something better than one sixth of a tractor in return. The gains are the higher level of consumption trade permits. As economist Paul Krugman, who won the Nobel Prize in economics in 2008 for his pioneering work on international trade, has written, "imports, not exports are the purpose of trade. That is, what a country gains from trade is the ability to import things it wants. Exports are not an objective in and of themselves; the need to export is a burden that a country must bear because its import suppliers are crass enough to demand payment."[6]

The theory of comparative advantage may seem straightforward and intuitive, but it is often misunderstood. Whenever someone claims it is unfair for workers in high-income economies to have to compete with low-wage workers in developing nations, they are not appreciating the benefits of international trade between rich and poor nations. It is still true, as we discuss later, that trade creates winners and losers within each country; however, the theory of comparative advantage suggests that when countries trade freely, the total gains to each country outweigh the losses. Similarly, when critics refer to international trade as a competition, implying that some nations win and others lose as if trade were a zero-sum game, they are misunderstanding the mutually beneficial gains that follow from comparative advantage. And if someone says that no nation can compete with China's exports of manufactured goods they are both missing the reality of global exchange and do not understand Ricardo's basic insight.[7]

[6]Paul Krugman, "What Do Undergrads Need to Know about Trade?," *American Economic Review* 83, no. 2 (May 1993).

[7]In "Ricardo's Difficult Idea: Why Intellectuals Don't Understand Comparative Advantage," in G. Cook, ed., *The Economics and Politics of International Trade* (London: Routledge, 1998), economist Paul Krugman demonstrates how even sophisticated policy makers and commentators often miss the basic insight of comparative advantage.

THE BENEFITS OF TRADE

The core of comparative advantage is that both countries gain from trade whenever the relative prices of commodities in each country differ in the absence of trade. Once the two countries begin to trade, the relative prices of commodities begin to shift until they are the same in the two countries. This is because the model assumes no tariffs, quotas, or other barriers to free trade and it assumes that transport costs are negligible. In the example in Table 18–1, the relative price of tractors in terms of tons of tomatoes settles somewhere between five and six. The final world price that will prevail in both countries under free trade will be closer to the initial price in the market of the country whose economy is larger.

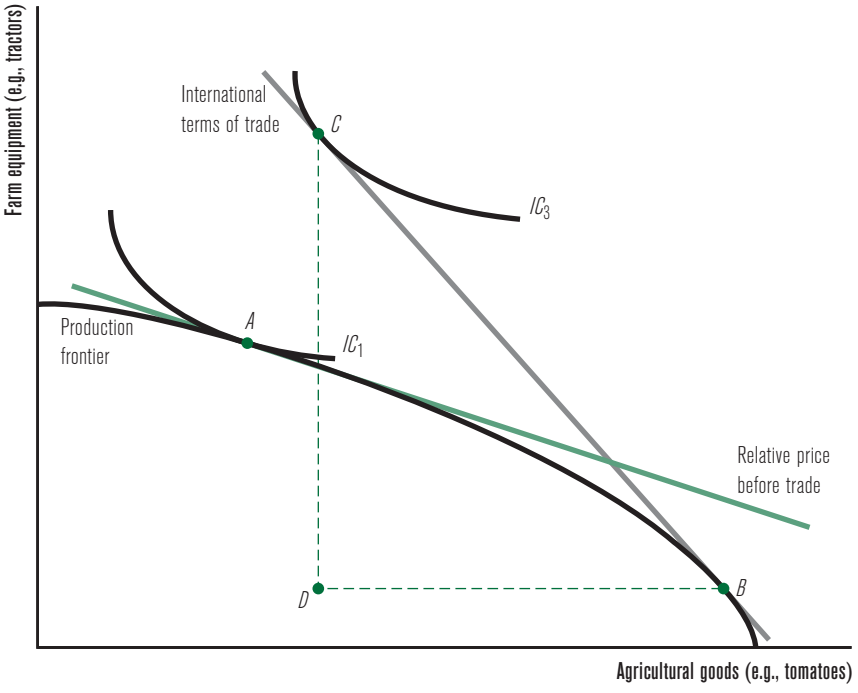
In our example, the final price in both countries would probably not settle at 5.5 tons, equidistant from the Mexican and U.S. autarky prices, but closer to 5 tons of tomatoes per tractor. This is because the U.S. economy is so much larger than that of Mexico and would exert more influence on both the supply and demand of each good. One implication is that small countries may benefit more from trade because the relative price of commodities shifts more, and therefore the gains from trade are greater. In our example, Mexico would be better off if the world price settled at 5.1 rather than 5.5 tons of tomatoes per tractor, because it would have to give up fewer tomatoes (and the labor time to produce them) for each tractor. Where world prices settle, often referred to as the **international terms of trade** (the ratio of the prices a nation receives for the goods it exports relative to the prices it receives for the goods it imports) is an important determinant of the gains from trade and is discussed in more detail later in this chapter.

The theory of comparative advantage as posed here is in the very simple form developed by Ricardo during the nineteenth century: two countries, two goods, and only one factor of production (labor).[8] Some of the complexities of the real world can be incorporated into the theory. A trading world of many countries can be handled by taking the home country, say, Kenya, and treating the rest of the world as its trading partner. Swedish economists Eli Heckscher and Bertil Ohlin expanded the theory during the first half of the twentieth century to deal with two factors of production, such as labor and capital. The Heckscher-Ohlin model leads to an extremely important result: A country tends to export products that use its abundant factors of production more intensively and imports products that require relatively more of its scarce factors.

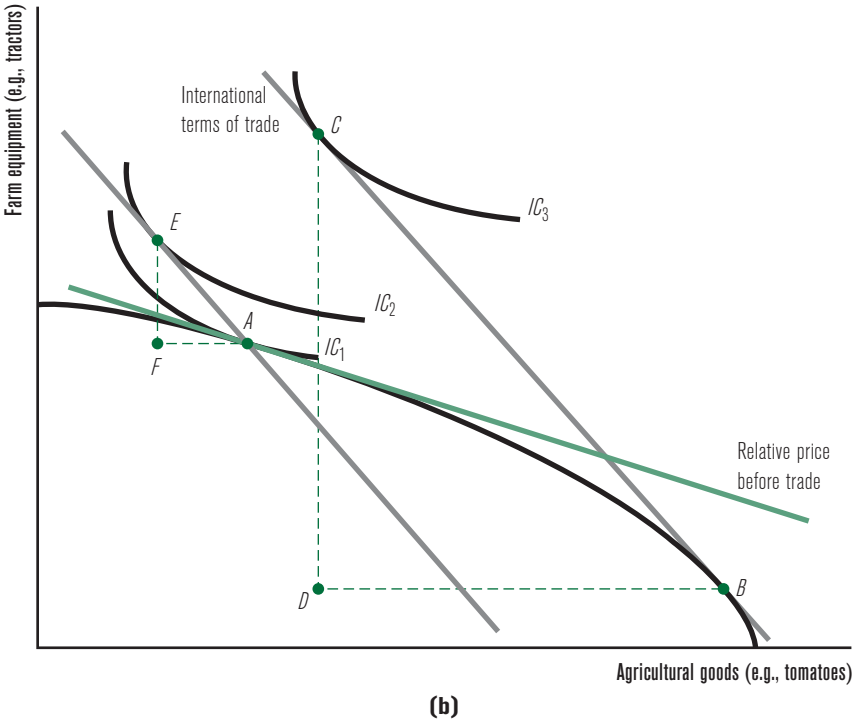
This more general approach to comparative advantage is encapsulated in Figure 18–6. The economy of the **home country** is divided into two goods, one that will be exported (tomatoes) and one that will be imported (tractors). The country's collective utility in consuming these goods is represented by the community indifference curves.

Without trade, the home country achieves its greatest utility by producing and consuming at point A, the tangency of the community indifference curve IC_1 and

[8]Ricardo's example, fitting his era, relied on England and Portugal trading wine and cloth.



(a)



(b)

FIGURE 18-6 Gains from Trade

the production frontier. Because it is producing on its frontier, all resources are fully employed. The slope at *A* determines the domestic relative price of tomatoes in terms of tractors. Assume that the rest of the world is better endowed with capital than labor relative to the endowments of the home country and that world consumers have tastes broadly similar to those of the home country. Then, in world markets, the relatively higher production of tractors compared to the demand for tractors drives their prices lower than in the home country, and the relatively lower production of tomatoes compared to the demand for tomatoes drives their price higher than in the home country. Because only relative prices matter, these two statements mean the same thing: In world markets, the price of tomatoes in terms of tractors is higher than in the home country leading to the international terms of trade being steeper than the autarky price line.

This difference in relative prices between the home country and the rest of the world presents an opportunity for the home country to improve its welfare through trade. *With trade*, the country, taking advantage of its factor endowment, can produce more tomatoes and fewer tractors and can sell tomatoes on the world market at the higher relative price. By producing more tomatoes, the home country moves along its production frontier from *A* to *B*, where the international terms of trade line is tangent to the production frontier. With tomatoes to export, the country can import tractors and consume more of *both* goods and be better off. The consumption point *C* is determined by the tangency of the international terms of trade line with the community indifference curve IC_3. The country now exports *BD* tons of tomatoes and imports *CD* of tractors; the amounts for both goods equal to the difference between consumption at *C* and production at *B.* (Note that the ratio CD:BD is an exact description of the international terms of trade.) As indifference curve IC_3 is above indifference curve IC_1, the country is better off with trade than without it. Note that both with and without trade the home country produces along its production frontier and all resources are fully employed. By pursuing its comparative advantage, the home country, exerts the same level of effort because all resources are fully employed in both cases but consumes more with trade than without trade.

Figure 18–6 can be used to gain some further insight into the sources of the **gains from trade**. In Figure 18–6b, the international terms of trade line is drawn through point *A*, the equilibrium without trade. This depicts the opportunities a nation faces by trading with other nations even when it does not change production. Under such a scenario, the nation could still gain from trade benefiting from the higher prices the rest of the world is willing to pay for its export good and the lower prices the nation would pay for its imports. The country could now move from the equilibrium at point *A* to the one at point *E*, exporting *AF* tons of tomatoes in return for *EF* tractors. This exchange permits the nation to move from indifference curve IC_1 to IC_2 and be better off. (In this example, note that the ratio EF:AF = CD:BD = the international terms of trade.) The implied increase in welfare is called the **gains from exchange** because it is due only to the difference between autarky and world prices.

But facing the opportunities trade provides, a nation will also reallocate its resources as it pursues its comparative advantage. At point *A* producers in the home country are not maximizing profits and consumers are not maximizing utility (both of which are requirements for equilibrium in this type of general equilibrium model). It will not remain producing at point *A* when other nations are willing to pay more for its exports and to sell imports for less. The change in production from point *A* to *B* maximizes producers' profits and results in an additional improvement in welfare depicted in the change in consumption from point *E* to *C*, and the movement from indifference curve IC_2 to IC_3. These are the **gains from specialization**. The gains from trade are therefore made up of the gains from exchange plus the gains from specialization.

Figure 18–6 shows the gains from exchange and specialization, but trade confers many more benefits to a nation that are not illustrated by the figure. First, trade exposes domestic firms to competition. Especially in small economies, there may be a limited number of producers of particular products. This may enable them to exercise market power, supplying less and charging more for their goods. Faced with competition from imports, market power is eroded, prices fall, and consumption and welfare increases. Trade disciplines domestic firms that wield market power. Competition from foreign firms can also increase the productivity of domestic firms. Faced with competition from abroad, firms may invest less in rent seeking, gaining special favors from government to protect their market power, and more in increasing efficiency and lowering costs. In competitive markets, the most productive firms survive and in so doing improve the use of scarce resources and raise output levels.

Second, trade, especially in intermediate goods including machinery and other forms of capital, often embodies new technologies that raise productivity. In our example of trade in tomatoes and tractors, if imported tractors contain better technology (for example, higher fuel efficiency or longer-lasting tires or other parts), Mexican farmers will become more productive, producing more tomatoes per hour, because of the imported technology embodied in imports that is absent in domestic farm equipment. Referring to Figure 18–6, trade-induced productivity effects such as these would shift the home nation's production frontier outward, a graphical reflection of economic growth. These benefits are in addition to trade enabling a nation to consume beyond its production frontier.

Third, trade increases not only the amount of goods a nation can consume but also the quality and variety of goods available. Before their transition, tropical fruits, especially bananas, were viewed as an exotic luxury in countries such as East Germany and Poland. An opening to trade made such products commonplace, increasing consumer welfare. In addition to final consumption goods like bananas, trade increases the variety of other goods available, including medicines. Cellphones provide an example of the benefits trade can confer to even some of the poorest workers. Without trade, cellphones might not be available because domestic firms may not have the knowledge and technology to produce them. But because of imports, cellphones can

be found in even remote areas. In India, traditional fishermen often carry cellphones with them and use them to communicate with other fishermen about where to fish and to find out about available prices for their catch at various local ports. Cellphones, a product that would have been unavailable unless imported, have eroded the market power of the fish buyers onshore and improved the livelihoods of the fishermen.

Fourth, trade also brings people into contact with one another. Whether this promotes peace, global understanding, and democracy remains an intriguing hypothesis.[9] Added to Ricardo's insight about exchange and specialization based on comparative advantage, there are many other important benefits poor nations can realize from international trade.

WINNERS AND LOSERS

Any country, whatever its size and stage of development, can benefit from trade. This is true for large countries like China and India, small countries such as Laos and Togo, high-income countries like Japan and the United States, and low-income countries such as Haiti and Mozambique. Ricardo's theory uses two goods and two nations with no restrictions on what type of nation is involved. As long as relative prices at home differ from those on world markets (or would differ in the absence of international trade), countries can increase their aggregate welfare by engaging in international trade.

Many of you may read these conclusions and remain skeptical and unconvinced. If the economic argument that trade is such a good thing is true, why is there so much opposition to globalization, the WTO, and free trade? What about sweatshops and the harsh conditions in the maquiladoras, a Spanish term used in Mexico and elsewhere that refers to factories, usually just over the border, that assemble goods for richer nations? What about the loss of jobs, factories, and a way of life throughout the old industrial belt in the European Union and United States? Trade played some role in these outcomes. Clearly the benefits of trade that stem from economic theory are missing an important part of the story.

The answer lies, in part, in the distributional consequences of trade. Comparative advantage as described by Ricardo focuses on the efficiency gains of trade, on how a given amount of resources can result in more total consumption. Although every country can gain from trade in the aggregate, *not all individuals or groups within each country necessarily gain.*

In the tomatoes for tractors example, Mexican vegetable growers gain from trade because they can sell more tomatoes at a higher price. Mexican consumers of tractors also gain, because they can now purchase imported tractors at a lower price. But, once trade begins, Mexican tractor manufacturers face competition from imports

[9]Douglas Irwin, *Free Trade under Fire,* 3rd ed. (Princeton: Princeton University Press, 2009), chap. 2, provides an excellent survey of the benefits of trade.

and hence sell fewer tractors at a lower price than before, whereas Mexican consumers of tomatoes must pay higher prices. (By an analogous argument, the price changes brought about by trade benefits tractor producers and tomato consumers in the United States but hurt U.S. tractor consumers and tomato growers.) If firms producing tractors (tomatoes) in Mexico (the United States) shut down in the face of imports, workers in those firms will be hurt, at least in the short run, and so may the communities in which they live. If there are such clearly defined losers from international trade, how can we be so sure that trade is a good thing?

The explanation lies in the core insight of comparative advantage. The theory tells us that the *aggregate gains* from trade outweigh the losses for the country as a whole. This is because the value of output in total rises, implying that the gains of the winners are greater than the losses of the losers. The winners gain enough to compensate the losses of the losers. But that may be cold comfort to Mexican producers of computers and consumers of tomatoes because there is no guarantee that they will be compensated for their losses. These are the distributional consequences of trade that are the seed of the political opposition to policies that promote freer trade.

The work of Heckscher and Ohlin leads to further important insights into the distributional consequences of trade.[10] If a nation's comparative advantage lies in those goods that intensively use a nation's abundant factor, than that factor gains from trade while the less abundant factor loses. If Mexico is labor abundant relative to the United States, then Mexican labor will eventually gain from trade no matter if it is employed in the production of tomatoes or tractors. This is because trade reallocates factors of production, from point *A* to *B* in Figure 18–6. Because the economy is now producing more of the labor-intensive good (tomatoes), the aggregate demand for labor rises and increases its return.[11] Of course such transitions are not seamless. Some workers, especially older ones, will find it hard to move and take on new jobs in the export sector. Trade often benefits the next generation, which is better able to take advantage of the new opportunities whether in farms now producing for export rather than subsistence, in factories making goods for overseas markets, or in call centers providing tradable services.

However suggestive the theory of comparative advantage may be for developing countries, it is only the beginning of an explanation of development through international trade. Ricardo's discussion of comparative advantage assumes *static conditions* that hold the factors of production in fixed supply and unable to cross borders. But economic growth is fundamentally a *dynamic process*. To explain growth and

[10]These insights are originally attributed to Wolfgang Stolper and Paul Samuelson, "Protection and Real Wages" *Review of Economic Studies* 9, no. 1 (1941).

[11]In the developed nation, which generally is capital abundant relative to its developing nation trading partner, comparative advantage lies in the capital-intensive good, and capital, not labor, is the factor that benefits from trade. This helps explain why labor in developed nations, especially less-skilled labor, and politicians who seek the vote of such groups, are often opposed to trade agreements and in favor of protecting domestic industries.

structural change we need to account for changes over time in the amount of capital, land, and labor available to producers as well as improvements in the quality or productivity of those factors. This is something comparative advantage alone cannot do. The theory of comparative advantage also does not capture the two-way relationship between trade and development. In one direction, increased trade can lead to improved welfare and higher incomes. In the other direction, advances in development go hand in hand with more highly skilled labor and higher-quality capital and machinery, opening up new opportunities for increased trade. To better understand how trade and development affect each other we need to examine some of the alternative trade strategies nations have adopted, some of which are consistent with comparative advantage and others not.

TRADING PRIMARY PRODUCTS

Most world trade involves manufactured goods. But primary products, whether agricultural raw materials, food, fuels, minerals, or ores, still represent about one third of the value of all traded goods. Figure 18–7 identifies economies that are dependent on exports of primary products, that is, where more than 75 percent of their merchandise exports is in primary products. This group includes 13 low-income, 27 middle-income, and 9 high-income economies. For most of these economies petroleum is the dominant export. But for others it might be coffee (Nicaragua), copper (Mongolia), or cotton (Burkina Faso). Given how many nations continue to depend on primary product exports, it is worth examining the role of trade in primary products on economic growth and development.

There are additional reasons for focusing on trade in primary products. First, for most developing economies, international trade often began with primary products, and subsequent thinking about the impact of trade on development has its roots in the exchange of primary products from the south for manufactured goods from the north. Second, the growth performance of resource-rich economies often has been disappointing, referred to as **the resource curse** and seemingly the opposite of the predictions of comparative advantage. The rest of this chapter considers trade in primary products as an engine of growth, its potential and its failures, as well as its broader implications for whether movements in the international terms of trade support developing economies. Trade in manufactured goods is considered in Chapter 19.

The promise of primary product exports is clear. Many low- and middle-income nations have resources that are unavailable or severely depleted in high-income economies. Tropical fruits and beverages, including bananas and coffee, cannot be grown in Europe, Japan, and North America (except in greenhouses) but are in high demand. Diamonds and oil are obvious examples of primary products in abundance in some low- and middle-income economies that face limited demand at home

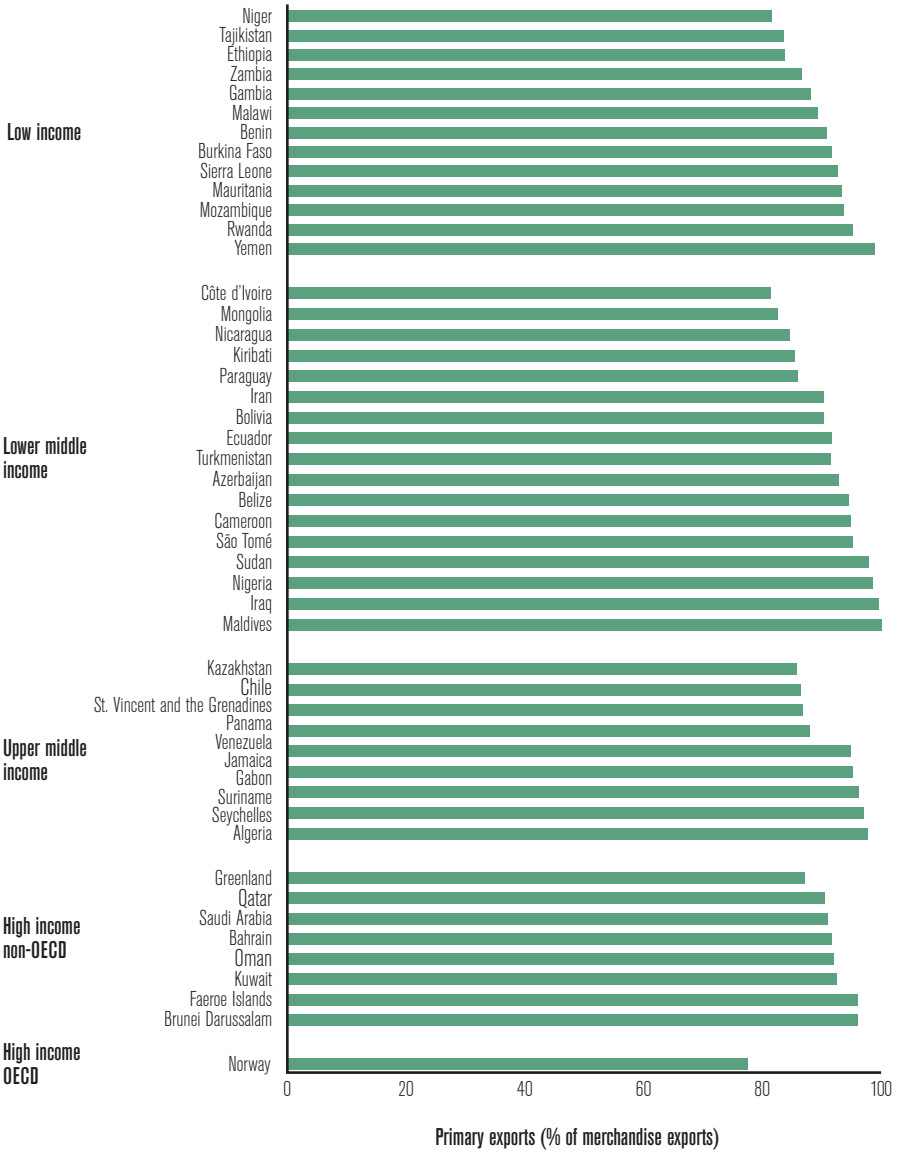


FIGURE 18–7 The Importance of Trade in Primary Products

OECD, Organization for Economic Co-Operation and Development. Included are countries where primary products account for 75 percent or more of all merchandise trade.
Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

but significant demand abroad. Copper deposits in Chile and Zambia, coltan in the Congo (a mineral used in making the electronics for cellphones), and palm oil grown throughout Southeast Asia offer further examples.

Referring to Figure 18–6, we can substitute resource-intensive primary products on the *x*-axis and manufactures on the *y*-axis. Without trade, a nation will produce

some combination of the two goods, as represented by point *A*. Once trade becomes an option, a relatively resource-rich nation will devote more of its resources to the primary goods sector, shifting capital and labor to produce more of its primary products, where it has a comparative, if not an absolute, advantage. This permits the economy to consume more, as represented by point *C*. The gains can be even greater if land or some other natural resource lies idle. If this is the case, without trade the nation might find itself in an interior point well within the production frontier. Trade creates an opportunity to take advantage of such idle resources and improve domestic consumption outcomes even more.

Burmese economist Hla Myint long ago observed that when parts of Africa and Asia came under European colonization the consequent expansion of their international trade enabled those areas to use their land or labor more intensively to produce tropical foodstuffs such as rice, cocoa, and oil palm for export. Myint applied Adam Smith's term **vent for surplus** to these cases.[12] The concept applies to a situation in which a country has the capacity to produce more than it can sell on the domestic market. Trade enables the country to employ either land or labor more fully and sell the goods produced with its "surplus" land and labor to the rest of the world. In many ways, this describes the economic development of Australia, Canada, and the United States during the nineteenth century.

In addition to helping an economy use its *current* factor endowments more intensively and more efficiently, the expansion of primary product exports can lead to the accumulation of additional factors of production, including capital and labor. Primary export–led growth can spur increases in foreign investment to complement the fixed factors of production, land, and natural resources. In the context of Figure 18–6, the country would be able to produce more because its production frontier would shift out.

Once profitable opportunities in tropical agriculture or natural resources become apparent, foreign investment is likely to be attracted to the country to exploit the country's comparative advantage. The influx of foreign investors has been a familiar story in all mineral-exporting industries and in many tropical-product industries based on plantation agriculture. Well-known historical examples include Standard Oil in Venezuela, British Petroleum in Iran, Anaconda in Chile, Alcoa in Jamaica, Firestone in West Africa, and United Fruit in Central America.[13] All these examples resulted in an increase in primary product production and exports. They also often involved political manipulation by these foreign firms in the domestic affairs of their host country, leaving a legacy that made other countries wary of foreign investment. The situation today, generally, is viewed differently, with governments able to negotiate

[12]Hla Myint, "The 'Classical Theory' of International Trade and the Underdeveloped Countries," *Economic Journal* 68 (1959), 317–37.

[13]Anaconda developed Chile's copper mines, Alcoa pursued Jamaica's bauxite deposits, Firestone operated rubber plantations in Liberia and elsewhere for use in tire production, and United Fruit focused on bananas throughout Central America.

better terms for the development of natural resources and the sharing of **resource rents**.[14] (Foreign direct investment was discussed in Chapter 10.)

Another potential benefit from primary product exports is the possibility of stimulating production in other, related sectors. The very notion of export-led growth implies that exports would lead to more broad-based economic growth. Several types of linkages to the rest of the economy are possible, including to upstream or downstream industries, increased production of consumer goods, enhanced infrastructure, more widely available skilled labor, and increased government revenues. Albert O. Hirschman coined the phrase **backward linkages** for the situation in which the growth of one industry (such as textiles) stimulates domestic production of an upstream input (such as cotton).[15] Expanded production of primary products also can stimulate **forward linkages** by making lower-cost primary goods available as inputs into other industries. In many developing countries, agricultural products are used as inputs to the food processing industry. Senegal and the Gambia process raw groundnuts (peanuts) into shelled nuts and cooking oil; in Indonesia, forest products are used in furniture production. Forward linkages can develop for mining and mineral products (such as, petroleum refining, plastics, or steel), but the more complex production techniques and demand for highly specialized capital and labor make it difficult for many developing countries to compete in these activities. Furthermore, developed nations often employ **tariff escalation**, imposing higher tariffs on imports of processed goods than on the raw material, making it harder for the developing nation to move up the value-added ladder.

Infrastructure linkages arise when the provision of overhead capital (roads, railroads, power, water, and telecommunications) for the export industry lowers costs and opens new production opportunities for other industries. Primary export sectors also may stimulate human capital linkages through the development of local entrepreneurs and skilled laborers. The growth of the Peruvian fishmeal industry, with its many small plants, encouraged scores of new entrepreneurs and trained many skilled workers to operate and maintain equipment. These resources then became available for subsequent development. Rubber, palm oil, and tin production for export encouraged entrepreneurs in Malaysia, and small-scale farming for export has proven an outlet for entrepreneurial talent in several African countries. But perhaps the best case for petroleum, mining, and some traditional agricultural crops is the fiscal linkage. Governments can capture large shares of the economic rents from these exports as taxes, dividends, or through leasing arrangements and use the revenue to finance development in other sectors.

[14]Economists use the term *rent* to refer to the excess of revenues over the costs of production plus the normal profit margin. When commodity prices rise, as in the case of petroleum exports, resource rents also tend to rise because the cost of oil drilling often does not change appreciably.

[15]Albert O. Hirschman, *The Strategy of Economic Development* (New Haven, CT: Yale University Press, 1958), chap. 6.

EMPIRICAL EVIDENCE ON PRIMARY EXPORT-LED GROWTH

Since the 1950s, some economists and the leaders of many developing countries have argued that, despite the possible benefits, primary exports cannot effectively lead the way to economic development. The most common arguments have been that the markets for primary products grow too slowly to fuel growth, the prices received for these commodities have been declining, earnings are too unstable, linkages do not work, and the fiscal windfall more often than not breeds corruption and civil strife rather than growth and development. We examine these arguments in the next section. Here, we discuss the direct empirical question, What has been the relationship between primary exports and economic growth in recent decades? A few resource-rich developing countries have performed relatively well, including Botswana, Indonesia, Malaysia, Mauritius, and some of the oil-rich states of the Persian Gulf. However, many others have grown very slowly or not at all, or have experienced increases in GDP per capita, which have been accompanied by little economic development. Such nations are Angola, Burma, Ecuador, Equatorial Guinea, Jamaica, Nigeria, and Zambia. At the same time, many of the fastest-growing Asian economies are resource poor—for example, Korea, Singapore, Taiwan, and, more recently, India and Vietnam.

Using regression analysis, economists Jeffrey Sachs and Andrew Warner examined the relationship between primary product exports and economic growth in a sample of 95 countries from around the world between 1970 and 1989.[16] They ran growth regressions that isolate the independent effects of various determinants of GDP per capita growth, including initial per capita income, openness to trade, government efficiency, investment rates, inequality, and the share of primary product exports in GDP. They found strong evidence showing that, on average and holding constant the other determinants of growth, countries with substantial primary product exports have grown much more *slowly* than resource-poor countries. They found that an increase of 10 percentage points in the ratio of primary exports to GDP was associated with a 0.7 percentage point *slower* annual rate of growth of per capita income. Figure 18–8 illustrates the basic negative relationship over a slightly longer time period (1970–96). For the 27 countries in which primary exports (in 1970) were 5 percent of GDP or less, annual per capita growth averaged 2.7 percent. By contrast, annual per capita growth averaged less than 0.5 percent in the 16 countries with primary exports equal to 20 percent or more of GDP. This

[16]Jeffrey Sachs and Andrew Warner, "Natural Resource Abundance and Economic Growth," NBER Working Paper No. W5398, Cambridge, MA, December 1995.

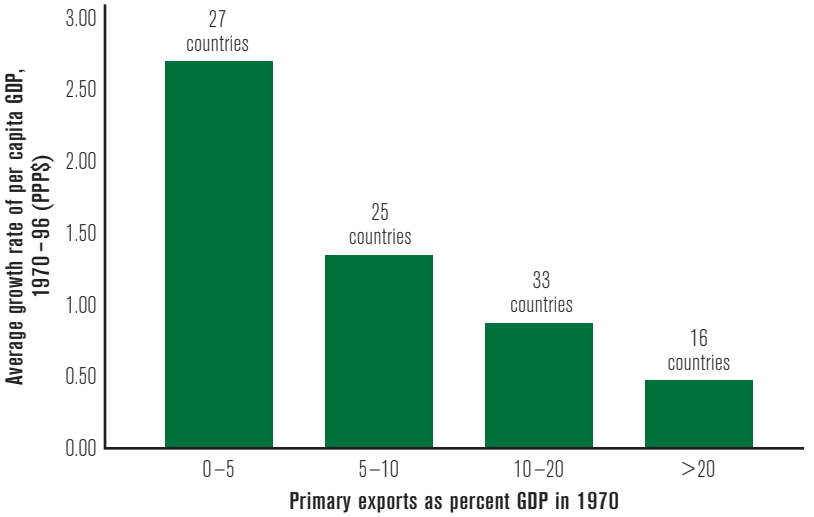


FIGURE 18–8 Natural Resource Abundance and Economic Growth

Countries with small amounts of primary product exports have grown more quickly than resource-abundant countries.

GDP, gross domestic product; PPP, purchasing power parity.
Source: Updated from Steven Radelet, Jeffrey Sachs, and Jong-Wha Lee, "Economic Growth in Asia," Harvard Institute for International Development, Cambridge, May 1997. Data from World Bank, *World Development Indicators* and *Penn World Tables*, mark 5.6.

represents a massive difference in the rate of economic growth and the potential for economic development.[17]

How can resource abundance be associated with slower rates of economic growth? Following the law of comparative advantage and capturing the dynamic gains from trade, primary product exports should help raise aggregate income and allow a country to increase imports and investment. There are several possible explanations for this apparent puzzle.

EXPORT PESSIMISM

Long before the growth regression results of Sachs and Warner and well before "the resource curse" became part of the vocabulary of development economics, there were reservations about primary product exports as an engine of growth for poor nations. It is worth reviewing this history because some of its themes still resonate today. What became known as **export pessimism** has its origins in work first published in the 1950s

[17]A recent study confirms these findings: "Developing countries, which in 1980 derived more than 70 percent of their export revenues from non-fuel primary commodities, increased their per capita GDP by only 0.4 percent a year between 1980 and 2006, and countries that mainly exported fuels raised their per capita GDP by 1.1 percent a year. By contrast, more-diversified exporters achieved per capita growth of 1.6 percent a year." World Bank, *Global Economic Prospects* (Washington, DC: World Bank, 2009), p. 98.

by economists Raul Prebisch, Hans Singer, and others.[18] These economists argued that, over the long run, prices for primary commodity exports on world markets tend to fall relative to prices of manufactured goods. Because at the time the Prebisch-Singer hypothesis was advanced, developing countries mostly exported primary products and imported manufactured goods, such a shift in the international terms of trade implied that, over time, developing countries would have to export more primary products to import the same amount of manufactured products. Countries relying on primary product exports, therefore, would continue to lag behind in the development process as the gains from trade would become more and more limited.

The root of the relative price movements suggested by Prebisch and Singer were traced to structural factors in the global economy. First, there is Engel's law (named after Ernest Engel, a nineteenth-century German statistician who first proposed the idea) that the demand for staple foods grows more slowly than income. For high-income countries, the income elasticity of demand for food is below one half, implying that the demand for food (including food imported from developing countries) tends to lag far behind income growth. At the same time, the demand for manufactured goods is income elastic, causing demand and hence prices to grow relatively more rapidly for such goods. The implication is that developing nations would wind up earning relatively less for their exports and paying relatively more for their imports.

Second, technological change in manufacturing also works against the demand for raw materials and the nations that produce them. New technologies and improved machinery help firms reduce wastage so that less raw material is needed in the production process. Modern looms waste less cotton yarn, new mining techniques are more efficient, and so forth. Perhaps even more significant, new technologies allow for the substitution of synthetics for raw materials. Synthetic rubber long ago replaced natural rubber in automobile tires and other applications; fiber optics has substantially reduced the use of copper in wire; and automobiles use fiberglass instead of steel, decreasing the demand for iron ore. There is widespread evidence that industrial production has grown more rapidly than the consumption of raw materials.

Third, Prebisch argued that manufacturing firms in developed countries tended to have market power, whereas primary producers in developing countries faced much greater competition. As a result, prices for manufactures were less flexible than those for primary goods. With productivity growth, primary producers would face lower prices due to competition from other producers, but this would not happen to the same degree with manufactured goods. In manufactures, productivity growth might raise profits rather than lower prices. This reasoning provided one more reason why poor nations might expect declining terms of trade and export pessimism.

[18]United Nations (Raul Prebisch), *The Economic Development of Latin America and Its Principal Problems* (Lake Success, NY: United Nations, 1950); Hans W. Singer, "The Distribution of Trade between Investing and Borrowing Countries," *American Economic Review* 40 (May 1950), 473–85. See also the later reflections of these two economists in G. Meier and D. Seers, eds., *Pioneers in Development* (Washington, DC: World Bank, 1985).

Prebisch, Singer, and the others concerned about trade and development based some of their findings on characteristics of the global economy that existed in the first half of the twentieth century and no longer hold. They also did not focus on such commodities as diamonds, natural gas, oil, tropical woods, and other primary products that face either relatively income elastic demand, shrinking supply, or few synthetic substitutes. They did not expect primary producers to be able to transition to manufactured goods as have many low- and middle-income economies. Nor did they envision the fierce global competition in manufactured goods that exists today or the cartels that influence the prices of some primary product markets. The export pessimism they expressed half a century ago, which went on to influence the trade strategies of many nations, is mostly rejected today. Too many low- and middle-income nations have benefited from exports as an engine of growth. But Prebisch and Singer's attention to the terms of trade and its implications for developing nations remains an important issue, one worthy of further examination.

DECLINING TERMS OF TRADE?

Concern over the direction of the terms of trade is easy to understand within the context of the theory of comparative advantage; but it is *not* a critique of that theory. Figure 18–9 reproduces Figure 18–6a and adds a decline in the terms of trade. The

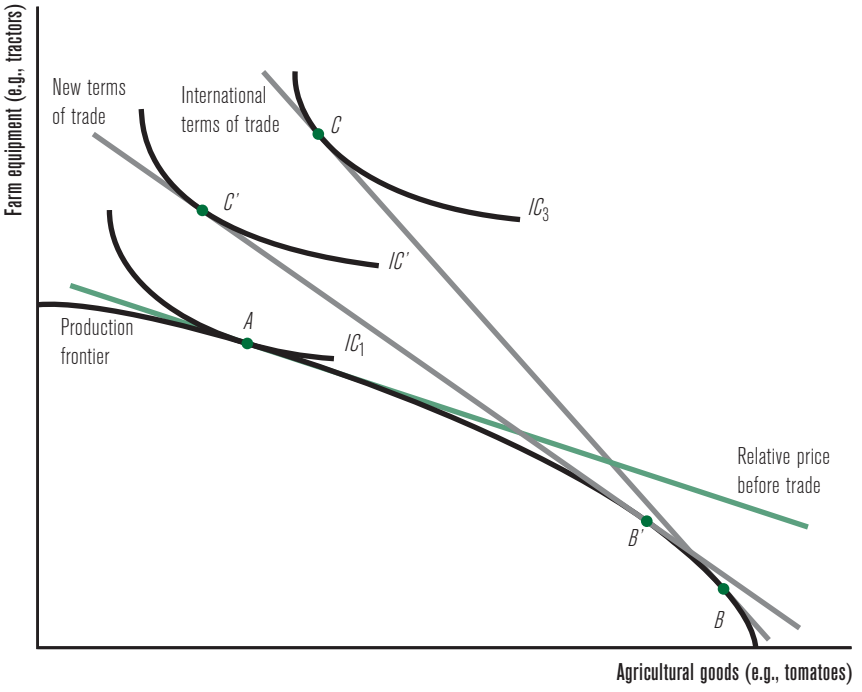


FIGURE 18–9 **Declining Terms of Trade**

counterclockwise rotation in the world price line indicates that the relative price of the exported good (tomatoes) has fallen relative to the price of the imported good (tractors). With these price changes, the nation exporting agricultural goods is worse off than before. It captures fewer gains from specialization, now producing at *B'* rather than at *B*, and lower gains from exchange and finds itself on a lower community indifference curve, *IC'*.

According to Figure 18–9, any country experiencing a decline in its terms of trade still realizes gains from trade. The nation is better than if it returned to autarky at point *A* and did not engage in any trade. But it would obviously prefer the initial terms of trade that permitted consumption at point *C* rather than at point *C'*. While the shift in the terms of trade in the figure diminishes the gains from trade for the nation exporting agricultural goods, the opposite would be true for the nation importing tomatoes and exporting tractors. For this nation, which in the Prebisch-Singer view would be a developed economy, the gains from trade would increase. No wonder some export pessimists called for a new international economic order, which would improve the benefits the developing nations received from global trade. Such concerns continue to be expressed by those both attending and protesting at meetings of the WTO.

In the simple two-good, two-country model the impact of the terms of trade on the gains from trade is straightforward. The empirical record, in a world of multiple products and nations, is somewhat less so. We begin by looking at trends in commodity prices. Figure 18–10 presents the trend from 1960 to 2009, in an index of commodity prices compiled by the United Nations Conference on Trade and Development (UNCTAD). The index refers to a broad range of commodities, excluding oil but including everything from coffee, cotton, timber, and tobacco to nickel and zinc. Food products account for about two thirds of the value of the index. Two trends are apparent. There has been a general secular decline in the price of non-fuel commodities over the past 50 years. The fluctuation in commodity prices is an even more pronounced trend.[19]

To capture movement in the terms of trade for primary product exporters, we need to know more than the trend in commodity prices. What about the price of imports? The terms of trade for these nations can be constructed by dividing a commodity price index, capturing the price of exports, by an index for the unit value of manufactures meant to reflect the price of imports. Trends in such a terms of trade index look similar to those depicted in Figure 18–10. Relative to manufactures, non-fuel

[19]These trends are reinforced if one looks at even longer time periods. Paul Cashin and C. John McDermont analyzed the behavior of commodity prices from 1862 to 1999, using a data set on industrial commodity prices compiled by *The Economist*. They found a downward trend in real commodity prices of 1 percent per year over the 140-year period. What was more pronounced was the degree of price volatility that increased over time. They concluded that the downward trend in commodity prices was "of little practical policy relevance." It was the variability in prices that had profound consequences for achieving macroeconomic stability. Paul Cashin and C. John McDermont, "The Long Run Behavior of Commodity Prices: Small Trends and Big Variability," IMF Staff Working Papers 49, no. 2 (2002).

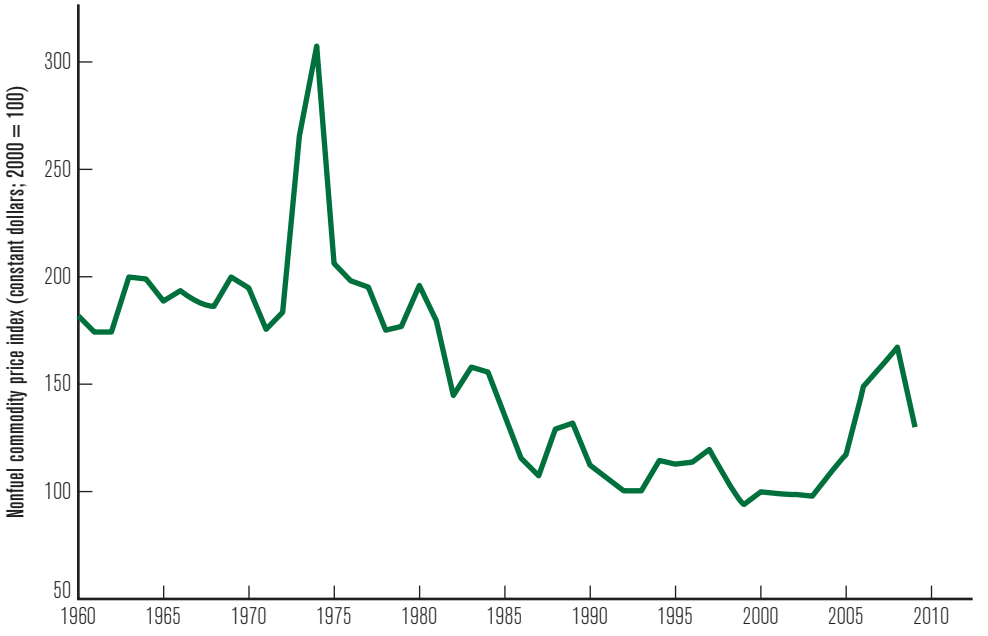


FIGURE 18–10 Nonfuel Commodity Price Trends, 1960–2009

Sources: United Nations Conference on Trade and Development, UnctadStat http://unctadstat.unctad.org/ReportFolders/reportFolders.aspx; International Monetary Fund, Primary Commodity Prices, *Index of Non-Fuel Primary Commodities,* available at www.imf.org/external/np/res/commod/index.aspx.

commodity prices have declined since the 1960s. Even more pronounced is the pattern of relative commodity booms and busts. In the 1970s there was a brief but sharp upturn in commodity prices; another commodity price boom began after 2003 but ended, at least temporarily, with the global recession in 2008.

Although the Prebisch-Singer hypothesis is properly addressed by looking at commodity data, the overall terms of trade for a country (including all products) is what really matters for its development. A commonly used measure of relative prices of traded goods is the **net barter terms of trade**, T_n. The term T_n is the ratio of two indexes: the average price of a country's exports (P_e), whether they are primary products or manufactures, and the average price of its imports (P_m). Figure 18–11 shows two different series of the net barter terms of trade since 1960. When oil exporters are included, the terms of trade of all developing countries rise dramatically after 1972 and remain high, despite the fall of oil prices during the 1980s and early 1990s. But the terms of trade for non-oil-exporting developing countries decline over the period by almost 30 percent. Over these decades, there is support for the Prebisch-Singer hypothesis about declining terms of trade facing non-oil exporting developing nations, though the experience of individual countries varies around this average trend.

More important, evidence on the poor long-run growth and development performance of most primary product exporters involves factors other than

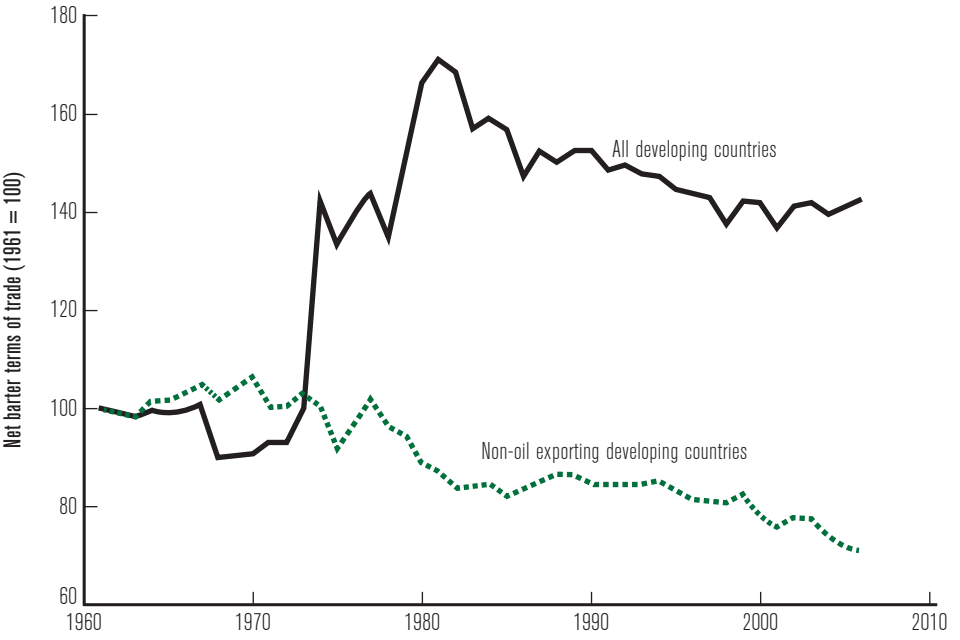


FIGURE 18–11 Net Barter Terms of Trade, 1961–2006

Source: IMF, *International Financial Statistics Yearbook 1987* (Washington, DC: IMF, 1987); IMF, *International Financial Statistics Yearbook 1996* (Washington, DC: IMF, 1996); and IMF, *International Financial Statistics Yearbook 2007* (Washington, DC: IMF, 2007).

movements in the terms of trade. For oil exporters and non-oil exporters alike, commodity price trends alone do not account for the resource curse. Other explanations are needed.

DUTCH DISEASE

Prebisch and Singer worried about declining terms of trade. But *improvements* in a nation's terms of trade brought on by booming primary export prices may be even more perilous for an emerging economy. The problem has been labeled **Dutch disease**, named after the experience of the Netherlands after 1960, when major reserves of natural gas were discovered. The ensuing export boom and the balance-of-payments surplus promised new prosperity. Instead, during the 1970s, the Dutch economy suffered from rising inflation, declining export of manufactures, lower rates of income growth, and rising unemployment. The oil boom of the 1970s and early 1980s produced similar paradoxes in a number of countries, including Mexico, Nigeria, and Saudi Arabia. Coffee booms in some African economies have followed a similar path. It is too early to tell if the commodity boom of the 2000s will leave a similar legacy, although there are indications that some of the mistakes of the past are not being repeated.

Economists began to realize that Dutch disease might be a very general phenomenon, applicable to all countries that enjoy export booms of primary

commodities.[20] Because the influx of foreign exchange itself causes the Dutch disease, the syndrome also can result from large inflows of foreign capital in any form. The import of gold and silver from the Americas, for example, may have helped retard Spain's growth in the sixteenth century. Similarly, large foreign aid flows to developing countries can undermine exporting firms' international competitiveness. How can this happen?

One key to this paradox is to understand how export booms affect a country's real exchange rate and how the real exchange rate, in turn, affects other industries. The official or **nominal exchange rate** is the price at which anyone holding foreign exchange, such as dollars, can convert it into local currency whether pesos in Mexico, naira in Nigeria, or rupiah in Indonesia. Most commonly, the nominal exchange rate is quoted in local currency per unit of foreign currency, say, pesos per dollar. When the exchange rate rises (that is, there are more pesos or naira or rupiah per dollar), the domestic currency is said to *depreciate* because it takes fewer dollars to buy the same amount of pesos, naira, or rupiah than before.[21] When the exchange rate falls, the domestic currency *appreciates* because pesos are now worth more in dollars than before.

The real exchange rate starts with the nominal exchange rate then incorporates the prices of tradable goods and **nontradable goods** to analyze how changing prices and exchange rates affect the incentives for production and consumption. Economists call any good that is or could be imported or exported a tradable, whether apparel or automobiles, oil or soybeans. Prices for tradable goods generally are determined on world markets and, in most cases, are not affected by domestic conditions in developing countries, each of which tends to be small relative to the size of the global market. Nontradables are goods produced domestically and not imported or exported. Prices for nontradables are determined by supply and demand conditions in the domestic market. Examples include construction, electricity, local transportation and personal services, the prices of which depend almost entirely on local conditions as these goods and services are not traded across borders. Even some manufactured goods or farm products, which are heavily protected and not subject to competition from imports, can be considered nontradables.

[20]Early theoretical work on the Dutch disease appears in W. Max Corden and S. Peter Neary, "Booming Sector and Deindustrialization in a Small Open Economy," *Economic Journal* 92 (December 1982), 825–48. Our late co-author, Michael Roemer, applied this theory to developing countries in "Dutch Disease in Developing Countries: Swallowing Bitter Medicine" (234–52), in Matts Lundahl, ed., *The Primary Sector in Economic Development* (London: Croom-Helms, 1985). A recent survey of the literature on Dutch disease and, more generally, the resource curse, is Frederick van der Ploeg, "Natural Resources: Curse or Blessing?" *Journal of Economic Literature* 49 (June 2011).

[21]The language on exchange rate changes sometimes is confusing. When the nominal exchange rate rises, it takes more domestic currency (naira) to buy foreign currency (dollars). Equivalently, this means that it takes fewer dollars to buy naira. In this latter sense, the naira is said to depreciate, or lose value, when the exchange rate rises.

The real exchange rate most commonly is defined as,

$$\text{RER} = (E_o \times P_T)/P_N \qquad [18\text{–}1]$$

where RER is an index of the real exchange rate, E_o is an index of the official (nominal) exchange rate, P_T is an index of the prices of tradable goods expressed in foreign currency (such as, U.S. dollars), and P_N is an index of the domestic price of nontradables expressed in local currency (such as, pesos.) Note that the RER is an index of prices, like the terms of trade, not a price itself. Also note that, because P_T is expressed in foreign currency, by multiplying by E_o, the numerator $(E_o \times P_T)$ becomes a domestic currency index of tradable's prices. Therefore, the RER can be thought of as the ratio of the price of tradable goods to the price of nontradable goods, all expressed in domestic currency. An increase in the RER—because of either an increase in the nominal exchange rate (more pesos per dollar), a rise in the dollar prices of tradable goods on world markets (more dollars per barrel of oil), or a fall in nontradables prices (fewer pesos for water, a haircut or school fees)—suggests that the relative price of tradables in domestic markets has risen. Such a shift should encourage firms to produce more tradable goods and discourage the consumption of tradable goods. Following the language of nominal exchange rates, a rise in the RER is referred to as a *real depreciation* of the peso, naira, or rupiah, and a decline in the RER is called a *real appreciation.* Official devaluations, which raise the peso price of dollars (E_o), cause the real rate to depreciate, at least initially. The same is true for an increase in world tradable prices. An increase in nontradables prices has the opposite effect, causing the RER to appreciate.[22] Thus while the nominal exchange rate tells you how many pesos you can buy with a dollar, the real exchange rate is an index describing the quantity of actual goods and services you can buy with a dollar when you trade with Mexico. When a country's RER appreciates (depreciates) that country becomes less (more) competitive in international export markets.

The Dutch disease is a wolf in sheep's clothing because it typically starts with what looks like a good thing: A boom in a country's raw material exports. But the boom in exports can cause a sharp appreciation of the RER, in two ways. First, the influx of foreign exchange from higher export earnings creates a surplus of foreign currency. Unless the central bank tries to maintain the official exchange rate at its former level, this increase in the supply of foreign exchange causes the market exchange rate to fall and the currency to appreciate in value. This is because the export windfall of dollars entering the country must typically be converted into local currency, and this increase in the demand for pesos bids up the dollar price of pesos, lowering the nominal exchange (E_o), causing a real appreciation of the peso relative to the dollar. Second, higher income from booming primary exports also spurs faster domestic price inflation. It does this because the additional income creates greater demand

[22]For a comprehensive discussion of the real exchange rate, see Sebastian Edwards, *Real Exchange Rates, Devaluation and Adjustment: Exchange Rate Policy in Developing Countries* (Cambridge: MIT Press, 1991).

for all goods and services in the economy. To the extent that this demand spills over into more imports, there is an outflow of foreign exchange but no inflation, because the price of imports is not much affected by demand in a single country. But prices for nontradable goods and services are likely to increase. Due to a limited supply of nontradables, especially in the first months or years of the boom, the greater demand results in higher prices for nontradables. (By way of an example, think of how a commodity boom could lead to rising domestic land and real estate prices.) From equation 18–1 we know that both the appreciation of the domestic currency and the rise in nontradable prices cause the RER to appreciate.

To understand how the real exchange rate becomes the key to the Dutch disease paradox, consider the impact of the real exchange rate on export industries *other than the booming primary export sector.* Booming commodity prices, stimulate more rapid domestic inflation and cause the RER to appreciate, rendering *other* exports less competitive and hence less profitable. Therefore, the disease is the deleterious effect of a commodity boom on other export sectors. Producers of tradables, both nonprimary product exporters and import competitors, face rising costs in their purchases of nontradable goods and services, including the wages of their workers. But they cannot charge higher prices because they compete with foreign producers, either as exporters or as import competitors. These farmers and manufacturers face a profit squeeze that causes some of them to reduce production and employment. The boom in primary exports and nontradables is partly offset by a contraction in other tradable industries. If the contracting industries have more long-run potential for productivity growth, the short-run benefits of the commodity price boom will have negative long-term growth consequences.

If it is relatively easy to move capital and labor between the booming commodity sector and other activities, and the booming sector can employ all the factors of production released from other, now less-profitable activities, then the commodity boom poses no major problem. As prices for the primary export product rise, labor and capital can move into the booming sector, and the economy is better off as a result. If the boom ends, the factors of production can move back to their previous activity. But this usually is not the case. If the booming sector is highly capital intensive (such as petroleum), few new jobs are created, so unemployment may go up (although the indirect employment effects of the boom in construction and other activities can mitigate this effect). More insidiously, when the boom ends (and booms always end), it is likely to be very difficult to move the factors of production back to their previous employment. Manufacturing activities cannot just restart overnight or, in many cases, can farming. In addition, the boom can lead to social or migratory shifts that are hard to unwind. Nigeria's oil boom in the 1970s encouraged rural workers to migrate to urban areas to look for new, high-paying construction and civil service jobs. When the oil boom ended, few urban workers wanted to return to rural areas. Many other oil exporters suffered similar fates when petroleum prices fell. (Box 18–1 illustrates the sectoral shifts common to Dutch disease.)

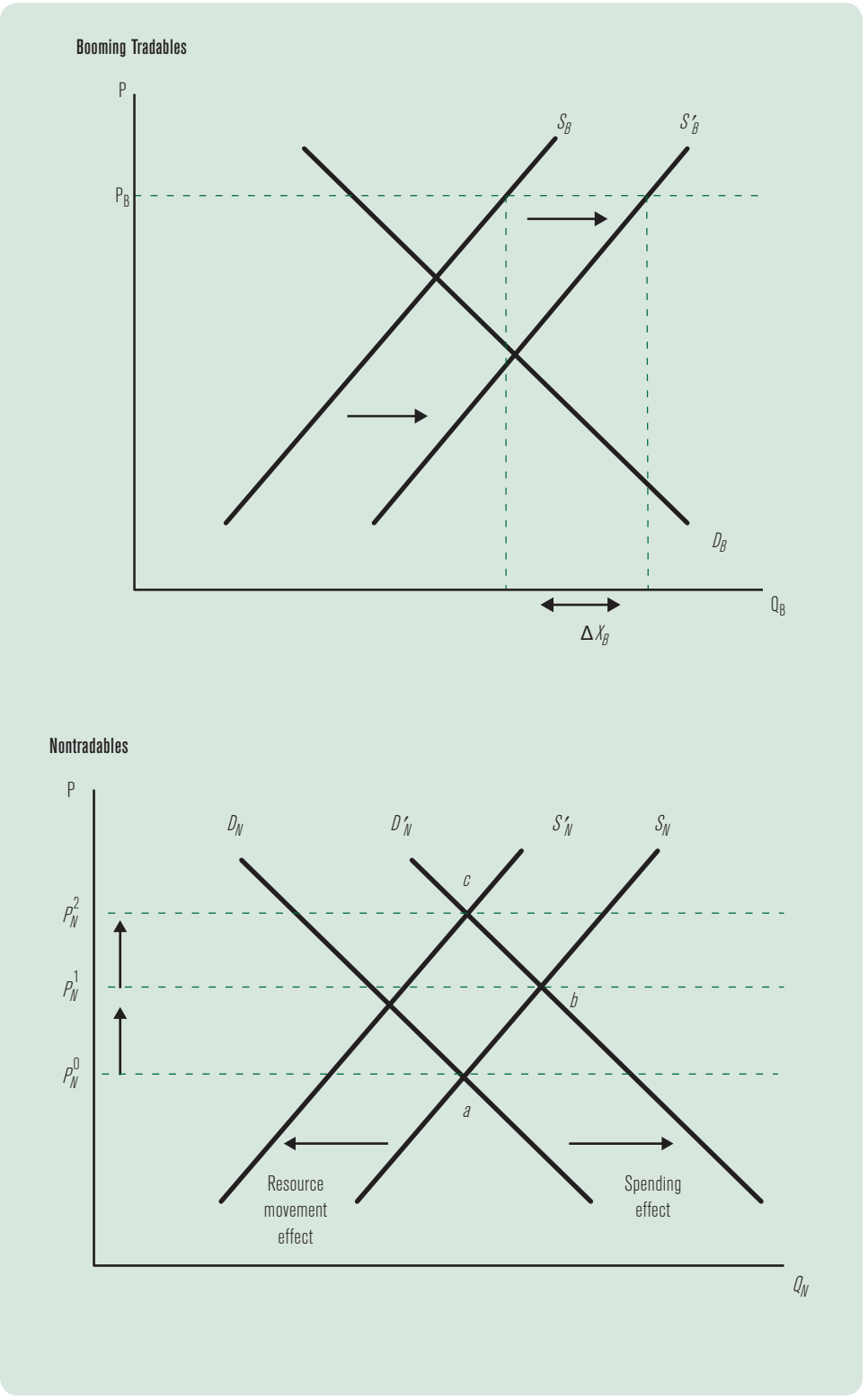
BOX 18-1 DUTCH DISEASE: A GEOMETRIC PRESENTATION

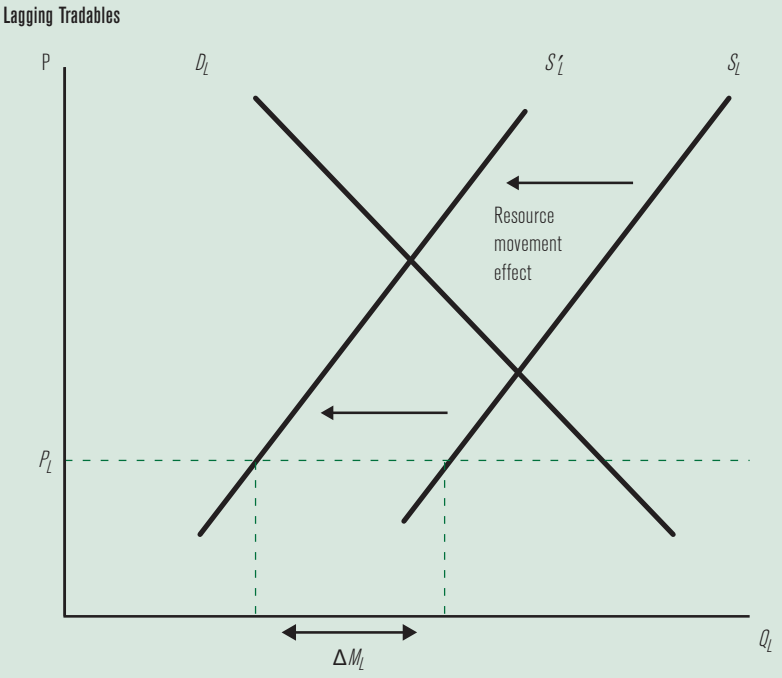
Dutch disease is best illustrated in an economy consisting of three sectors. Two of these sectors are tradables, the booming sector (B) and the lagging sector (L); the third sector is nontradables (N). Output in each sector is produced using a factor of production specific to that sector (for instance land in agriculture) and labor. Workers are assumed to move freely between sectors, implying that the equilibrium wage rate is the same in all three.

Suppose that the country experiences a positive trade shock, for example the discovery of large reserves of offshore oil as occurred in Ghana in 2007. We can understand the potential consequences of this shock by considering its effects on supply and demand in each of the economy's three sectors. The figure "Booming Tradables" illustrates the initial shock as a rightward shift in the supply curve in the booming tradables sector, from S_B to S'_B. This results in an increase in the quantity exported at the world price P_B by ΔX_B. (The positive trade shock could also occur due to a sudden increase in the world price of the booming sector export good, P_B, which would similarly increase the quantity of exports.) The country enjoys a windfall of new income.

Most likely, some of this windfall will be spent on nontradables. This spending effect is illustrated by the rightward shift in the demand curve for nontradables, from D_N to D'_N, in the figure "Nontadables." (Remember, although this small country takes the prices of tradable goods as given by world markets, the price of its nontradable goods is determined exclusively by supply and demand within the country.) This spending effect increases the price of nontradable goods from P^0_N to P^1_N, shifting the equilibrium from point a to b. The spending effect generates inflation, and by increasing the price of nontradables relative to the price of tradables also causes the real exchange rate to appreciate. Inflation and real appreciation are the initial symptoms of Dutch disease.

Meeting the additional demand for nontradables requires that workers migrate from the lagging tradables sector into nontradables. Additional workers also migrate from the lagging tradables sector into the booming tradables sector. This outflow of workers from the lagging sector is shown in "Lagging Tradables" as a leftward shift in the supply curve from S_L to S'_L, increasing the quantity imported at the world price P_L by ΔM_L. Similarly, there may be a net outflow of workers from the nontradables sector into the booming sector, resulting in a leftward shift in the supply curve for nontradables, illustrated in "Nontradables" as a shift from S_N to S'_N. These shifts are called the "resource movement effect." This creates an excess demand for nontradables at equilibrium point b, causing still more domestic price inflation and more appreciation of the real


Booming Tradables
P
S_B
S'_B
P_B
D_B
Q_B
ΔX_B
Nontradables
P
D_N
D'_N
S'_N
S_N
c
P_N^2
P_N^1
b
P_N^0
a
Resource movement effect
Spending effect
Q_N



exchange rate as the nontradables market reaches a new equilibrium at point *c*, with the price of nontradables now increasing to P_N^2.

While the net impact of these effects on the quantity of nontradables is ambiguous, the quantity produced in the lagging sector falls. This is an important part of why the Dutch disease is considered an illness. Once workers leave that sector, they may never return, even after the boom ends. In practice, the lagging sector in many episodes of Dutch disease in developing countries is agriculture.

Source: Based on W. M. Corden, "Booming Sector and Dutch Disease Economics: Survey and Consolidation," *Oxford Economic Papers* 36, no. 3 (November 1984).

Because mineral and other primary sectors typically pay high taxes, commodity booms can lead to a swelling of government revenue. This *fiscal linkage* can be used to stimulate development, especially if the additional revenue is invested in public services, such as infrastructure, education, and health, or to promote efficient investment in tradable sectors, notably agriculture and manufacturing, that have been rendered less competitive by the primary export boom. Although some governments use their fiscal resources effectively during commodity booms, the record has often been dismal, with significant resources wasted on frivolous projects. Precisely

when fiscal resources are generously available, finance ministers have the most difficult time resisting political and social pressures for higher expenditures. Once again, the big difficulties arise when the boom ends, and government must quickly reduce spending and reverse the commitments made during the good times. Some countries borrowed heavily from abroad during the commodity booms of the 1970s on the assumption that export prices would remain high and were left with large foreign debts and a significantly reduced capacity to pay when the boom ended.

The depredations of Dutch disease and other problems with primary product exports often are fatal to development aspirations, as the case study of Nigeria illustrates (Box 18–2). But this does not have to happen. Some countries have turned their resource windfalls into sustained development, including Botswana, Chile, Indonesia, Malaysia, Thailand, and the United Arab Emirates. The negative relationship between primary product exports, commodity booms and economic growth is a tendency, not an absolute straightjacket. Determined governments can take several steps to avoid some of the most difficult problems associated with abundant natural resources and commodity price cycles.

BOX 18–2 NIGERIA: A BAD CASE OF DUTCH DISEASE

In 1973–74, the oil embargo imposed by Arab countries, followed by the activation of the Organization of the Petroleum Exporting Countries (OPEC) as an effective cartel, quadrupled the price of petroleum on world markets. In 1979–80, the price doubled again, so that by the end of 1980 the terms of trade for oil exports, relative to the price of imports, was nearly seven times the level in 1972. In Nigeria, higher export prices generated an oil windfall, which added 23 percent to nonmining gross domestic product (GDP) in the middle 1970s and again in the early 1980s. Looking back, Nigeria's oil revenues did little to promote economic development.

Nigeria's political history has been marked by intense competition among ethnic groups, including the Biafran war of the late 1960s. One of the major battlegrounds of this strife, which continues today, has been the incidence of taxation and government expenditure. Under this kind of pressure, the Nigerian government spent its oil windfall. Public investment rose from 4 to 30 percent of nonmining GDP, and the average pay for civil servants was doubled in 1975. Much of the newfound revenue was squandered on wasteful projects; some disappeared into the offshore bank accounts of powerful members of the government and military. The second oil windfall only whetted fiscal appetites even more: From 1981 to 1984, the budget deficit averaged 12 percent of nonmining GDP.

Fiscal excesses exacerbated the tendency of export windfalls to create inflation. Prices rose while the central bank kept the nominal exchange rate fixed, so that, by 1984, the real exchange rate had appreciated to nearly three times its level in 1970–72. Over the decade ending in 1984, Nigeria's non-oil exports fell almost 90 percent in nominal terms, a classic if extreme symptom of Dutch disease. Agriculture suffered worst. Because rural constituencies were politically weak, little of the oil windfall was invested in agriculture, while vast amounts were spent, and wasted, on infrastructure and industry. From 1973 to 1984, the quantity of agricultural exports fell by more than two thirds, while agriculture output per capita and total caloric consumption per capita both declined. From 1972 to 1981, growth in nonmining GDP was a respectable 5.3 percent, but this was only 60 percent of the growth rate during the five years before the oil price boom.

Oil prices declined during the 1980s and did not rise significantly again until the 2000s. Throughout this period, Nigeria oil revenues per capita kept growing even as Nigeria's population increased from almost 50 to 125 million people. But despite rising oil revenues, Nigeria experienced little economic growth and saw its poverty rate more than double and inequality soar. All these difficulties cannot be explained by Dutch disease alone, although during some time periods an appreciating real exchange rate contributed to the nation's woes. It can be argued that Nigeria might have been better off without oil.

Source: Henry Bienen, "Nigeria: From Windfall Gains to Welfare Losses?" (227–60), in Alan Gelb, ed., *Oil Windfalls: Blessing or Curse?* (New York: Oxford University Press for the World Bank, 1988); Frederick van der Ploeg, "Natural Resources: Curse or Blessing?" *Journal of Economic Literature* 49, no. 2 (June 2011).

The best prevention for Dutch disease effects is to avoid or reverse the initial real appreciation of the currency. In most cases, this requires a devaluation of the currency, accompanied by strong restraints on government spending and money creation by the central bank, both aimed at curbing inflation. The government needs to resist demands for expansion and save its newfound revenues until there is time to plan sensible, well-targeted projects with high returns. More-successful governments channel new investments into health, education, and infrastructure development that improve overall well-being, enhance productivity, and open up economic opportunities outside the primary goods sector. Such an investment policy accomplishes two things. First, it harnesses export windfalls to finance sound, long-term development. Second, by delaying the new expenditures, the government acts *countercyclically* and so helps stabilize the economy by spending less during the most inflationary period of the export boom and more after the boom has faded. Indonesia pursued some of these strategies in the 1970s and 1980s

and, facing the same increase in prices for oil exports as did Nigeria, had far better growth and development outcomes (Box 18–3). Early evidence on the commodity boom of the 2000s suggests that better macroeconomic management has been pursued by many primary product exporters as they attempt to minimize Dutch disease. On average, fiscal spending by resource dependent economies has been

BOX 18–3 INDONESIA: FINDING A CURE

At first, the oil boom affected Indonesia much as it did Nigeria. The 1973–74 boom added 16 percent to nonmining gross domestic product (GDP) and the 1979–80 price surge raised the windfall to 23 percent. From the first windfall, the government spent more than 60 percent. From 1974 to 1978, the real exchange rate appreciated an average 33 percent over its preboom level, a bit more than in Nigeria at that time. Yet the outcome was very different in Indonesia. Throughout the boom period, the government was required to balance its budget each year, and because all controls had been removed from foreign exchange transfers, stringent management of the money supply was necessary to protect foreign exchange reserves. These self-imposed restraints limited the impact of the windfalls on inflation.

The Indonesian government adopted two policies that took advantage of oil windfalls for national development goals. First, investment in agriculture had a high priority, especially the goal of achieving self-sufficiency in rice production. The government financed irrigation systems, encouraged the adoption of new rice varieties, subsidized fertilizer and pesticide sales, provided credit to farmers, invested in rural health and education facilities, and built roads and other infrastructure in rural areas. By 1982–83, Indonesia's total food output per capita was a third above its 1970 level—a performance far above average for all developing and industrial countries over the period. The second policy was to devalue the exchange rate enough to avoid real appreciation. Major devaluations were imposed in 1978, 1983, and 1986, after which the rate was managed flexibly to maintain its real value. At the end of the oil boom period in 1984, the Indonesian real exchange rate had *de*preciated 8 percent from its 1970–72 average. Consequently, over the period from 1971 to 1984, the quantity of non-oil exports grew by over 7 percent a year, and from 1972 to 1981, nonmining GDP expanded by over 8 percent a year.

Shrewd policy played a major role in effecting an early cure for incipient Dutch disease. The government also placed great stress on integrating the multitude of ethnic groups of this diverse country and avoided the conflicts over fiscal resources that paralyzed Nigeria. It undertook a series of reforms that laid the

basis for rapid growth. Import tariffs were substantially reduced, the banking sector was opened to foreign competition, and new regulations were set up to create an active stock market. These reform measures served to align Indonesia more closely to the global economy. The result was a period of rapid growth. Between 1988 and 1996 GDP, measured at constant U.S. dollars, grew at an annual rate of 6.8 percent, reducing poverty levels from the 1976 level of 54 million to 22.5 million by 1996. Indonesia, which at one time had been considered a laggard among the East Asian economies, was now considered a development miracle. Part of its success was due to policies that avoided succumbing to Dutch disease.

Source: Bruce Glassburner, "Indonesia: Windfalls in a Poor Rural Economy," in Alan Gelb, ed., *Oil Windfalls: Blessing or Curse?* (New York: Oxford University Press for the World Bank, 1988).

more prudent and the RERs of these countries have appreciated less than in past commodity booms.[23]

Still, primary commodity exporters tend to grow more slowly than more diversified economies. Policies of restraint, essential to stabilization, are seldom popular and often vigorously opposed by political pressure groups. Although the judicious use of commodity revenues is easy to prescribe and essential for an economy's health, the advice often is not taken in sufficient doses to affect a cure. The relationship between primary commodities and governance is considered next.

THE RESOURCE TRAP

Indonesia and Nigeria are both populous countries that export oil. Oil represents both an opportunity and a challenge to these economies. As described in Boxes 18–2 and 18–3, Indonesia managed earlier resource booms well and prospered; Nigeria did not. The same can be observed when comparing Botswana and Sierra Leone, two countries that export diamonds or Chile and Zambia, two copper exporters. The populations of Botswana, Chile, and Indonesia have experienced economic development; those of Nigeria, Sierra Leone and Zambia have not. Clearly, governance and politics played a large role in these different outcomes. Managing Dutch disease and making productive use of export revenues was not simply a technical matter but one of political will and effective governance.

[23]World Bank, "Dealing with Changing Commodity Prices," in World Bank, *Global Economic Prospects* (Washington, DC: World Bank, 2009), discusses recent trends. It notes that countries experiencing a commodity boom for the first time, often because of newly discovered resources (Chad, Equatorial Guinea, Sudan, Yemen) or recent statehood (Azerbaijan and Kazakhstan), are showing the same kind of macroeconomic volatility that characterized more established producers during earlier booms. Those developing economies with prior experience with commodity price cycles appear to be better managing the macroeconomic consequences of the current boom.

To understand the relationship between primary product exports and governance, it is useful to turn to the work of British economist Paul Collier, who has written extensively on commodity exports and economic development. Some of his early work was on the consequences of the coffee price boom of the late 1970s on Kenya's economic development. In 2007, Collier published *The Bottom Billion*, the title of which refers to the populations of 58 countries, mostly in Africa, whose people are "living and dying in fourteenth century conditions." What distinguishes these economies from other developing nations is their failure to grow over the past three to four decades. Some have experienced long-term economic stagnation; others have seen per capita incomes decline. Collier argues that these economies have fallen into one of several **traps** that have prevented them from growing. Dependence on natural resource exports is directly implicated in his analysis.

Collier specifically identifies a natural resource trap, which incorporates problems we already have discussed. Dutch disease and the appreciation of the RER can crowd out nonprimary product exports with more long-term growth potential. Price volatility during the boom often leads to public expenditures wasted on white elephant projects, expansions in civil service employment motivated more by political patronage than service delivery, or outright theft by corrupt officials. Commodity booms often are also associated with increased foreign borrowing and the accumulation of debt. If borrowed funds are spent wisely, this is a good strategy. If the borrowed funds are misspent, then during the inevitable and subsequent bust in commodity prices and government revenues, the economy faces a debt burden and must service its debt often by cutting essential development expenditures. Underlying these problems of commodity price cycles and macroeconomic mismanagement is poor governance, which according to Collier and others, is tied to the resource revenues themselves.

Simply stated, resource revenues worsen governance. Evidence for this is provided by the World Bank, which has developed indexes of corruption. Oil and mineral exporters, on average, have more corruption than agricultural exporters, who in turn are more corrupt than diversified exporters (the latter include economies that export a range of fuel, non-fuel primary commodities and manufactures).[24] Corruption then acts as a tax, often a large one, on economic activity. It encourages rent-seeking behavior rather than productive investment and growth suffers accordingly. Collier closes the circle among dependence on resource revenues, widespread corruption and poor governance, and a lack of economic growth by explaining why resource revenues breed corruption. Revenue from oil, diamonds, minerals, and certain other commodities reduce the need to tax citizens.[25] The government secures revenues by nationalizing the resource industry or by making leasing arrangements

[24]World Bank, "Dealing with Changing Commodity Prices," p. 109.

[25]Not all primary commodities are equally vulnerable to these problems. The more concentrated production and revenues, the easier it is for the state to appropriate resource revenues. This is what makes oil, diamonds, gold, silver, uranium and other minerals, and some plantation crops different from cotton, rice, fish, and livestock exports.

with multinational extractors. Either way, the revenues bypass the population in a way that ordinary tax collection does not. If citizens are not directly taxed they may be less likely to hold public officials accountable.

When resource rents are available, political patronage often becomes the means of political competition, far more effective in garnering votes or remaining in power than by improving the provision of public services or fostering national development. Collier argues that resource revenues can pollute governance in autocratic regimes and democracies alike, and can be inimical to development in both low-income countries (Congo, Sierra Leone) and middle-income nations (Russia, Venezuela.) He worries that when resource revenues worsen governance, the trap is hard to break free from.

In addition, Collier describes a **conflict trap** that ensnares the bottom billion. Civil wars have been endemic, particularly in sub-Saharan Africa. Collier argues that the existence of mineral wealth significantly increases the likelihood of civil war, as greed motivates many rebel groups and mineral wealth itself helps finance conflict. Competition over the control of valuable natural resources for export certainly played (and in some instances continues to play) a role in the civil wars in Angola, Congo, Sierra Leone, and Sudan. The expression *blood diamonds*, stems from some of these conflicts. In Fiji, Collier argues, control over the valuable mahogany and other tropical wood concessions, was central to the armed insurgency that took place in 2000 and temporarily overthrew the elected government. According to Collier, low income, slow growth, and dependence on primary commodity exports significantly increase the probability that a nation will experience a civil war.

BREAKING THE RESOURCE CURSE

Collier paints a bleak picture, far bleaker even than that of Prebisch and Singer, of the prospects for the resource-rich nations among the bottom billion. His critics agree that correlations between resource abundance and poor development outcomes are relatively robust, but find far less statistical support for the causal channels Collier identifies.[26] And even if a resource trap exists, there are examples of nations that have escaped. If Botswana leveraged diamond revenues into rapid growth, so can Angola. If Indonesia, despite its own pervasive corruption, managed its oil revenues well, then so can Nigeria. The resource curse, if one exists, can be overcome. Some recent approaches include the use of sovereign wealth funds and wider adoption of the Extractive Industries Transparency Initiative (EITI) and similar international charters.

Sovereign wealth funds refer to the accumulation of foreign assets, usually financed by primary product export earnings. These funds are distinct from the foreign reserve holdings of central banks that are intended to help with short-term macroeconomic stabilization. By keeping resource earnings in assets denominated in foreign currencies, RER

[26]See William Easterly's review of *The Bottom Billion*, "An Ivory Tower Analysis of Real World Poverty," *The Lancet* 307 (October 27, 2007) 1475–76.

appreciation and Dutch disease can be reduced. The funds also can be used countercyclically and smooth the expenditure flows that follow from commodity price cycles and can provide a national nest egg for when non-renewable commodities are depleted. But they are not a substitute for good governance. Sovereign wealth funds can be mismanaged or abused just like any other source of government revenue. Venezuela, for example, has borrowed heavily against its sovereign wealth fund minimizing its potential benefits.

EITI, launched in 2002 by the British government, is an attempt to impose some accountability on both international companies engaged in oil and mining and on governments who receive payments for these natural resources. Without an international charter like EITI, negotiations for drilling or mining rights are often held behind closed doors and beyond public scrutiny. As Collier puts it, such deals often are good for the companies and for the minister doing the negotiations but not for the country.[27] EITI is an international coalition of governments, companies, multilateral organizations, and civil society that creates a global standard for the verification and publication of both company payments and government revenues obtained from oil, gas, and mining. It requires firms in extractive industries to disclose payments for oil and mineral rights, governments to disclose receipt of payments, and independent international audits of these transfers. The hope is that greater transparency will reduce corruption and make governments more accountable in their use of resource revenues. In 2009, oil-rich Azerbaijan became the first country validated as compliant with all EITI requirements. More than two dozen other low- and middle-income nations are following Azerbaijan's lead and 40 of the world's largest oil, gas and mining companies are participating in the EITI process.

EITI is one example of an international charter Collier recommends as a means of improving accountability and governance among the bottom billion. He also recommends the use of verifiable international auctions for granting mineral rights. This would complement EITI and further reduce the back-room deals that govern the award of many such contracts. On the expenditure side, there also needs to be transparency in how government revenues are spent. In Uganda, the ministry of finance reported how much money was released to a given locality. Local media outlets were informed and individual schools were sent a poster informing them of what they should expect to receive. One study found that these steps increased the flow of resources to their intended recipients from only 20 to 80 percent of the initial allocation.[28] Collier believes that a package of international standards and charters can

[27]From, "Paul Collier on *The Bottom Billion*," TED, May 2008, available at www.ted.com/talks/paul_collier_shares_4_ways_to_help_the_bottom_billion.html.

[28]Ritva Reinikka and Jakob Svensson, "The Power of Information: Evidence from a Newspaper Campaign to Reduce Capture," World Bank Policy Research Working Paper No. 3239, Washington, DC, March 2004. See also Paul Hubbard, "Putting the Power of Transparency in Context: Information's Role in Reducing Corruption in Uganda's Education Sector," Working Paper Number 136, Center for Global Development, Washington, DC, December 2007, which argues that the improved flow of resources to intended beneficiaries was the product of many reforms in government practice, including but not solely the result of, more transparent information.

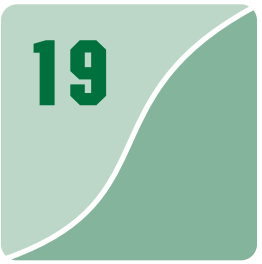
impose some checks and balances on bad governance and help transform natural resource exports from a curse to a means for promoting economic development.

David Ricardo would probably recognize some contemporary debates about trade between nations. He would still argue on behalf of the potential benefits of trade for improving aggregate output and welfare. The gains from trade are as real today as they were in the early nineteenth century. But trade is not a panacea. As the discussion of primary product exports demonstrates, economic growth and development entails far more than just the pursuit of comparative advantage.

SUMMARY

- World trade has grown rapidly in recent decades. Trade between high-income economies still dominates world commerce, with developing nations (including China) accounting for about one quarter of the total. With imports and exports from low- and middle-income economies growing more quickly than their GDPs, trade has contributed to a rise in globalization and to the importance of trade for most nations.
- The share of trade in GDP tends to be higher the smaller the economy and the better its access to seaports and major world markets. Landlocked nations, of which there are 44, tend to trade less and grow more slowly than nations with coastlines. As economies grow, the composition of trade tends to change, with exports shifting from primary products toward manufactured goods.
- Comparative advantage, an idea developed by David Ricardo, allows two countries to gain from trade with each other, even when factor inputs are more productive for all products in one country than the other. By pursuing comparative advantage, countries experience gains from exchange and gains from specialization, and are able to raise national income.
- Beyond these gains, trade encourages competition and reduces the market power of domestic firms, facilitates the adoption of new technologies, can spur investment, and increases the quality and variety of goods available. But trade is not without costs; there are both winners and losers. In the short run, firms and their workers in the export sector see their business expand and in the import-competing sector contract. Trade tends to benefit a nation's abundant factors of production with the gains of the winners surpassing the losses of the losers.
- Most world trade involves manufactured goods, but primary products continue to dominate the exports of many low- and middle-income economies. Unfortunately, many of these economies have suffered from a resource curse, by which a dependence on oil and other primary good exports often has been associated with economic stagnation and even decline.

- Pessimism about primary product exports and development has a long history. In the 1950s, economists such as Prebisch and Singer argued that the terms of trade steadily and systematically would turn against primary products. This, they argued, was because of structural features in the global economy, including that demand for these exports typically grow more slowly than for manufactures.
- Today, the arguments of Prebisch and Singer no longer are seen as explaining the resource curse. More often, the problem is not steadily declining terms of trade but the cycles of price booms and busts in some commodities. This can lead to the phenomenon known as Dutch disease, by which a country rich in natural resources suffers slower growth. Dutch disease most often is associated with an appreciation of a nation's real exchange rate, which harms growth in other sectors, especially in nonprimary good exports.
- Beyond Dutch disease, lies an even more troubling explanation for the resource curse. Resource revenues tend to worsen governance in a way that export earnings from manufactured goods do not. In weak states lacking good institutions, resource rents can perpetuate civil wars, fuel endemic corruption, and undermine the necessary checks and balances required for good governance.
- Nations can and have overcome the resource curse by pursuing prudent macroeconomic policies and by taking the necessary steps to prevent waste and corruption in the use of primary export revenues. International charters like the EITI are a mechanism for supporting the improved governance of oil and other resource exporters.



Trade Policy

In 2009, the Pew Research Center, conducted a survey of global attitudes. It sampled populations in 55 countries. One question asked, "What do you think about the growing trade and business ties between your country and other countries—do you think it is a very good thing, somewhat good, somewhat bad or a very bad thing for our country?" Over 2,000 people were surveyed in India and 96 percent replied that trade was "very good" or "somewhat good." In China and South Korea, 93 and 92 percent, respectively, felt the same way. The percentages in Nigeria and Brazil were 90 and 87. Somewhat lower approval ratings of international trade were voiced in France (83 percent), Russia (80 percent), Japan (73 percent), and the United States (65 percent). In every country surveyed, a majority of respondents thought trade was "very" or "somewhat good."

Given the benefits of international trade discussed in the previous chapter, it is not surprising that trade is positively viewed by the general population. It can raise national income, result in lower prices for many goods, and increase product quality and variety. In some nations, exports are perceived of as the engine of economic growth. But a favorable view toward trade is not the same as support for free trade. Trade may be viewed favorably at the same time that workers and firms call for protection from imports. Politicians in rich and poor countries often find it in their interest to support trade protection. In many nations, policies that favor exports over imports are pursued, a strategy reminiscent of the mercantilism that Adam Smith and David Ricardo argued against hundreds of years ago.

World trade may today be more open than at any time in the past 50 or even 100 years, but the trade strategies that nations adopt continue to mix openness with

elements of protection or other forms of support for individual firms and sectors. The ideal trade strategy to promote economic growth and development remains a subject of continued debate, influenced by both economic theory and real world experience.

There exists a continuum of trade strategies a nation might pursue. At one end is economic isolationism, sometimes referred to as autarky, by which a nation elects to be self-sufficient. Japan pursued something close to this strategy before its opening to the West in the 1850s after the arrival of Commodore Matthew Perry and the black ships of the U.S. Navy. More recent examples include Albania (1976–91) after the principles set forth by Communist leader Enver Hoxha, and Burma (1962–88) under the ideology of military ruler Ne Win. Today, North Korea comes closest to pursuing a strategy of autarky. Its official state ideology, *juche*, includes economic self-sufficiency as a key element. But even North Korea engages in some limited international trade, exporting minerals, animal products and, allegedly, illicit goods such as missiles, narcotics and counterfeit cigarettes to secure critical imports of fuel and food, as well as luxuries for the nation's elite.

At the other end of the continuum from autarky is free trade. Free trade refers to the unfettered movement of goods and services across national borders. It implies that goods move as freely between nations as they do within nations. Once again, there are few examples. Hong Kong, especially before its return to Chinese sovereignty in 1997, is cited as having been one of the world's most open economies with few barriers to trade or support of one industry over another. Today, free trade characterizes the internal trade among members of the European Union (EU). But trade between EU members and the rest of the world is subject to numerous barriers, including tariffs, quantitative restrictions, and subsidies. Free trade agreements, such as the North American Free Trade Agreement (NAFTA) between Canada, Mexico, and the United States, often include restrictions on the exchange of some goods between member nations. It took 15 years for NAFTA to eliminate restrictions on trade in corn, sugar, and other commodities. U.S. anti-dumping duties have been applied to cement from Mexico and softwood lumber from Canada, whereas Mexico has imposed similar duties on high fructose corn syrup from the United States. Trade between NAFTA members in these and other commodities has been less free than internal trade within each nation's borders.

If autarky and free trade are rare and at opposite ends of the continuum of trade strategies, then most nations must adopt trade strategies that fall somewhere in between. The policy debate centers on how inward looking or how outward oriented an economy should be. The pendulum of which approach to follow has swung back and forth. The historical development of all advanced economies, whether Britain, Japan, or the United States, included periods of trade protection followed by a more open approach. A relaxation of restraints on trade characterized the early twentieth century only to be reversed after World War I and the Great Depression. The infamous Smoot-Hawley tariffs, passed into law in 1930 by the U.S. Congress, set off a round of retaliatory tariffs in other countries that contributed to a dramatic reduction in global

commerce and a prolonged decline in world output. After World War II, many nations signed the 1947 Global Agreement on Tariffs and Trade (GATT), resulting in a reduction in tariffs and other trade barriers and an expansion in international trade.

In the developing world, Latin America's primary export markets were severely disrupted by the Great Depression and by the scarcity of commercial shipping during World War II. Having built up domestic manufacturing capacity during this period, Argentina, Brazil, Colombia, Mexico, and other countries systematically erected barriers to keep out competing imports after the war. In Asia and Africa, most countries adopted inward-looking strategies in the 1950s and 1960s, as newly independent states wanted to develop their own industrial capacity and distance themselves from their former colonial rulers. By the 1960s, import substitution was the dominant strategy of economic development.[1] But this strategy did not live up to its promise and by the 1980s had generally been abandoned in favor of more outward-looking approaches. The early export-oriented emerging economies of East Asia (Hong Kong, Korea, Singapore, and Taiwan) had demonstrated the advantages of an export orientation.

Broadly speaking, governments in developing countries have employed two different trade strategies, import substitution and export orientation. Import substitution (IS) is the production of goods and services that replace (or substitute for) imports. Because new firms in developing countries often are unable, at least initially, to compete on world markets, import substitution protects domestic firms from international competition by erecting trade barriers that make imported products more expensive or more difficult to purchase, with the aim that over time firms will become more efficient and competitive. **Export orientation** shifts the focus to producing for global markets. This strategy is designed to make producers internationally competitive by relying on market forces; strengthening key institutions; and, in some cases, using subsidies, managed exchange rates, and other policy instruments. The core idea behind import substitution is that newly developing industries cannot survive at first without some protection from imports, giving them the chance to learn and grow. In contrast, the core idea behind an outward-looking trade strategy is for firms to compete internationally to gain access to new technologies, increase efficiency, and enlarge the scope of their potential market. We consider each strategy in turn.

IMPORT SUBSTITUTION

The basic idea of IS is straightforward. For sustained economic development, countries need to shift from primary production to manufacturers to prevent prolonged specialization in low-productivity activities. Some early proponents saw IS as both

[1]Recall the arguments of Raul Prebisch and Hans Singer discussed in the previous chapter. Their concerns and explanations for declining terms of trade facing developing nations provided some of the intellectual support for the import substitution trade strategies that many governments pursued.

a trade strategy and as a way to structurally transform an economy and referred to it as import substitution industrialization (ISI). Proponents of IS argue that firms are unlikely to be able to compete in manufactures immediately and require government assistance to get started. Ideally, the first step is to identify products with large domestic markets, as indicated by substantial imports, and relatively simple production technologies that can be mastered quickly, rather than products requiring advanced machinery and highly skilled labor. Governments then introduce either tariffs or quotas to increase the price of competing imports. **Tariffs** are taxes imposed on imports at the border; **quotas** are quantitative limits on specified categories of imports. These protective barriers by raising the price of imports, reduce the quantity of imports in the domestic market, and allow domestic manufacturers to charge higher prices, compensating them for higher production costs and making their operations profitable. The strategy comes at a cost to consumers and other firms, who must pay the higher prices with commensurate welfare and profit losses as a result.

This stylized approach (focus on relatively simple products with large domestic markets) implies that consumer goods, such as processed foods, beverages, textiles, clothing, and footwear, should be the first targets. By contrast, capital goods should not be heavily protected because they require more sophisticated skills and raising their costs hurts all downstream industries and investors that buy capital goods. Nevertheless, many developing countries have used IS to protect steel, machinery, and other capital goods, usually with little success.

Today, most economists are critical of the IS strategy, but there are valid arguments in favor of this approach, the most compelling of which is the concept of an **infant industry**. Entrepreneurs opening new production facilities in a developing country must compete against firms from industrialized countries that have long experience and have mastered both production technologies and marketing. The managers and workers of the new, or infant, industry must learn to use these technologies efficiently to compete. This process of learning by doing can take several years. Advocates of IS argue that without some form of assistance these investments are unlikely to take place and developing countries will be unable to learn the skills needed eventually to compete with imports on equal footing. For this strategy to work, an infant industry must "grow up" to be capable of eventually competing without protection against imports. This suggests that tariffs should be *temporary* and decline toward zero over time, as productivity increases and production costs fall. For this strategy to be economically worthwhile, the eventual (and discounted) benefits to society of establishing the new industry should exceed the current costs to the economy of protection.

To more fully understand IS and trade strategies in general, we first explore the impact of tariffs, quotas, and subsidies on imports, production, and consumption. We then examine the relationship between exchange rate policy and trade, before returning to an overall evaluation of IS.

PROTECTIVE TARIFFS

The most direct effect of a protective tariff, whether it is imposed by a developing or a developed economy, is to raise the domestic price of the good above the world price. For the importing country, the world price is the cost at the port of entry, usually called the *c.i.f. price* (including cost, insurance, and freight) or *border price*. If we assume the country is small relative to the world market and that the imported good is a perfect substitute for the good produced domestically, then the country faces a world supply curve that is perfectly elastic at the world price. We also assume that the domestic market is competitive, permitting us to represent the domestic market using standard supply and demand curves. In Figure 19–1, at the world or border price, P_w, consumers demand Q_1 and local producers produce Q_2; the balance, $M_1 = Q_1 - Q_2$, is imported. If an **ad valorem** (that is, percentage) tariff, t_0, is placed on the good, the percentage increase in the domestic price, $t_0 = (P_d - P_w)/P_w$, is referred to as the **nominal rate of protection**.[2] The tariff reduces quantity demanded to Q_3 and increases domestic production to Q_4, the latter being the goal of import protection. Imports are reduced to $M_2 = Q_3 - Q_4$.

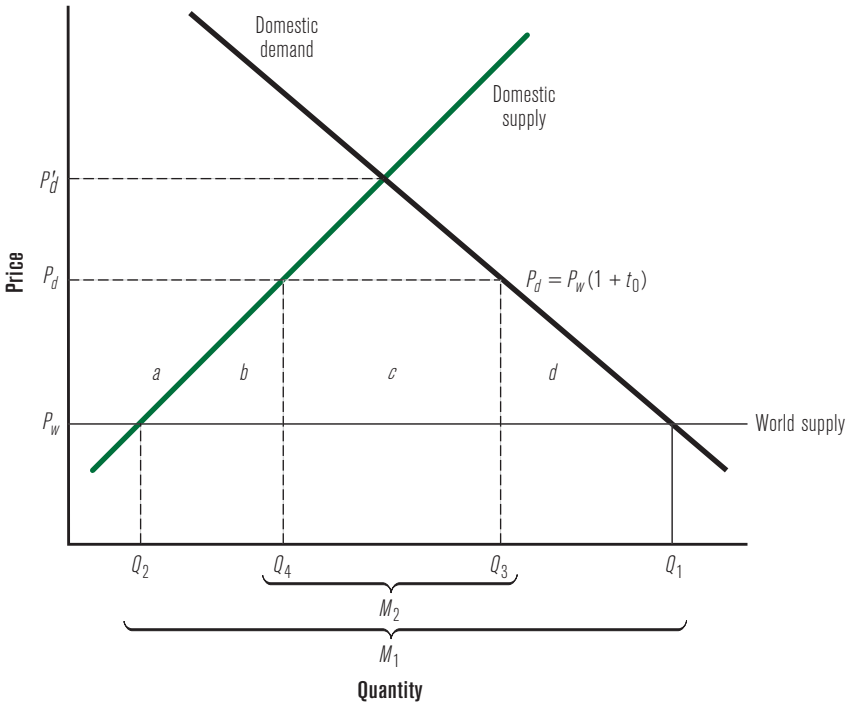


FIGURE 19–1 **Nominal Tariff Protection**

[2]Tariffs sometimes are applied not in percentage terms but as specific duties—for example, instead of, say, 25 percent of the c.i.f. price, 15 pesos or 75 rupees are added to the c.i.f. price. When a specific duty, t_d, is applied, the nominal rate of protection is t_d/P_w.

The increase in domestic output from Q_2 to Q_4 has two effects. First, it increases the *producers' surplus* by an amount given by trapezoid *a*, which captures the extent to which producers receive a price, P_d, that exceeds their marginal cost of production, represented by the domestic supply curve. Second, the increase in production entails a *resource cost*, given by triangle *b*, because it uses labor, machinery, and raw materials that could be used more efficiently to produce something else. The resource cost can best be understood by thinking about the opportunity cost of increased domestic production. Without the tariff, the nation can import an amount equal to Q_2 to Q_4 by paying the world price and costing, $P_w(Q_4 - Q_2)$. To produce this added amount at home costs more. The cost equals the additional cost of all these units, which is the area under the supply curve from Q_2 to Q_4; an amount equal to: $P_w(Q_4 - Q_2) + b$. Because it costs more to produce at home than to import, the society is using its resources inefficiently. Society could have used its resources to import the same quantity of the good, Q_2 to Q_4, and applied the extra resources represented by triangle *b*, to produce and consume something else.

Consumers pay for protection both by the higher price, P_d, they pay on all purchases and by reducing consumption from Q_1 to Q_3. Part of what they pay is transferred to the government as tariff revenue that can be used by society for other purposes. This amount is a transfer within society rather than a loss. Tariff revenue is equal to the new quantity of imports, M_2, multiplied by the increase in the import price resulting from the tariff, represented by rectangle $c = (t_0 P_w) \times M_2$. The total cost to consumers is represented by the loss of *consumer surplus*, the difference between consumer demand and the market price, equal to area $a + b + c + d$. The loss to consumers is only partially compensated by the gains to either producers (trapezoid *a*) or the general public (through the tariff revenues represented by rectangle *c*). Area $b + d$, two triangles representing welfare losses not compensated by gains to anyone, called the *deadweight loss*, represents the efficiency loss to the economy.

We already identified area *b* as a loss resulting from the inefficiency of producing the good at home rather than importing it; area *d* is the inefficiency resulting from higher prices that reduce the choices available to consumers, sometimes referred to as the *consumption cost* of the tariff. Both these deadweight losses increase as the tariff rate rises. The maximum losses occur under a prohibitive tariff, which drives the world price all the way to the autarky price, P'_d, and imports to zero. The protective effects of tariffs are really more complex than suggested by Figure 19–1 because the diagram focuses only on output markets. Much depends on the relationship between tariffs on both the final good and any imported inputs needed to produce that good (Box 19–1).

IMPORT QUOTAS

Industries also can be protected through **quantitative restrictions** on imports, usually in the form of quotas. With quotas, the government determines in advance the quantity of imports it wants to allow, whereas with tariffs, the quantity of imports depends

BOX 19-1 EFFECTIVE RATES OF PROTECTION

How much do tariffs affect producer's incentives to shift production? The tariff rate on competing imports gives us a clue but does not tell the whole story. The tariff rate determines the nominal rate of protection, which focuses on the effect of tariffs on *output* prices. But producers also care about the effect of tariffs on *input* prices and, ultimately, on the margin between revenues and input costs. Whereas a tariff on competing imports helps raise profits, tariffs on intermediate goods reduces them.

The **effective rate of protection (ERP)** measures the protection afforded by the entire structure of tariffs on inputs and outputs.[a] It focuses attention on the impact of trade policies on *value added*, the difference between the selling price of the good and the unit cost of intermediate goods.[b] Value added can be measured in domestic prices as the difference between the domestic price of the product (P_d) and the domestic cost of material inputs per unit of output (C_d), while the value added at world prices is the difference between the world price of the product (P_w) and the costs of inputs valued at world prices (C_w). The ERP is the ratio between the two:

$$\text{ERP} = \frac{\text{Value added (domestic prices)}}{\text{Value added (world prices)}} - 1 = \frac{P_d - C_d}{P_w - C_w} - 1$$

The key difference between domestic prices and world prices is tariffs (and other trade barriers such as quotas that can have similar effects on prices). The tariff on competing imports (t_0) raises the domestic price of the product above the world price, $P_d = P_w(1 + t_0)$, while the average tariff on inputs (t_i) raises the costs of producing it, $C_d = C_w(1 + t_i)$. Thus domestic prices and costs can be expressed in terms of world prices and tariffs:

$$\text{ERP} = \frac{P_w(1 + t_0) - C_w(1 + t_i)}{P_w - C_w} - 1 \text{ or } \text{ERP} = \frac{P_w t_0 - C_w t_i}{P_w - C_w}$$

[a]The concept of the effective rate of protection is due to Max Corden, "The Structure of a Tariff System and the Effective Protection Rate," *Journal of Political Economy* 74 (June 1966), 221–37.

[b]Within this margin, the manufacturer must pay all the factors of production: wages, rents, interest on borrowed capital, and profit. Together these payments are equal to the value added.

Because the denominator is the value added calculated at world prices, the ERP measures the impact of all tariffs on the value added, rather than on prices. The measurement of ERP easily can be extended to include quotas, subsides, or other policies that influence output and input prices and thus the value added. To see the powerful insights from the ERP concept, suppose a producer manufactures a product that sells for $100 and costs $60 to produce on world markets, so that value added is $40. Consider the impact on value added of three stylistic cases of tariff policy:

- *The tariff on output is exactly the same as the tariff on inputs*. Say both t_0 and t_i equal 10 percent. You might think there would be no impact, but there is one. The sales price rises to $110 and domestic costs increase to $66, so value added increases by 10 percent to $44. A uniform tariff, 10 percent in this case, provides effective protection on value added of 10 percent, exactly equal to the nominal rate of protection.
- *The tariff on output is higher than the tariff on inputs*. If $t_0 = 10$ percent but $t_i = 0$, the domestic price rises to $110 and the valued added rises from $40 to $50. Thus the ERP $= 25$ percent, much higher than the nominal rate of protection of 10 percent. More generally, if t_0 is greater than t_i, the ERP exceeds the nominal rate of protection. Note the key point here: What seems like a small tariff can have a huge impact on the margin between costs and prices and thus a big impact on producers' incentives to reallocate resources away from sectors with no protection to those with protection. Tariff structures in many countries tend to follow this pattern, escalating from relatively low rates on inputs to higher rates on finished products. Under these circumstances, ERPs of 100 percent are not uncommon.
- *The tariff on output is lower than the tariff on inputs*. If there is no tariff on competing imports but a 10 percent tariff on inputs ($t_i = 10$ percent), domestic costs rise to $66 but output prices remain at $100. Thus value added falls to $34 and the ERP is a *negative* 15 percent. Although a tariff structure that undermines investment incentives seems unlikely, it is actually fairly common, particularly for agriculture and export products. Governments like to keep food prices low to keep urban consumers happy and so are reluctant to impose duties on food imports. Therefore, protective tariffs on fertilizer, seeds, or irrigation equipment undermine the profitability of agriculture and switch investment out of agriculture. As for exports, governments cannot use tariffs to raise output prices because exporters sell on world markets. A tariff on inputs raises exporters' costs and reduces incentives to invest in export industries. This pattern of protection is a critical problem undermining the international competitiveness of exporting firms in many developing countries.

on the reactions of producers and consumers to higher prices, as captured by the elasticities of supply and demand. A quota that limits imports to the same quantity as a tariff has many of the same effects. In Figure 19–1, a quota limiting imports to M_2 forces the domestic price up to P_d, increases domestic production to Q_4, and decreases consumption to Q_3, just as with the tariff. The quota permits M_2 units of imports at every possible price, in essence, shifting the domestic supply to the right by an amount, M_2. Equilibrium in the market occurs at the intersection of domestic demand and domestic supply augmented by the quota. In Figure 19–1 this occurs at price P_d and quantity Q_3, identical to the effects of the tariff t_0. The loss in consumer surplus, the deadweight losses, and the gain in producer surplus are also the same for quota, M_2, as for ad valorem tariff, t_0.

But in two critical respects import quotas have different effects from tariffs. First, the government does not necessarily collect revenue equal to area *c*. To enforce the quota it must issue licenses to a limited number of importers, giving them the right to purchase up to M_2 of imports. If the government simply gives away the licenses with no fee (which many governments do, using the licenses as patronage for politically well-connected importers), license holders earn a windfall profit equal to the area of rectangle *c*. This is because the importer can purchase the goods at price P_w on the world market and sell them at P_d domestically. This windfall, often called a **quota rent**, can be substantial, so importers expend a great deal of effort to obtain them, including possibly offering large bribes, in effect sharing the quota rent with the right government official. (Under a tariff, importers can also bribe customs officials to avoid collection of all of the required tariffs.) Alternatively, and preferably, the government could sell the import licenses at auction. Assume the government decides to auction 1,000 import licenses, each license entitling the holder to import M_2/1,000 units of the good. Potential importers would be willing to pay up to an amount *c*/1,000 for each of these licenses because that is the amount of the potential revenue windfall they could earn. A well-functioning auction could yield the government the same revenues as a tariff, although the distribution of quotas often is seen as more costly to administer than tariffs.[3]

Second, market dynamics that shift the domestic supply and demand curves or the world price have very different impacts with tariffs and quotas. Consider a fall in the world price P_w. With a tariff, both P_w and P_d fall, giving consumers the benefit of the lower price. Consumption increases, domestic production declines, and imports rise to fill the gap. Under a quota, however, imports cannot rise unless the quota is increased. Domestic production and consumption remain unchanged, and the domestic price remains at P_d. The lower world price simply increases the quota rent, with the benefit going directly to import license holders. Put more generally, under

[3]Distributing quota rents can be complex. The U.S. government uses quotas as both an instrument of protection for the U.S. sugar industry and as a tool of foreign policy. The United States has protected its domestic sugar industry since the 1930s, resulting most years in much higher prices for sugar in the domestic market than in the world market. U.S. quotas are allocated to sugar-exporting nations, with Brazil, Dominican Republic, and Philippines receiving the largest amounts. The governments of these nations then decide how to allocate the valuable quota rents. In essence, the U.S. government, through its use of quotas, takes some of the money U.S. consumers pay for higher-priced sugar and gives that money to other nations.

most (but not all) circumstances because tariffs allow changes in imports on the margin, domestic producers and consumers react on the margin to market changes as they would in an open economy; with quotas, they react on the margin as they would in a closed economy. Because of these consequences of quotas, economists prefer tariffs, and trade reforms and international trade agreements often start with the elimination of quotas or at least conversion to their equivalent tariffs.

TRADE PROTECTION AND POLITICS

From the partial equilibrium model of a tariff (or quota) developed in Figure 19–1, it seems almost illogical for trade protection ever to be used. Consumers lose consumer surplus equal to $a + b + c + d$ because of the tariff but producers and government gain only $a + c$. A general equilibrium approach reaches similar conclusions. (Box 19–2). If the costs of protection outweigh the benefits, why would a government employ such measures?

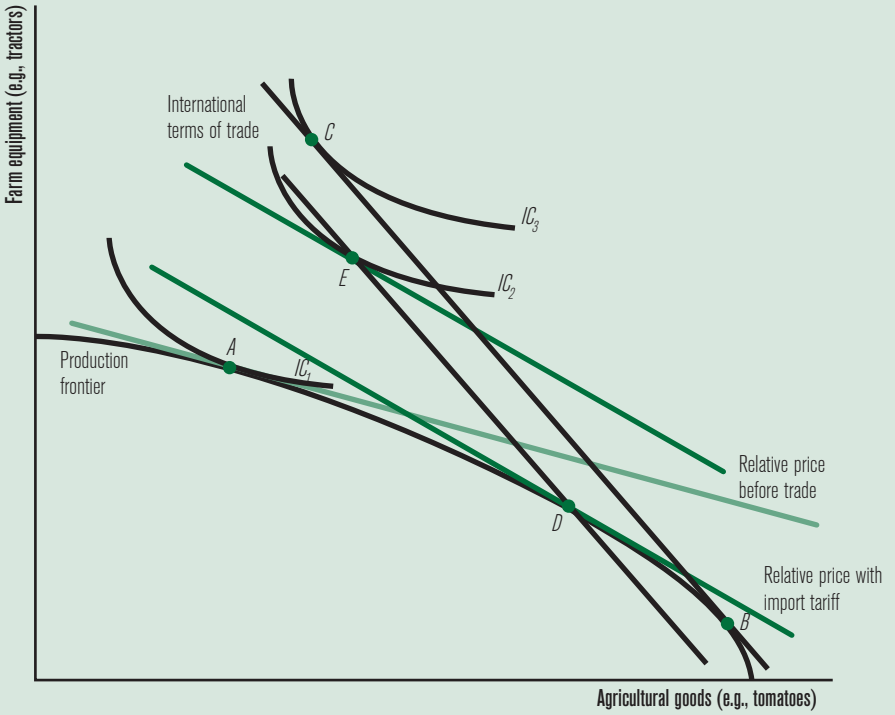
The infant industry argument acknowledges that domestic costs of production, captured by the domestic supply curve, are higher than production costs elsewhere. That is precisely why trade protection is needed. It enables the new industry to gain experience, bringing costs down, causing the domestic supply curve to fall, and eventually making the domestic industry competitive. Short-run costs, according to this line of reasoning, are worth paying for expected productivity growth and any positive spillover effects once the infant industry matures.

Another and in some ways better explanation for the widespread use and persistence of trade protection involves its distributional consequences and the incentives such protection creates. Consider Thailand's 60 percent tariff on imported beer. With the tariff, domestic producers can sell more beer and can sell it at a higher price. The tariff keeps lower-priced beer from Laos and Vietnam out of Thailand. Domestic breweries earn more revenue and greater producer surplus. They employ more workers because the tariff expands production beyond the level comparative advantage warrants. But consumers lose. They face fewer alternatives and pay a higher price for beer. In the aggregate, according to our models, consumers lose more than producers gain. To explain the persistence of the tariff, what may really matter is the extent of *individual* gains and losses.

With the tariff, individual consumers have to pay a few more baht (the Thai currency) for a bottle of beer. They would prefer to pay less, but the amounts involved for consumers are small relative to their total expenditures. It is hard to imagine mobilizing the public to protest over the beer tariff. For Thai beer producers the incentives are different. They are relatively few in number and the potential gains from a tariff are large. Millions of Thai beer drinkers each pay a bit more for beer, but these losses when aggregated become the gains shared by only a few firms. It is easy to imagine these firms mobilizing to seek political support or the favor of influential government

BOX 19-2 THE TWO-COUNTRY MODEL WITH A TARIFF

We can also examine the impact of a tariff using the two-country, two-good, two-factor model developed in Chapter 18. Mexico is again trading tomatoes and tractors with the United States (or with the rest of the world.) Mexico has comparative advantage in tomatoes and imports tractors. If Mexico engages in no trade, it consumes and produces at *A*; with free trade it produces at *B* and consumes at *C*. Trade increases utility from the level at IC_1 to that at IC_3.



Assume Mexico embarks on an import substitution strategy and decides to protect its tractor industry with a tariff, t_0. The tariff changes the slope of the price line facing domestic producers. Before the tariff, the slope is, Px/Py; after the tariff it is, $Px/[Py(1 + t_0)]$. This causes the price line to rotate and become flatter. Because of the tariff, it becomes more profitable to produce tractors in Mexico and resources move out of tomatoes. Production moves from *B* to *D*. If the tariff is high enough, a so-called prohibitive tariff, the price line will rotate by even more and relative prices will be the same as under autarky. The economy will neither import nor export and will produce and consume at *A* with a corresponding loss in utility.

Assuming the tariff is not prohibitive and Mexico's trading partner does not retaliate with a tariff on tomato imports, how much will Mexico import and export, and where will it consume? Assuming Mexico is a small economy and a price taker on world markets, then the international terms of trade remain unchanged and are now drawn through point *D*. To find consumption, recognize that Mexican consumers, like Mexico's producers, no longer face international prices; they now face relative prices, which include the tariff the government placed on tractors. The economy consumes at E, a point on the international terms of trade line that is also at the tangency between the price line including the tariff and IC_2. In the figure, both imports and exports, identified by the difference between production and consumption at points *D* and *E*, are lower with the tariff than under free trade.

The model predicts that Mexico is worse off with a tariff, consuming at a lower level of utility, IC_2, than under free trade, IC_3. This is because it has lost some of the potential gains from exchange and specialization. The loss in utility is analogous to the findings of the partial equilibrium model of the tariff in Figure 19–1, which identifies deadweight losses and lost consumer surplus resulting from the tariff.

officials to maintain the tariff and the trade protection of their industry. The more concentrated the industry, the fewer the firms who share in the benefits of protection. Some of the resulting profits may then be spent supporting the government in power, which, in return, maintains the trade protection.

The political economy of trade protection applies not only to developing nations. It also helps explain the use and persistence of tariffs, quotas, and subsidies in developed nations. The political calculations rather than the economic merits explain Japan's protection of its rice farmers and the European Union's protection of its dairy industry. Politics in the United States explains trade protection for everything from steel to sugar. Because of domestic political interests, it often is difficult to remove trade protections once they are introduced. One solution is to include tariff reductions as part of multilateral treaties. In the case of the Thai beer tariff, it expired in 2010 as part of the ASEAN Free Trade Area, a regional trade treaty among nations in Southeast Asia.

PRODUCTION SUBSIDIES

Direct subsidies are an alternative to tariffs or quotas as a means of protecting domestic manufacturers. Many high-income economies use large subsidies to support agricultural production and protect it from global competitors, often at the expense of

potential competitors from developing countries. Emerging economies use subsidies as well, often to shelter more capital-intensive manufacturing from import competition. The impact of subsidies is similar to tariffs in some ways: A 20 percent protective tariff on imports and a 20 percent subsidy on output have identical effects on producer surplus. But the effects on consumers and the government budget can be quite different.

A subsidy of s_0 percent, equal to the earlier tariff of t_0, effectively moves the supply curve from S to S', as shown in Figure 19–2. Production shifts out from Q_2 to Q_4, but prices do not change, so consumers still purchase Q_1 at the original price P_w. Thus a big advantage of using subsidies is that there is no loss of consumer surplus. The total cost of the subsidy is $a + b$, but in this case, the funds come from the government budget rather than being imposed only on consumers of the particular good. Producers are equally happy with a subsidy or an equivalent tariff. As with the tariff, they gain producer surplus a, with a resource loss of b. Thus the deadweight loss is only triangle b, which is smaller than the amount $b + d$ with a tariff or quota.

If protection is to be employed, economists generally prefer subsidies over tariffs (and tariffs over quotas) because consumers are able to purchase more of the good while society pays no more for production than under an equivalent tariff. The cost is borne by taxpayers as a whole, not just the consumers of the product. Subsidies usually appear as an expenditure item in the government's budget, so there is an annual accounting for the costs of protection. Unfortunately, government officials tend to prefer quotas or tariffs over subsidies, precisely because the costs do not appear on

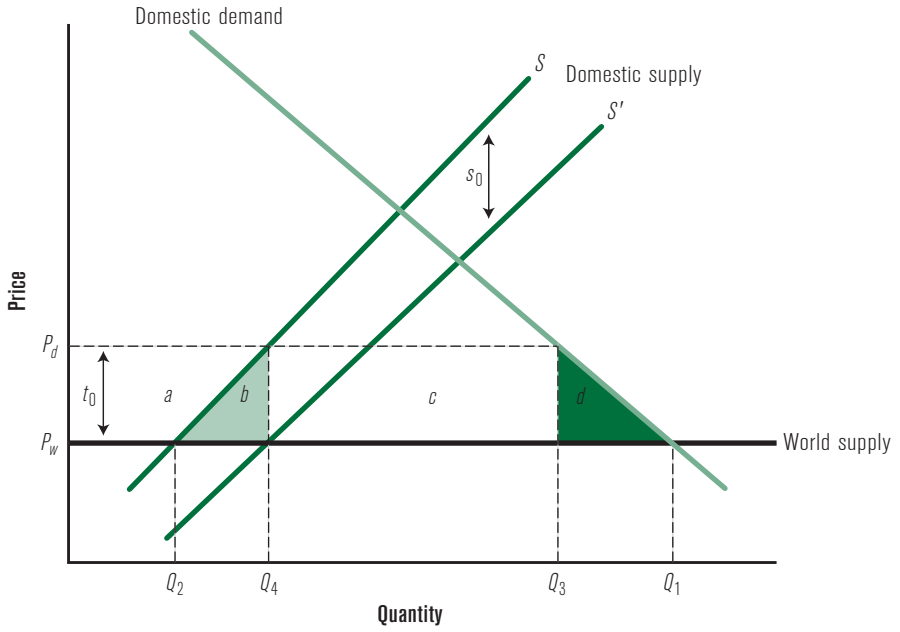


FIGURE 19–2 **The Impact of Subsidies on Firms Competing with Imports**

the budget and are less obvious to society as a whole. Subsidies also suffer from the same problem facing tariffs and quotas. What sectors warrant protection, whatever the form? How does a government decide which industry to support? How long must the subsidy last? While there are examples of governments identifying and supporting industries that went on to be successful, there are even more examples of when this approach failed. "Picking winners" is never easy.

EXCHANGE-RATE MANAGEMENT

Tariffs, quotas, and subsidies are only some of the tools policy makers can use to influence trade. One of the most powerful instruments they can employ is the exchange rate, which often affects more transactions than any other single price in an economy and directly affects the domestic prices of everything traded. Tariffs, quotas, and subsidies are used to support very specific products and sectors by changing the relative price between those products and all other products. The exchange rate, by contrast, has a uniform effect on the prices of all tradable goods but alters the price between tradables and nontradables.[4]

Figure 19–3 illustrates the market for foreign exchange, in which exporters supply and importers demand foreign currencies. The vertical axis is the exchange rate measured in local currency; for example, pesos per dollar. With a floating exchange rate system, a rate of e_e is an equilibrium rate that clears the market. IS strategies have a tendency to cause the domestic currency to appreciate. Tariffs and quotas on imports reduce the demand for foreign exchange at every possible exchange rate. In Figure 19–3, this leads to a shift in the demand curve and to a new equilibrium exchange rate, e_e'. The new exchange rate acts like a tax on exports because it now costs more foreign exchange to buy a unit of domestic currency or, equivalently, a dollar's worth of exports now returns fewer pesos. This is an important insight. Anytime a government uses trade protection it affects both imports *and* exports. Many IS strategies were designed to foster industrial development but at the same time often hurt traditional exports, including agricultural goods, which often are produced by poorer households.

In formulating trade strategies, governments also often intervene directly to fix their exchange rate, causing it to be either overvalued or undervalued. Historically, many governments have overvalued their currency and intervened to hold the official exchange rate below e_e at a rate like e_0. Such a policy can directly harm domestic producers of goods that compete with imports. For instance, suppose that a ton of corn can be imported for $150. At an (equilibrium) exchange rate of 10 pesos to the dollar, a ton of imported corn would cost 1,500 pesos. Yet, if the peso was overvalued

[4]As discussed in the previous chapter, tradable goods include exports, imports, and all goods and services that could be exported or imported, with prices determined on world markets. Nontradable goods are produced locally and not subject to competing imports, with prices determined by local supply and demand conditions.

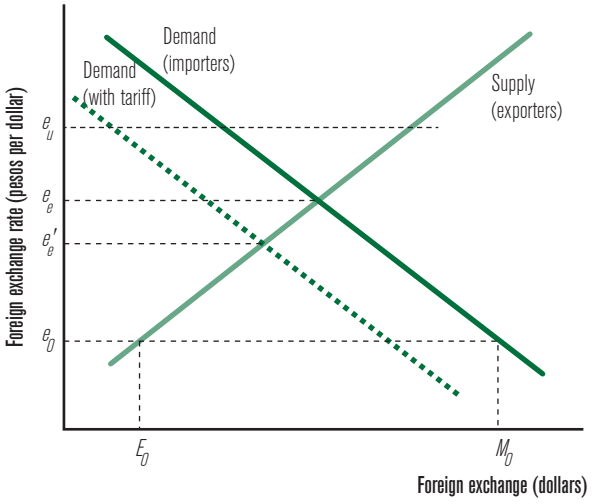


FIGURE 19–3 Overvalued and Undervalued Exchange Rates

relative to its equilibrium, say at the rate of 8 peso per dollar, that same ton of imported corn would cost only 1,200 pesos. This might make consumers happy, but local corn farmers would have to compete against artificially cheap imports, resulting in a 20 percent decrease in their earnings. At the same time, anything exported and sold in dollars earns fewer pesos. The result is an increase in import demand and a fall in exports. As contradictory as it might at first seem, IS strategies often employed overvalued exchange rates as a policy instrument. The goal was not to make *all* imports cheaper because the intent of IS is to protect domestic industries. Overvalued exchange rates were used to make capital goods and other critical imports, such as food or fuel, cheaper in terms of domestic currency. Nonessential imports are kept out of the domestic market via high tariffs and quantitative barriers. Ironically, there can even be a vicious circle in which tariffs lead to overvalued exchange rates, which reduce the cost of imports and thus create a need for even higher tariff rates to compensate.

A quick look at Figure 19–3 identifies some of the problems associated with employing overvalued exchange rates. Like any price ceiling, it creates excess demand. At e_0 the excess demand for foreign exchange equals, $M_0 - E_0$. Because the foreign exchange market now fails to clear, it becomes necessary to ration foreign exchange, requiring licenses and the administrative controls to oversee them. Such licenses easily become a source of rent seeking and corruption. Overvalued exchange rates also give rise to black markets for foreign exchange.[5] This is because overvaluation creates chronic excess demand for foreign currencies. Incentives are created to deal foreign

[5]In the 1970s and 1980s, travelers to many developing nations, including the authors of this text, invariably were asked by taxi drivers from the airport if they wanted to exchange money, a sure sign of an overvalued exchange rate and the black markets they encouraged.

exchange at rates higher than the official one and in so doing may frustrate the intended goals of the IS and exchange rate policy, and generate more rent seeking.

As recently as the 1970s and 1980s, the combination of overvalued exchange rates and pervasive tariffs and quotas was quite common. Some of this was because of IS strategies. But in many other cases, governments overvalued their exchange rates primarily to lower the price of imported goods for those who were politically well connected, including the political elite, urban consumers, and the military. Today, following one of the core precepts of the Washington Consensus discussed in Chapter 5, official exchange rates set at overvalued levels are relatively uncommon.

Some governments follow the opposite strategy and intervene to hold the exchange rate above e_e at a rate like e_u. They **undervalue their currency**. This approach makes all imports more expensive and, at the same time, makes exports more profitable by increasing their price in domestic currency. Undervalued exchange rates help stimulate exports and provide protection to firms competing with imports by raising the price of competitive products. They have been used in some East Asia economies, most notably China, as part of more outward-looking strategies. We evaluate their use later in the chapter.

OUTCOMES OF IMPORT SUBSTITUTION

Import substitution has the potential to be an effective strategy for certain sectors over a limited period of time. Almost all countries have tried it at one stage or another, and many have achieved some success. Even Latin America, which is often cited for the failures of IS, had more rapid growth in the 1960s and 1970s under IS regimes than it did for the next 20 years as it moved in the direction of more liberal trade. But in Latin America and elsewhere, all too often, the basic conditions for prolonged success are not met. In many countries, IS protected too many activities and remained in place for too long. Developing countries are littered with infants that never grew up and were never able to compete internationally, such as the petrochemical industry of Colombia, the automobile industry of Malaysia, and the textile industry of Kenya. Firms in these industries have tended to require protection indefinitely, at continuous cost to the rest of society. Sometimes IS fails because of an initially poor choice to protect the wrong kinds of activities (the microcomputer industry in Brazil); often it stems from the reluctance of governments to remove the protection given to politically well-connected industrialists.

Most developing countries have relatively small internal markets, either because per capita incomes are low or populations are relatively small. This makes them poor candidates for IS because the domestic market quickly becomes saturated. The next step, expansion to export markets, may never happen because of low productivity growth which may itself be the result of the trade protection intended to help the industry. Facing a small domestic market there may be little domestic competition and few incentives to innovate. A limited market means that firms cannot take advantage of

economies of scale and may produce at less than their minimum efficient size. And by reducing their commercial links with the rest of the world, import-substituting countries limit their exposure to new technology and ideas.

IS has had some limited success in bigger economies with large internal markets, at least for a short period of time. But even the very large economies of China and India had more success when they stopped looking only inward—for the Chinese economy after the reforms that began in 1978 and for India after 1991. Of course, these reforms entailed more than a change in trade strategies, and it is difficult to pinpoint the specific contribution of the change in trade policy.

Ironically, many countries that try IS run into balance-of-payments problems from growing trade deficits. Even though the strategy is designed to replace imports with domestic production, not all imports can be replaced (especially capital goods). Because the strategy effectively discourages exports, foreign exchange earnings lag, especially because IS often is accompanied by an overvalued exchange rate. Therefore, many countries following this strategy have had to borrow heavily and found it difficult to meet their debt service requirements. Better macroeconomic management could have avoided these problems, but the instruments of IS did not help.

Underlying the protective regime also is a set of incentives that reward political lobbying, corruption, and bribery more than economic efficiency and competitiveness. When higher domestic costs, reduced import prices, or better-quality foreign goods erode the competitive position of domestic firms, a natural reaction is for these firms to turn to the government for enhanced protection. This option blunts the competitive instincts of entrepreneurs, who normally would have to cut costs, improve quality, and thus raise productivity. In this environment, the most successful managers are those who have the political skills or connections with which to bargain effectively, or simply bribe, officials who determine tariff rates, administer import quotas or distribute foreign exchange. The pursuit of such unearned benefits is known as *rent-seeking behavior*. Rent seeking also hurts economies by diverting real resources from productive uses.

EXPORT ORIENTATION

Since the late 1980s, many countries have shifted the balance of their trade policies away from IS toward more export-oriented trade policies, in which firms compete on global markets. This strategy has many names, including *outward orientation, openness*, *export promotion,* and *export orientation*.[6] The idea is to introduce policies that

[6]Joining a free trade area, as Mexico did with Canada and the United States under the North America Free Trade Agreement (NAFTA) can also be seen as a move toward greater outward orientation. However, some free-trade areas or other types of economic integration can be more like import substitution if the combined economies are small and members build tariff walls around themselves, reducing trade with nonmembers.

encourage firms to produce products that are competitive on world markets, especially labor-intensive manufactured exports and agricultural products but also competitive substitutes for imports. The key difference with IS is that this strategy uses global competition to encourage investment, productivity gains, learning, and new technology to support growth.[7]

In the typical pattern, in the early stages, firms manufacture and export relatively simple labor-intensive products, such as textiles, clothing, shoes, toys, electronic equipment, and furniture. Some countries also export agricultural products and labor-intensive agroprocessing goods, such as fresh vegetables, fruit juices, or cut flowers, and increasingly certain services including data entry, basic accounting, and call centers. Then, over time, as workers learn new skills and gain access to improved technology, firms begin to shift to more-sophisticated products (which pay higher wages to match greater productivity and skills). As the country's comparative advantage gradually shifts, the mix of exports changes as well to include more-advanced electronic devices, automobile parts, higher-end clothing, steel, and many other consumer and capital goods.

The shift toward more outward-oriented policies among developing countries began in the 1960s with the Four East Asian Tigers: Hong Kong, Korea, Singapore, and Taiwan. They demonstrated that developing countries could compete on world markets by carving out niche markets in producing labor-intensive manufactured products. Their rapid growth in trade was accompanied by accelerating economic growth, reductions in poverty, and other advancements. Following their example, most of the very rapidly growing developing countries in recent decades have introduced some form of this strategy, including Chile, China, Indonesia, Malaysia, Mauritius, Poland, Thailand, Tunisia, and Vietnam. India's growth rate also accelerated in the 1990s as it shifted to a more outward-oriented strategy.

But not all countries that have shifted to policies that are more outward oriented achieved the success they had hoped for. Bolivia introduced many steps to open its economy, and while growth has recovered from the negative rates that prevailed in the early 1980s, it has averaged only about 1.5 percent since the late 1980s. Many of the countries formed by the breakup of the Soviet Union pursued more open trade strategies but realized little growth in either trade or output, especially during the 1990s. Without the other elements that encourage investment and entrepreneurship, including institutions that protect property rights and enforce the rule of law, trade reforms alone had minimal impact. In the 1990s, Zambia implemented reforms that resulted in one of the most liberal trade regimes in all of Africa. But the hoped for growth in exports did not follow. Part of the problem was macroeconomic instability, including high inflation and a volatile real exchange rate. Under

[7]For an overview of the shift in thinking about trade policy and development between the 1960s and 1990s, see Anne O. Krueger's presidential address to the American Economic Association: "Trade Policy and Economic Development: How We Learn," *American Economic Review* 1, no. 87 (March 1997).

such conditions, firms were reluctant to invest in new ventures whether in exports or import-competing sectors.[8]

REMOVING THE BIAS AGAINST EXPORTS

Policies that favor imports, including tariffs and overvalued exchange rates, implicitly are biased against exports. Recent decades have seen a retreat from such policies. As shown in Figure 19–4, there has been a clear global trend to reduce average tariffs. In South Asia, average tariffs fell from over 50 percent in 1987 to under 15 percent by 2007. Tariffs in all regions fell. Some changes in individual countries have been dramatic. In Costa Rica, the average statutory tariff was 55 percent in 1980 and only 6.2 percent in 2007; in Turkey it fell from 44 to 1.9 percent over the same time period. Currencies officially set at overvalued rates once common are also now rare.

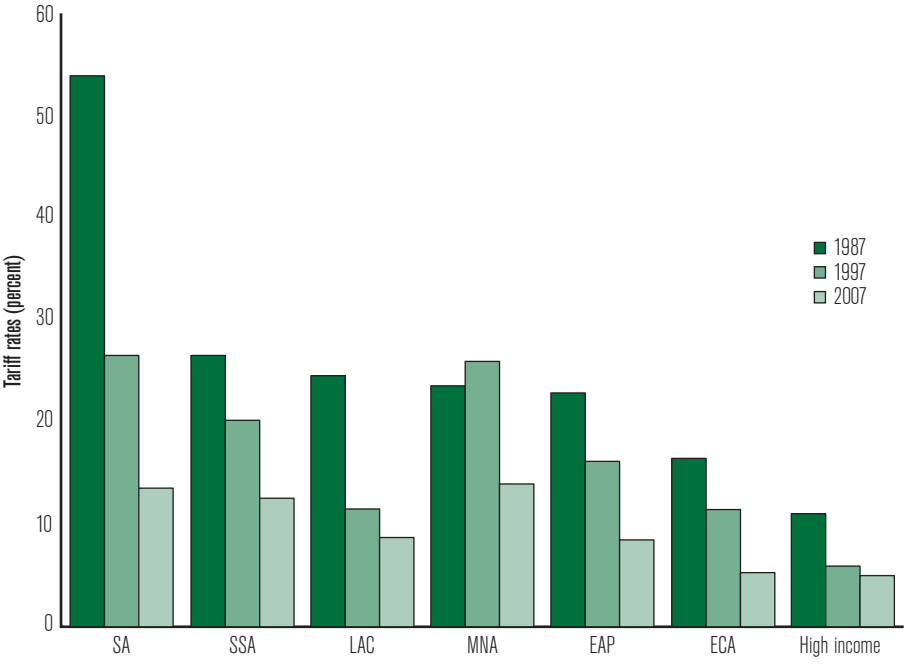


FIGURE 19–4 Average Unweighted Tariff Rates by Region

EAP, East Asia and the Pacific; ECA, Europe and central Asia; LAC, Latin America and the Caribbean; MNA, Mideast and North Africa; SA, South Asia; SSA, sub-Saharan Africa.
Source: World Bank, "Trends in Average Applied Tariff Rates in Developing and Industrial Countries, 1981–2007," available at http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:21051044~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html.

[8]World Bank, "Trade Liberalization: Why So Much Controversy?" in *Economic Growth in the 1990s: Learning from a Decade of Reform* (Washington, DC: World Bank, 2005).

The basic steps toward full-fledged liberalization aimed at removing the bias against exports will sound familiar:

- Remove quotas and tariffs and other forms of protection, especially on capital and intermediate goods.
- Allow the currency to float with a market-determined exchange rate, and ensure macroeconomic stability through prudent monetary and fiscal policies.
- Reduce unnecessary regulatory burdens, bureaucratic costs, and red tape that add to business costs.
- Keep factor markets flexible, especially for labor and credit, with market determined wages and interest rates.

In a full liberalization strategy, these steps are augmented with general approaches to strengthen infrastructure (especially roads, ports, electricity supplies, and telecommunications to connect firms with the rest of the world and reduce production costs), invest in education and public health (to ensure a well-educated, healthy, and productive workforce), and improve governance (so that property rights are respected).

These basic policies will not surprise any student of neoclassical economics or simply someone who has been reading this textbook. Outward orientation is often seen as synonymous with broader market-based and liberalized economic policies. The neoclassical approach does not prescribe a bias toward exports but rather a regime that is neutral in its treatment of exports, import substitutes, and nontraded goods. In this view, the strategy is not aimed at exports per se, but export growth is likely to follow because the distortions that inhibited exports and favored other activities are removed.

FAVORING EXPORTS

While liberalization has theoretical appeal, few countries actually have followed a pure neoclassical strategy, and some of the most successful countries were far from it in some respects. Hong Kong, at least before its return to China in 1997, was the closest, where the economy was very open, with few government interventions and distortions. Singapore also followed a fairly liberalized model but nevertheless diverged from the neoclassical norm in that many of its largest export industries are government-owned service firms in telecommunications, port services (including one of the world's best-run airports), and air transport (the world-renowned Singapore Airlines).[9] Korea was much more interventionist. It erected stiff protective barriers against many imports,

[9]Hong Kong and Singapore often are compared. They are both city-state, island economies that have transitioned to high-income status and relied heavily on trade to do so. Hong Kong, however, has come considerably closer to free trade than more interventionist Singapore. See Lawrence Krause, "Hong Kong and Singapore: Twins or Kissing Cousins?" *Economic Development and Cultural Change* 36, no. 3 (April 1988).

especially consumer goods. A visitor could literally see the effect. In the 1980s one could stand for hours at one of the busiest street corners in Seoul, where many lanes of traffic intersect, and not see a single imported car. The streets were choked with automobiles, but only those produced by Daewoo, Hyundai and a few other large Korean conglomerates. Tariff and other barriers kept Fords, Toyotas, VWs and every other international brand out of the domestic market.

Korea's government also controlled interest rates at below-market levels, directing cheap credit to favored industries and firms, including exporters. What made the approach outward-looking is that interventions were used to induce, and sometimes force, firms to meet ambitious export targets. Success or failure was dictated to a large degree by the ability to compete on world markets, rather than domestically.[10] Korea went beyond simple outward orientation and neutrality and introduced a bias toward exports. In contrast to an inward-looking regime, a Korean firm could not take advantage of protection or other preferential treatment unless it met stringent export targets. Korean car manufacturers, for example, became competitive on global markets and began to export while selling to a protected local market. Some of the tools of protection were used to help fledgling exporters who were learning by doing: adopting new technologies, learning how to master them, reducing costs of production, finding and entering markets overseas, and eventually competing on equal terms with foreign firms.[11] But not all firms succeeded. Korea's drive to expand its heavy and chemical industries in the 1980s met with more limited success.

The strong bias and intervention to support exporters followed in Korea was not the rule for all the Asian exporters. Indonesia, Malaysia, Taiwan, and Thailand used a more mixed approach.[12] Heavily protectionist and interventionist in certain areas, some of the industries explicitly supported never became internationally competitive, such as Malaysia's Proton automobiles and Indonesia's IPTN aircraft. What was more successful were the policies that insulated export industries from the distortions of the home market and permitted firms to buy inputs and sell output at close to world market prices, as the neoclassical strategy dictates, permitting a wide range of exporting firms to become competitive in international markets.

China has been more interventionist and pursued outward-oriented policies far from free trade and the full-fledged liberalization described earlier. As economist Dani

[10]See Marcus Noland and Howard Pack, *Industrial Policy in an Era of Globalization: Lessons From Asia* (Washington, DC: Institute for International Economics, March 2003).

[11]For a wonderful description of how Korean firms implemented "learning by doing" read Ira Magaziner and Mark Patinkin, "Fast Heat: How Korea Won the Microwave War," *Harvard Business Review* 67 (January–February 1989).

[12]The revisionist school on Korea, questioning the neoclassical approach, was led by Alice Amsden, *Asia's New Giant: South Korea and Late Industrialization* (New York: Oxford University Press, 1989). Taiwan's interventionist strategy is documented by Robert Wade, *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization* (Princeton, NJ: Princeton University Press, 1990). The World Bank attempted to reconcile neoclassical and revisionist thinking in its much-discussed book, *The East Asian Miracle: Economic Growth and Public Policy* (Washington, DC: World Bank, 1993).

Rodrik has put it, "China's policies resemble more those of a country that messed up big time than those of a country that became a formidable competitive threat in world markets to rich and poor countries alike."[13] In 1993 after a decade and a half of rapid export and economic growth, average tariffs still remained at 40 percent, the sixth highest level in the world at that time. (By 2007, six years after joining the World Trade Organization [WTO], Chinese tariffs had fallen below 10 percent.) But even more than its prior tariff regime, China, along with Korea and Taiwan in earlier years, used and still uses exchange rate policy to favor exporters by undervaluing its currency.

By pegging the exchange rate above the market-clearing rate, imports become more expensive and exports more profitable. This policy contributed to China's rapid export growth, its large trade surplus, and the resulting accumulation of foreign exchange reserves. An undervalued currency acts as an indirect subsidy for Chinese exporters. Like all subsidies, it comes at a cost, making imports more expensive and thus reducing consumption and consumer welfare for the average Chinese. Reserve accumulation can be helpful to a point, but reserves, by definition, are resources that a society puts aside and does not use for current consumption or investment. They are subject to exchange rate and other risks and may lose their value, as demonstrated during the financial crisis of 2008.

The costs of China's exchange rate policy also have external ramifications. China frequently is cited as engaging in unfair trade practices because of the way it manages its exchange rate. Legislation has been proposed in the U.S. Senate to impose tariffs on Chinese goods equal to the extent of undervaluation of its currency, a quantity that is subject to some debate. Some, including Ben Bernanke, chairman of the U.S. Federal Reserve Bank, have a different concern. They argue that China's huge foreign reserves, a consequence of its exchange rate policy, were part of the underlying causes of the 2008 global financial crisis. These reserves, Bernanke suggests, contributed to the global economic imbalances that drove interest rates down and created the conditions that encouraged the huge risk taking and subsequent collapse of financial institutions in search of higher returns. Continuation of such policies also is seen as an impediment to restoring the global imbalances that remain today and that may limit global economic recovery (Box 19–3).

BUILDING EXPORT PLATFORMS

Another approach taken in almost all of the countries that have achieved rapid export growth was to establish specialized export platform institutions that enabled exporters to import and sell at close to world market prices, even in the presence of more widespread distortions. The platforms took a variety of forms, including export processing zones (EPZs), bonded warehouses, duty exemption programs, and science and

[13]Dani Rodrik, "What's So Special About China's Exports?" *China in the World Economy* 14, no. 5 (September–October 2006).

BOX 19-3 IS CHINA'S EXCHANGE-RATE POLICY UNFAIR?

China has been criticized by some for maintaining an undervalued exchange rate that makes its exports more profitable; others argue that the currency is not seriously misaligned. China for a decade kept its currency, the renminbi (RMB), fixed against the U.S. dollar at a rate of 8.28 yuan to $1. Then in 2005 China introduced a managed float. and the yuan rose in value to 6.83 to $1. As the world moved into recession in 2007 and 2008, China again fixed its exchange rate ending further revaluation. Despite fixing the exchange rate, however, Chinese exports fell sharply in 2009 due to declining world demand for Chinese manufactures. With the beginning of recovery from the world recession in 2010, China's exports recovered rapidly and it announced in June of that year that it would allow its currency to float; it was expected that the Chinese currency would appreciate in value.

Several pieces of evidence suggest that the RMB was still undervalued in mid-2010—that is, kept at above the market clearing rate, as with exchange rate, e_U, in Figure 19–3. China's current account surplus (exports minus imports of both goods and services) had grown to 8 percent of gross domestic product (GDP) in 2008 and remained at a very high level of 4 percent in 2009, as predicted by Figure 19–3. Normally, a current account surplus is matched by an outflow of foreign capital, as countries invest the proceeds abroad. But China attracted net capital *inflows* of an additional 1.7 percent of GDP in 2009. Because this current account surplus has existed for many years (since 1994) China accumulated very large foreign exchange reserves of US$2.45 trillion as of the end of the second quarter of 2010, equivalent to roughly two years of imports.

China's exchange rate is not the only policy that affects these balances. The government maintains an array of restrictions on capital flows, which tend to discourage outflows of capital. In the absence of these controls, the surpluses would not be so large. Even taking these arguments into account, many economists still conclude that the RMB is undervalued, perhaps by 15 to 25 percent.[a]

An undervalued RMB makes Chinese exports more profitable, and indeed exports rose fourfold between 2000 and 2009, much of it through importing components from other countries in the region and assembling finished products for export. This export growth is a key reason for China's overall economic growth and its impressive reduction in poverty, as many unskilled workers have found jobs in the special economic zones (SEZs; much like export processing

[a]Morris Goldstein and Nicholas Lardy, *The Future of China's Exchange Rate Policy: Policy Analyses in International Economics 87* (Washington, DC: Institute for International Economics, 2009).

zones, discussed later in this chapter). Moreover, the buildup in reserves also allows China to maintain macroeconomic stability, as it is unlikely to face the kind of capital account crisis many of its neighbors faced in the late 1990s, as described in Chapters 12 and 13. China's buildup in reserves is partly a defensive move in response to these crises.

But the policy of maintaining an undervalued exchange rate has its costs, especially the longer it is in place. Restraining consumption reduces the welfare of ordinary Chinese by making imports of consumer goods more expensive. Foreign exchange reserves equivalent to two years of imports are also far greater than China requires to defend against a future run on its currency as happened elsewhere in Asia in 1997 and 1998. These large reserves are also mainly invested in government bonds (one third of China's reserves are invested in U.S. Treasury bonds), which in recent years have paid a low rate of interest.

But perhaps most important the undervalued exchange rate brings threats of protectionism from trading partners: Individual U.S. senators have proposed putting a high tariff on all Chinese imports unless China revalues the RMB substantially. China's exchange rate policy has become a convenient scapegoat for the U.S. trade deficit, even though China's exchange rate contributes relatively little to the U.S. deficit compared to other factors more in the control of the United States, particularly its budget deficit and low personal saving rate. Undervalued exchange rates and a rapid growth of exports do not usually lead to threats of retaliation when a country's total foreign trade is small as was the case in China in the 1970s when its total exports were under US$10 billion a year. In those years, threats of protection were made against the much larger exports of Japan, which was also believed to have an undervalued currency. But in 2009 China passed Germany to become the largest exporting country in the world and hence the focus of all those concerned about the impact of rising imports on their domestic industries.

China's exchange rate policy has contributed to its rapid economic growth and poverty reduction in recent years, and its large reserve holdings help stabilize the economy and reduce the chances of a major capital account crisis. An undervalued exchange rate also helped China postpone the date when exporters of products depending on cheap labor, because of rising Chinese wages, would have to give way to exporters in other countries where wages are lower. But, over the longer run, there will be little extra benefit for China to continue to accumulate reserves at the expense of consumption and possible protectionist retaliation. China, like Japan and the other East Asian exporters of manufactures who depended for a time on undervalued exchange rates, is likely to let its exchange rate rise to a point at which the renminbi no longer is undervalued.

technology parks. These institutions are designed to create an enclave for fledgling export industries, where they can be insulated from the controls and price distortions of the protected domestic market and be better able to compete on world markets.[14]

EPZs are located physically or administratively outside a country's customs' barrier, typically as fenced-in areas near a port, that provide exporters access to duty-free imports, reduced business regulations, expedited customs clearance, and reasonable infrastructure. Bonded warehouses essentially are single-factory EPZs that can be located anywhere, yet still have many of the other advantages of EPZs. These factories are called *bonded warehouses* because firms usually post a bond as a guarantee against any duties that might be applicable to imports diverted to the domestic market, and if they sell output locally instead of export it, they are liable to pay the import duties. Duty exemption systems allow qualifying firms to import their inputs duty free. Science and technology parks are like EPZs, but with infrastructure specially suited for electronics, pharmaceuticals, biomedical products, or other products related to science and technology.

The basic idea is straightforward. Potential exporters face a variety of challenges, including high tariffs or quotas on their inputs, poor infrastructure, and red tape. The neoclassical solution is to remove all import tariffs and quotas, improve infrastructure, and reduce red tape. The problem is that, for a variety of institutional and political reasons, no country realistically can implement all these steps quickly. Most governments lack the large number of trained personnel to take on so many issues at once, and political forces are likely to fight certain steps, like reducing tariffs. It can take many years for a country to build the infrastructure and introduce all the policy changes needed for firms throughout the economy to become competitive. The export platform approach is to allow one part of the economy to be competitive, while steps are taken over time to introduce economywide changes. Providing an enclave allows some firms, mainly producing labor-intensive manufacturers, to become competitive on world markets before all the required steps are in place for the economy as a whole.

Most of the successful exporting countries established at least one, and in most cases more than one, of these programs. Malaysia relied mainly on EPZs that provided reliable infrastructure and allowed exporters (mainly in electronics) to import and export without being taxed; the government also established bonded warehouses and a duty exemption system for other exporters. Indonesia established an agency in the Ministry of Finance that granted exemptions from import licensing restrictions and drawbacks (rebates) for duties paid on imported inputs. The vast majority of Tunisia's manufactured exporters operate as bonded warehouses, whereas Mauritian exporters are located mainly in export processing zones. China's special enterprise

[14]See David L. Lindauer and Michael Roemer, eds., *Asia and Africa: Legacies and Opportunities in Development* (San Francisco: ICS Press, 1994), esp. chaps. 1 and 11. Also see Steven Radelet, "Manufactured Exports, Export Platforms and Economic Growth," CAER Discussion Paper No. 42, Harvard Institute for International Development, Cambridge, MA, September 1999.

zones, located up and down the coast, have been the source of most of its manufactured exports. In the Dominican Republic, exports from EPZs increased by a factor of five during the 1990s, and by the end of the decade accounted for over 80 percent of all exports. EPZ employment reached nearly 200,000 people by the end of the 1990s, about 17 percent of the Dominican workforce. About 60 percent of employees are women, a pattern that is typical of many EPZs.[15]

Export platforms have not always been successful. When they are located far from ports in an attempt to spur development in isolated areas, production costs usually are too high for firms to compete. If firms in export platforms still face high regulatory costs or must wait long periods of time for goods to clear customs, they will not be successful on global markets. And export platforms cannot overcome poor macroeconomic management that leads to high inflation, overvalued or volatile exchange rates. Egypt's EPZs have not spurred faster export growth because they still face high production costs. Kenya's bonded warehouses began to expand in the 1980s, but when macroeconomic policies deteriorated and the currency became overvalued in the 1990s, 60 of 70 bonded warehouses closed. Export platforms, whatever their forms, are no silver bullet, but they provide examples of how innovative institutions shaped to a country's specific needs can help support market-based growth.

TRADE STRATEGY AND INDUSTRIAL POLICY

Almost all nations recognize the importance of increased integration into the global economy and the need to look outward. But an outward orientation is not the same as free trade or laissez-faire. Policy neutrality regarding production for exports versus the domestic market has been the exception rather than the rule. The debate revolves less around inward versus outward orientation and more around the use of **industrial policy**, a broad set of interventions including, but not limited to, undervalued exchange rates and EPZs, which governments may use to favor one set of economic activities over another.

Most economists remain wary of industrial policies. They recognize the various and pervasive market failures that constrain the growth of new firms and activities, from credit markets unwilling to finance new products to the lack of complementary inputs to service export initiatives, but are unsure governments are up to the challenge of improving the situation. Governments tend not to have the knowledge and information to figure out what firms really need, and more often than not industrial policies breed corruption or promote inefficiency. Such economists recognize that there are examples of success with industrial policy. POSCO, a Korean steel producer, was formerly a state-owned enterprise that grew behind high trade barriers into one

[15]Staci Warden, "A Profile of Free Trade Zones in the Dominican Republic," Harvard Institute for International Development, Cambridge, MA, 1999.

of the world's largest and most efficient private steel corporations. Farmed salmon in Chile, today a highly successful industry, was initiated and fostered by a quasi-public agency. But for each of these positive examples, there are many more where industrial policy has failed from airplanes in Indonesia to tanneries in Ethiopia.

Harvard University economist Rodrik, takes the opposite view. He acknowledges the failures of past strategies. Import substitution, planning, and state ownership of enterprises were tried and, in some cases, succeeded for a while but in the end failed to generate sustained economic growth and development. But liberalization and full reliance on market forces *also* delivered less than what was hoped for, especially in Latin America and sub-Saharan Africa. Even where there has been success, cursory examination of the economic miracles of past decades reveals how far they were from the neoclassical ideal and how central a role government played in favoring exports, specific sectors, and even specific industries and firms. In East Asia and Latin America it was the combination of private initiative and public interventions to support exports that characterized most cases of successful trade and development strategies.

Rodrik and others see the role of government as extending beyond the traditional areas of ensuring macro stability, providing public goods, and resolving externalities (such as pollution). Government also has a coordinating role to play in facilitating the structural change inherent in the development process. Economies need to move away from traditional products and to diversify their economies, but this is not easy to do. As Rodrik puts it, the problem is one of "self-discovery." Comparative advantage and factor endowment theory offer important insights into the benefits and direction of trade, but they alone are insufficient to explain the specific goods different nations export. Labor abundant economies may all export labor-intensive goods, but why did entrepreneurs in Bangladesh specialize in T-shirts and in Pakistan in soccer balls? Why is Korea the dominant exporter of microwaves while Taiwan, an economy with similar factor endowments, exports none but dominates in bicycles? In each of these examples individual entrepreneurs figured out that they could succeed in a particular product. Their success was then imitated by others and the industry expanded.

Rodrik argues that in most developing countries, governments need to do something to encourage this process of self-discovery.[16] Entrepreneurs need to determine which existing goods and technologies from abroad can be adapted to local conditions and successfully exported in global markets. But they may be reluctant to do so for many reasons. For one, the gains of their individual efforts may be undercut by the inevitable imitation that will follow their success. In this instance, self-discovery presents a classic public good problem: Information on what products a firm might successfully export is both nonexcludable and nonrival; therefore, not enough of this information will be forthcoming. Government intervention can help. Some form of protection or subsidy must accrue to the initial investor, and not the copycats, to

[16]Dani Rodrik, *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth* (Princeton, NJ: Princeton, 2007), 4.

encourage the self-discovery that may be absent on its own. Other coordination problems may be resolved by government action, for example, in the setting of product standards for exports or relieving infrastructure constraints.

Rodrik is not alone in his support of industrial policy as part of a developing nation's trade policy. In their exhaustive review of the literature on trade and industrial policy for a recent volume of the *Handbook of Development Economics*, economists Ann Harrison and Andrés Rodríguez-Clare do not go as far as Rodrik does in supporting hard industrial policies, including trade protection in the form of tariffs, quotas, and subsidies. But they do see a need for soft industrial policies aimed at resolving the coordination problems that nations face trying to break into established export markets. They argue for partnerships between the public and the private sectors in identifying and resolving the binding constraints facing entry into in new markets.[17]

TRADE, GROWTH, AND POVERTY ALLEVIATION

The promotion of exports and the use of industrial policy is not the only area where there is disagreement among development economists over trade policy. A lively debate remains over the centrality of trade in generating economic growth and of the consequences of increased trade for alleviating poverty. This may seem surprising given the many theoretical reasons discussed in the previous chapter for how trade can improve factor productivity and encourage economic growth and for why trade favors a nation's abundant factor and, in the case of poor labor abundant economies, is expected to reduce poverty. The empirical evidence, however, is less straightforward. We begin by discussing the empirical evidence on trade and growth; in the next section, we turn to the evidence on trade and poverty alleviation.

Hundreds of academic papers have been written examining the relationship between trade and growth. It is important to understand why it is so difficult to establish an empirical pattern linking trade, trade policy, and economic growth. Problems of reverse causality, omitted variable bias and the measurement of trade policy confront the economist trying to sort out the underlying relationship. (If you have taken a course in econometrics some of these common problems may be familiar to you.)

Economists Jeffrey Frankel and David Romer tackle the difficult issues of the direction of causality and omitted variable bias in a paper aptly titled, "Does Trade Cause Growth?"[18] A strong positive relationship between trade and per capita income does not prove that the former is the *cause* of the latter. It is entirely plausible

[17]Ann Harrison and Andrés Rodríguez-Clare, "Trade, Foreign Investment, and Industrial Policy for Developing Countries," in Dani Rodrik and M. R. Rosenzweig, eds., *Handbook of Development Economics*, vol. 5 (Amsterdam: North Holland, 2009).

[18]Jeffrey Frankel and David Romer, "Does Trade Cause Growth?" *American Economic Review* 89, no. 3 (June 1999).

that the causation runs the other way: Income and productivity growth, by increasing productive capacity and reducing costs, can make a country more competitive in world markets and lead to faster growth of manufactured exports. Also possible is that both export growth and economic growth are simultaneously caused by something else, such as improved macroeconomic policies, more stable political systems, reduced corruption, or increased savings. If one does not fully account for these factors (the omitted variables) it is possible that trade statistically is picking up their influence even if it is not the causal determinant of higher income levels.

Frankel and Romer address these issues by tracing the portion of trade due to geographical characteristics (such as a country's size, its location relative to its trading partners, and whether it is landlocked), which tend to be weakly correlated or uncorrelated with other possible determinants of growth. They show that this geographical component of trade has a large and positive effect on income. The study supports the direction of causality running from trade to growth but acknowledges the result is not highly significant; a finding echoed in subsequent studies. Causality is hard to untangle because, to a large extent, export growth and economic growth probably support each other in a virtuous circle: Exporting countries have greater access to new machinery and technology that support growth, while faster economic growth provides the means to finance investments in the new machinery, technology, and infrastructure that support exports.

The study by Frankel and Romer uses the ratio of exports plus imports to gross domestic product (GDP) as their measure of trade. This is a common way of measuring *trade volumes*. In the survey article by Harrison and Rodríguez-Clare cited earlier, the authors reviewed hundreds of articles, many that relate trade volumes to either income levels or growth rates while adjusting for other possible determinants of growth and employing a wide range of econometric specifications. They conclude that there is a strong correlation between increasing trade shares and country performance, ceteris paribus. This is not a surprising finding. Although trade may not be a panacea, one would be hard pressed to think of any successful low- or middle-income nation where trade has not become a large part of its economy.

Trade *volumes*, however, are not the same as trade *policies*. Trade policies, whether reducing tariffs, managing the exchange rate, or creating EPZs, are the instruments governments have at their disposal. Trade volumes are a consequence of these and other policies. For the policy maker, understanding the impact of alternative trade policies on growth would be of great value. But this is not easy to determine. Harrison and Rodríguez-Clare's exhaustive survey finds no significant correlation between lower tariff levels on final goods and country performance. Does this suggest that trade liberalization is a bad idea? Certainly not.

The problem is what alternative measures of trade policy can tell us. Measuring the links between specific trade policies and growth is complex because a range of policies affect trade and countries often employ instruments of import protection combined with policies for export promotion. Looking at any one policy in isolation from others can provide ambiguous results, as no single policy can fully reflect

a country's range of trade policies. Consider import tariffs: China, Korea, Malaysia, and other countries kept tariffs high on many products, while some other countries that reduced average tariff rates were not rewarded with rapid export growth. But the most successful countries did more than just lower average tariffs. The exact policy combination differed, although they all had the common element of zero tariffs on imported inputs and capital goods for exporters, and they took other steps to reduce costs for exporters and to integrate firms with the global economy. Because the policy mix differed, it is that much harder to measure for research purposes and to prescribe the precise steps that other countries might want to take.[19]

An alternative approach is to examine combinations of trade policies. An often cited example is a study by Jeffrey Sachs and Andrew Warner.[20] They consider the economic growth performance of 79 countries around the world during the period 1970–89 and find that countries with more open policies and less-biased exchange rates grew about 2 percentage points faster than did closed economies. As we have highlighted throughout the text, 2 percentage points added to the growth rate is a huge amount with the potential to rapidly alleviate poverty and transform a society. Can greater openness to trade deliver this much? The Sachs-Warner measure of trade policy considered a country to be "open" if it passed five criteria: (1) its average tariff rate was less than 40 percent, (2) its nontariff barriers (such as, quotas) covered less than 40 percent of imports, (3) the premium on the unofficial parallel market exchange rate did not exceed 20 percent, (4) there were no state monopolies on major exports, and (5) it was not a socialist economy. This is a broad interpretation of what openness entails and goes well beyond trade policy.

It is not surprising that the Sachs and Warner study has been subjected to a lot of scrutiny and debate. One of the key criticisms concerns their index of openness and their conclusions about trade policy. Poor management of the exchange rate (item 3) is more likely a call to improve macroeconomic policy than to liberalize trade. Similarly, reliance on state monopolies (item 4) and the socialist economy criteria (item 5) may have more to do with governance than with trade per se. Several researchers have found that it is precisely these items (items 3, 4, and 5) in the Sachs-Warner openness index that explain most of the difference in growth performance during the 1970s and 1980s. Reducing trade barriers (items 1 and 2) had little independent explanatory power.[21] Today, socialism, state monopolies over exports, and gross

[19]Lant Pritchett, "Measuring Outward Orientation in LDCs: Can It Be Done?" *Journal of Development Economics* 49, no. 2 (1996), 307–35.

[20]Jeffrey Sachs and Andrew Warner, "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity* 1 (1995), 1–118.

[21]The seminal articles in this debate are Ann Harrison and Gordon Hanson, "Who Gains from Trade Reform? Some Remaining Puzzles," *Journal of Development Economics* 59, (1999), 125–54; Francisco Rodríguez and Dani Rodrik, "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence," in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2000* (Cambridge: NBER, 2001). An updated version of the Sachs and Warner approach covering a much longer time period, 1950–98, finds results confirming the original study. See Romain Wacziarg and Karen Welch, "Trade Liberalization and Growth: New Evidence," *World Bank Economic Review* 22 (2008), 187–231.

overvaluation of exchange rates have generally disappeared. The Sachs and Warner findings appear to offer little guidance on how to conduct trade policy.

A skeptical reader might conclude that economists know little about the impact of trade and trade policy on economic growth and that it does not matter what strategy a nation pursues. This is the wrong conclusion to draw. Economists believe that increased trade confers benefits but that the steps necessary to stimulate economic growth and increased trade go beyond simply reducing tariffs and quotas and include strong macroeconomic management (such as a sensible exchange rate policy), steps to strengthen key economic and governance institutions, and policies that more broadly improve the environment for investment and productivity growth. The institutional innovations needed to spark both trade and growth may be different from those prescribed by a pure neoclassical model, and innovations that work in one country may not be easily replicable in others. Different paths to trade reform have been successful in different settings. One approach is not suited to all nations. But even those who are most skeptical about the relationship between trade and growth do not suggest that openness to trade is an ill-advised strategy or that broad trade barriers are conducive to long-run growth.

TRADE REFORMS AND POVERTY ALLEVIATION

Once again it will be useful to distinguish between trade volumes and trade policies and their respective impacts on the poor. We know from the previous section that increases in the share of exports (or of exports plus imports) out of GDP are well correlated with economic growth. In Chapter 6 we presented evidence that growth generally is good for the poor. If average incomes rise, as long as income inequality does not increase by too much, the incomes of the poor will also rise, lifting many above the poverty line. Taken together, this suggests that increasing trade is associated with poverty alleviation. The expansion of exports and increased foreign direct investment has brought a decline in poverty in many countries across all continents. This is most apparent in Asia's success stories of development. Over the past two decades, China, India, Indonesia, and Thailand have experienced rapid growth in exports and a decline in poverty. Simulations conducted in the early 2000s concluded that a global move toward freer trade could lift more than 300 million people out of poverty within 10–20 years through trade's impact on income growth.[22]

The mechanisms by which increased trade can reduce poverty are relatively straightforward. A more outward orientation reduces poverty by increasing the demand for labor. Trade in unskilled labor-intensive products, whether in agriculture or manufacturing, has the potential to create substantial employment opportunities for members of households living below or near the poverty line. Many of

[22]William Cline, *Trade Policy and Global Poverty* (Washington, DC: Center for Global Development, 2004); World Bank, "Market Access and the World's Poor," in *Global Economic Prospects and the Developing Countries 2002: Making Trade Work for the World's Poor* (Washington, DC: World Bank, 2002).

these workers tend to be young women in their late teens and 20s. This is what happened in many of the more outward-oriented economies in East Asia and elsewhere. New job opportunities increased the wage income of poor households and reduced the number of individuals living in poverty. Many of the poor also benefited as consumers. Trade reduces the price of imports, including that of necessities, and can increase the real incomes of poor households. Trade can also raise government revenues and improve social spending directed at the poor.

But trade has other impacts as well. In Chapter 18 we discussed how trade produces both winners and losers. Trade creates new employment opportunities in the expanding export sector but also eliminates jobs as some domestic firms find themselves unable to compete with cheaper imports. Trade can lower the prices of goods; this is beneficial to the consumer of the good but can harm the producer. After the NAFTA agreement between Canada, Mexico, and the United States, corn exports from the United States to Mexico increased, helping poor, often urban households who now faced lower food prices but hurting some small-scale Mexican corn producers who now received lower prices for their crops. Because of these varied effects, the impact of trade, especially of trade reforms, on the poor is varied and complex.

Recent case studies on a number of countries help identify the ways in which trade policy impacts poverty.[23] After Colombia joined the WTO in 1981, the nation's policy makers drastically lowered tariff and nontariff barriers in manufacturing. In 1984, the average tariff on Colombian manufactured good was 50 percent; by 1998 it had fallen to 13 percent. Workers in import-competing sectors, where tariffs tended to be the highest, suddenly faced significant price competition. These workers started to experience rising levels of unemployment and/or of informal employment, falling incomes and increasing poverty. But other Colombians had the opposite experience. Those working in the expanding export sector gained more formal sector jobs, increased incomes, and lowered rates of poverty. Something similar happened in India. After the trade reforms of the early 1990s, the rural poor gained less than the urban poor, and the more dependent a rural district was on trade the smaller the decline in poverty. What the results from Colombia and India demonstrate is that not all poor people are the same and that one set of policies may affect different groups of poor households differently, even within the same country.

The results also seem to contradict the basic theoretical arguments for why trade should help the poor in developing nations. Theories that identify factor endowments as the basis of comparative advantage, discussed in Chapter 18, predict that opening to trade favors a nation's abundant factor. If poor nations are abundant in unskilled labor, poor households should gain from trade as the economy moves to export more labor-intensive goods. This has happened throughout much of East

[23]Ann Harrison, ed., *Globalization and Poverty* (Cambridge: NBER, 2007). See in particular the chapters on Colombia by Pinelopi Goldberg and Nina Pavenik, on India by Petia Topalova, and on Mexico by Gordon Hanson.

Asia. Why are the findings different in Colombia and India? One reason is that factor endowment models assume that labor is mobile between sectors. If workers can easily move out of the contracting import-competing sector and into the expanding export sector, they should experience the gains from trade. What the studies on Colombia and India reveal is that if workers cannot easily relocate, often because of labor regulations that constrain the expansion of new jobs, then the gains from trade may not be transmitted in the ways the factor endowment model predicts.[24] Whether this is a short-run outcome or a long-run one is something not yet understood. Parents may be unable to take advantage of new opportunities, but maybe when their children grow up they will be better able to respond to the new opportunities trade provides.

The studies on Colombia and India do not suggest that trade reform in these economies was a bad idea. What they do suggest is that trade reforms *alone* often cannot be relied on to reduce poverty. Complementary policies have to be pursued. In Colombia and India, policy reforms were needed to improve worker mobility across sectors and regions. In Zambia, trade reforms were not enough to help farmers exploit new export opportunities. They also needed help accessing credit and obtaining extension services to aid them in growing nontraditional crops. Also needed are social safety nets to catch those who are hurt by trade reforms because they are unable to adjust to new market opportunities. In rich and poor settings alike, increased integration with the global economy produces winners and losers. Most economists argue that the gains of globalization far outweigh the losses, but that does not mean everyone gains. Some will be worse off, sometimes requiring safety nets to help them as their nation transitions to a more integrated global economy.

KEY ISSUES ON THE GLOBAL TRADE AGENDA

INCREASED GLOBAL COMPETITION AND THE RISE OF CHINA (AND INDIA)

The benefits for an individual developing country from greater outward orientation may be smaller in the future than those achieved by the original Four East Asian Tigers. When Hong Kong, Korea, Singapore, and Taiwan began to compete on global markets, there was little competition from other developing countries, so their firms were able to expand and multiply quickly. Today, many more developing countries have adopted this strategy, including the population giants China and India, generating concerns

[24]Trade models that assume that factors are not perfectly mobile between sectors are referred to as *specific-factor models* and are an alternative to the Heckscher-Ohlin model discussed in the previous chapter.

about tougher competition and falling prices for export products. As new highly efficient producers enter a market, prices often fall in the short run, as new entrants create overcapacity and drive out higher cost producers. Examples include semiconductors when prices fell in the mid-1990s as world productive capacity grew rapidly and coffee when prices dropped after Vietnam started producing, hurting some traditional coffee exporters such as Ethiopia and Honduras. After the expiration of global agreements to manage world trade in clothing, China's expansion into garment exports put pressure on firms as far away as tiny Lesotho in southern Africa. There is no question that some firms and some nations lose while others gain in a world of globalized trade.

The long-term evidence, however, suggests that world trade can expand very quickly and accommodate many new firms. Since 1950, world exports grew much more quickly than world output, and exports of manufactured products grew most quickly of all. The United States was a huge force in world markets, accounting for 25 percent of world output after World War II, much larger than China today, and there was concern that U.S. firms would dominate all trade. But while the reduction of trade barriers by the United States and its European allies put pressure on some firms as world prices dropped, the changes created many new opportunities for other firms. Similarly, in today's world, as more countries trade and transport and telecommunication costs continue to drop, the opportunity for firms to specialize increases. Global production networks allow firms from many different countries to contribute to the production of one finished good, with each firm specializing in a particular phase of the production process. Many exports from China are assembled and finished products built with components imported from other countries. A computer from China might include a screen made in Malaysia, a circuit board from Singapore, and a keyboard made in the Philippines.

As developing countries open to world markets, they also become consumers for other country's exports, not just competitors. Trade is a two-way process. China and India are rapidly increasing their imports as well as their exports and are becoming two of the largest markets in the world for other countries. China is now one of Indonesia's top five export markets importing oil and gas, timber, fish, and even some electronics products. China recently edged out the United States as Brazil's major trading partner. China is also aggressively establishing trade with nations throughout sub-Saharan Africa and today runs an overall trade deficit (imports exceed exports) with the region.

Still many nations worry that with the expansion of China's trade, and to a lesser extent India's, that there is little room for their exports of labor-intensive goods in the global economy. Low-income economies believe they cannot match the combination of low wages and high productivity in China's factories. There is more than a little truth to such fears. The explosion of Chinese garment exports has driven firms in many nations, previously operating behind the protection afforded by the 1995 global Agreement on Textiles and Clothing, out of business. But this is also a static view of trading opportunities. Wages in China are increasing as the average productivity of

Chinese workers rises. News stories increasingly report labor shortages, rising wages, and growing numbers of strikes by Chinese workers. These are all signs of a growing scarcity of labor, a result fully expected and hoped for as China develops. This also means that China's comparative advantage is changing, creating new opportunities for lower wage economies. Reebok, New Balance, and other manufacturers of running shoes, for example, are expanding in Indonesia in response to rising wages in China.

Some middle-income nations, like Mexico, face similar problems that may be more challenging. Relative to the United States, Mexico's factory endowment is unskilled labor. But Mexican wages are too high relative to those of China, and its relative productivity level not high enough, to compete with China in many goods destined for the U.S. market. Its comparative advantage, in part, lies with its locational advantage. Proximity to U.S. markets means goods from Mexico can get to the United States in a fraction of the three weeks typically needed to ship goods across the Pacific. This especially favors heavier products such as auto parts for which shipping costs are higher. Mexico's gains from trade certainly are limited by China's export drive but comparative advantage still works and the basic rationale for an outward-oriented strategy remains.

DOES OUTWARD ORIENTATION CREATE SWEATSHOPS?

Some critics argue that outward orientation leads to a global sweatshop economy, in which corporations pit workers around the world against each other in a race to the bottom to see who will accept the lowest wages, benefits, and environmental standards. It is understandable how someone looking at the factories that produce shoes and textiles in low-income countries could reach this conclusion. By rich world standards, the wages are very low, often only a few dollars a day or less. Workers labor for long hours in repetitive tasks, often with only limited breaks. In many cases, the conditions are deplorable. This is also true for firms producing for local markets, for agricultural workers, and others.

Low wages are a reflection of the extent of poverty in low-income countries. Wages are set not by what firms pay in other countries but by the productivity of local workers (which establishes the demand for labor) and the wages paid to workers' next-best opportunity (which determines the supply of labor). As we have seen, billions of the world's population live on less than $2 a day. Many toil in backbreaking work in agriculture to try to feed their families, and others work in market stalls, as servants, dock laborers, or in other difficult positions. Many have no job and little or no income at all. A 25-year-old woman machine operator in a textile firm in Nairobi earns about one 13th of what her counterpart earns in Amsterdam or Taipei.[25] This huge gap in pay is not because the Kenyan worker is exploited by her employer or because the Dutch or Taiwanese workers are 13 times more skilled. It is because pay differences across countries reflect differences in the average level of *economywide*

[25]Union Bank of Switzerland, *Prices and Earnings* (Zurich: UBS AG, 2009).

productivity. Workers in Kenya have less physical and human capital to work with and, on average, output per worker is low. Therefore, the market-determined wage is going to be substantially lower than in high-productivity Taiwan or the Netherlands.

For many people, the opportunity to work in a steady factory job for a few dollars a day is a big improvement over other options and can represent an important first step in rising above subsistence living and out of poverty. Without such options, poverty could be worse, rather than better. Efforts to substantially and immediately raise wages to industrialized-country standards could backfire. Firms forced to pay well above the marginal product of labor hire fewer workers, substitute machines for workers, or simply decide to relocate their operations, making workers even worse off. No one advocates that factory jobs with relatively low wages (by global standards) should be seen as the end goal. Rather, it should be seen as the first step on what ideally will be a dynamic path on which job skills, wages, and standards of living can grow steadily over time. The United States, Japan, and the nations in the European Union all went through a phase of very low wages and difficult working conditions during their transition from agrarian economies to industrialized nations. This is not meant to glorify this phase. It suggests only that there are no shortcuts to achieve economic development. Wages rise as the average productivity of the economy rises, the result of productive investments in both capital and people. Success stories in recent years are those nations that grew rapidly and were able to get through the very low wage and difficult working conditions phase quickly.

The image of a race to the bottom suggests that wages steadily *fall* in countries that compete to attract foreign firms that export to world markets. Sometimes wages fall at first for some workers, as uncompetitive protected firms close down and where wages effectively were subsidized by consumers who paid a higher price for protected products. But the typical pattern in countries that shift to outward orientation is not a steady fall in wages; instead, wages rise over time as workers gain new skills and factories become more productive. Evidence suggests that, in countries that attract foreign direct investment, multinational corporations tend to pay higher average wages and have better average working conditions than do domestic firms engaged in the same activities. Nike subcontractors in Indonesia paid nearly three times the average annual minimum wage, and workers in foreign-owned apparel and footwear factories in Vietnam rank in the top 20 percent of the population by household expenditure. In Mexico, export-oriented firms pay upward of 50 percent higher wages than similar nonexporting firms. Pay for workers in EPZs generally is higher than for similar activities outside of zones.[26]

The concern about the race to the bottom goes beyond wages and includes working conditions and labor standards. Advocates of stronger labor standards argue

[26]Drusilla Brown, Alan Deardorff, and Robert Stern, "The Effects of Multinational Production on Wages and Working Conditions in Developing Countries," Working Paper No. 9669, National Bureau of Economic Research, Cambridge, MA, April 2003; Theodore H. Moran, *Beyond Sweatshops: Foreign Direct Investment and Globalization in Developing Countries* (Washington, DC: Brookings Institution Press, 2002), p. 7; Edward Graham, *Fighting the Wrong Enemy: Antiglobal Activists and Multinational Enterprises* (Washington, DC: Institute for International Economics, 2000), chap. 4.

that they are necessary to guard against corporations imposing poor working conditions, such as excessively long hours, short breaks, gender discrimination, and poor ventilation and other health and safety hazards. Opponents argue that the imposition of stricter standards raises producer costs and discourages investment, thereby hurting workers by costing them their jobs. Despite the sometimes polemic debate, there is ample middle ground. Some standards are widely shared. The International Labour Organization (ILO) refers to these as **core standards**: abolishing forced labor; ending discrimination based on gender, race, ethnicity, and religion; eliminating child labor where it is harmful to the child; and permitting freedom of association of workers and the right to collective bargaining. For other improvements in conditions, which often are referred to as **cash standards**, the costs to producers are often small, including providing reasonable work breaks and improving safety conditions (such as unlocking doors, supplying fire extinguishers, and improving ventilation). In other words, some working conditions can be improved without adding significantly to costs and discouraging investment.

There is no question that working conditions are worse in low-income countries than in richer countries, and this is the case not just for firms engaged in trade but for those selling in protected domestic markets and for agriculture. The goal is to improve those conditions over time. Does outward orientation and greater trade make these conditions worse or better? The evidence is not complete, but overall it suggests that working conditions generally improve as economic growth proceeds in outward-oriented countries.[27] This outcome is not always the case. There are situations in which investors threaten to go elsewhere if workers are allowed to unionize or individual workers advocating for better conditions lose their jobs, as is the case with nonexporting firms. But in countries that expanded trade over a sustained period of time, such as Korea, Indonesia, Malaysia, and Mauritius, working conditions improved and are much better than they were in the late 1980s and early 1990s.[28] There is evidence that working conditions are improving across China as factories increasingly compete for workers.

In countries where there is little investment and growth, worker conditions tend to stagnate. Export-oriented firms tend to include a large share of foreign investors, and most studies show that worker conditions tend to be better in foreign-owned factories compared to similar domestically owned factories. This hardly suggests that advocacy efforts are misplaced. Better conditions in foreign-owned firms are partly in response to advocacy: Multinational firms respond to adverse publicity about poor worker conditions and attempt to bring about improvements (Box 19–4). But it is hard to find systematic evidence of a decline in worker standards and a race to the bottom as countries become more open and expand trade.

[27]Kimberley Ann Elliot and Richard B. Freeman, *Can Labor Standards Improve under Globalization?* (Washington, DC: Institute for International Economics, 2003); Brown et al., "The Effects of Multinational Production."

[28]The evolution of wages and working conditions in Korea is discussed in David L. Lindauer et al., *The Strains of Economic Growth: Labor Unrest and Social Dissatisfaction in Korea* (Cambridge, MA: Harvard Institute of International Development, 1997).

BOX 19-4 LABOR ACTIVISTS AND LABOR OUTCOMES IN INDONESIA

Anti-sweatshop campaigns and other forms of labor activism pressure firms in developing nations to improve working conditions, especially firms producing brand name goods for export. No company wants their subcontractors exposed as operating sweatshops because consumers in advanced economies may respond to such negative publicity by buying fewer products. Public relations campaigns targeting Nike, the Gap, and others have been effective in getting these corporations to impose codes of conduct on the factories that produce their footwear and apparel. But are the workers in these factories always helped by such actions? By raising labor costs are jobs lost? How often do firms close down in the face of labor activism? Are such firms foot-loose and able to relocate to countries where there is less activism and attention to labor standards?

Whether labor activists wind up helping or hurting the workers they intend to aid is an empirical question. Ann Harrison and Jason Scorse investigated this question using econometrics for Indonesia during the 1990s. They employ a "difference in difference" approach, which permits the researcher to separate out underlying trends from the behavior of a treatment versus a control group. Harrison and Scorse were interested in how wages and employment changed over time in districts where there was a concentration of firms exporting textiles, footwear, and apparel (TFA). They compare these changes with districts in which there was a concentration of firms producing for the domestic market or engaged in products other than TFA.

Indonesia provided an excellent opportunity to study the impact of labor activists because of Jeff Ballinger, a U.S. labor activist, who in the 1990s exposed the harsh working conditions and poor pay of workers in Indonesia's footwear industry. Ballinger targeted plants that produced running shoes for Nike, reasoning that by drawing attention to Nike, pressure could be brought to bear on other footwear and clothing producers. In a 1992 article in *Harper's* magazine, Ballinger published the pay stub of Sadisah, an Indonesian worker in a Nike subcontractor. In one week she worked 63 hours and earned 76,120 ruppiah ($37.46 at the market exchange rate.) Ballinger noted that it would take Sadisah 44,492 years to earn the $20 million Michael Jordan earned for his endorsement of Nike products. Ballinger's story and subsequent exposés were reported by many U.S. newspapers, magazines, and television networks.

Harrison and Scorse concluded that Ballinger's efforts were successful: "wages increased systematically more for exporting and foreign TFA plants in districts where activists concentrated their efforts." Equally important, they found no adverse total employment effects—that is, employment growth in targeted

TFA districts was not hurt relative to employment growth in non-TFA districts. This somewhat counterintuitive finding is explained, in part, by the very low share of Indonesian labor costs, around 5 percent, of the final retail sales price of a Nike shoe. It is also explained by the very low wages that TFA workers received at the start of the period. Mostly young women worked in these factories, and they received wages well below those of other manufacturing workers, making it easier for Nike and its contractors to absorb the increases in their pay. The evidence on labor activism in Indonesia appears to be win–win: Low-paid workers received higher wages and did not suffer any job loss. But the authors of the study are quick to point out that these results may be temporary, not permanent, because Nike shifted some of its outsourcing to other low-cost producers, including Cambodia and Vietnam. High-end running shoes are also a somewhat unique product and labor activism on goods with less brand name identity and smaller profit margins may not have the same effects.

Sources: Ann Harrison and Jason Scorse, "Improving the Conditions of Workers?: Minimum Wage Legislation and Anti-Sweatshop Activism," *California Management Review* 48, no. 2 (2006); Ann Harrison and Jason Scorse, "Multinationals and Anti-Sweatshop Activism," *American Economic Review* 100, no. 1 (March 2010); Kimberly Ann Elliott and Richard B. Freeman, "White Hats or Don Quixotes?: Human Rights Vigilantes in the Global Economy," in Richard Freeman, Joni Hersch and Lawrence Mishel, eds., *Emerging Labor Market Institutions for the Twenty-First Century* (Chicago: University of Chicago Press, 2004).

EXPANDING MARKET ACCESS

Implementing a more-open trade strategy is successful only if there are markets in which developing country exporters can sell their products, and the most important markets remain those in the European Union, Japan, and the United States. Generally speaking, these countries have low trade barriers, with relatively low average tariffs and quotas by world standards. But their largest trade barriers and greatest protection are precisely on the products in which low-income countries have a comparative advantage: textiles, apparel, and agriculture. Although the world's richest countries advocate for free markets and recommend that poor countries shift toward greater outward orientation, they also maintain trade barriers that impede that process. A key question is whether industrialized countries will react to their own macroeconomic problems and to the greater competition on world markets from China, India, and many other developing countries by maintaining protectionist measures or by imposing new ones. Or will the advanced economies allow even greater market access to goods produced in developing countries.

Limits on market access are reflected in tariff revenues. In most years, the United States collects more tariff revenue on exports from poor countries than rich

countries. Consider U.S. imports from Bangladesh and France. In 2009, Bangladesh paid $563 million in tariffs to the United States, whereas France paid half as much, $285 million. This alone is regressive, with the poorer nation paying more than the richer one. But the implicit tax rate of these tariffs is even more striking. Bangladesh exported only $3.7 billion worth of goods to the United States, including underwear and towels; France exported nearly *10 times* as much, $34 billion, including wine and airplanes. The implied average tariff on French products is under 1 percent; on Bangladesh over 15 percent.[29]

Tariff peaks refer to relatively high tariffs amid generally low tariff levels. For industrialized countries, tariffs of 15 percent and above are generally recognized as tariff peaks and often are applied to imports from developing nations. The United States charges a 32 percent tariff on acrylic sweaters, a Bangladeshi export, but no tariff on artwork imported from France. **Tariff escalation** refers to having no or low import duties on raw materials and higher duties on the finished goods that use these raw materials. For example, the United States does not charge a tariff on cocoa beans, but charges more than 25 cents a pound on imports of certain kinds of chocolate. Tariff escalation makes it harder for firms in developing countries to compete in the markets for some finished goods. Such mercantilist practices were denounced by Adam Smith over 200 years ago but remain features of developed nation trade policies.

High-income nations have worked to reduce protection on most manufactured goods, including imports of textiles and apparel. For decades, textile and clothing imports were restricted via a complex system of quotas and tariffs under an agreement imposed by the industrialized countries called the Multi-Fiber Agreement (MFA), later amended as the Agreement on Textiles and Clothing (ATC). Under these arrangements, each industrialized country allowed only a certain amount of imports of textiles and clothing from each developing country. Some exporters, like China, Indonesia, and Thailand, easily filled their quota and were not allowed to export more. This helped other developing countries, as investors began to look at countries like Bangladesh, Lesotho, Mauritius, and Sri Lanka to locate production facilities and export to the industrialized countries.[30] But, because of the quotas, the rich countries imported far less overall than they would have with open markets, which drove up domestic prices, hurting domestic consumers while protecting domestic textile and apparel makers (and their workers) at the expense of firms and workers in developing countries. Most of these multilateral arrangements expired during the

[29]Edward Gresser, "U.S. Trade Preference Programs: Options for Reform," Senate Committee on Finance, March 9, 2010.

[30]Cambodia is another country that benefited from the quota system. But it adopted a unique position, in part, because of pressure from the U.S. government. Cambodian garment exporters agreed to follow strict guidelines on worker rights, hoping that the United States and other developed nation consumers would be willing to pay more for garments certified to have been produced in factories complying with a number of nonwage labor standards. With the end of the system of managed trade in garments, it remains uncertain whether consumers in industrialized countries will pay more for Cambodian products or simply shop for the lowest-priced goods available.

past decade but national tariff and nontariff barriers on textiles and apparel imports remain, leaving firms in developing countries at a significant disadvantage.

Agriculture is another heavily protected sector, in rich and poor nations alike, and one that probably raises more controversy than any other economic sector. Canada, the EU, Japan and the United States provide substantial protection to their agricultural producers through both tariff and nontariff barriers and direct subsidies to producers. As shown in Table 19–1, the estimated magnitude of protection is quite large, ranging from a tariff equivalent of about 20 percent in the United States to over 80 percent in Japan. According to this analysis, in Canada, the EU, and Japan high trade barriers on agriculture are much more significant than production subsidies; the opposite holds in the United States.

A wide range of products is protected, including cotton, dairy products, maize, peanuts, rice, soybeans, sugar, and wheat. Sugar is among the most distorted markets in the world, with the industrialized countries subsidizing domestic producers and imposing strict quotas to limit imports. In recent years, U.S. consumers have paid as much as three times the world price of sugar because of U.S. sugar quotas, subsequently affecting the price of everything from candy to processed foods to soft drinks produced in the United States. But it is not only U.S. consumers who are supporting domestic sugar producers, growers of sugar cane in the Caribbean, Central America, and elsewhere are deprived from earning more by the limits placed on exports to the United States and other developed nation markets.

Trade protection provides farmers in industrialized countries with a significant advantage over potential competitors, and weakens the incentives for increased production in developing countries. Production and export subsidies also encourage *increased production* in the industrialized countries, which adds to world supply putting downward pressure on world prices. One World Bank estimate suggests that the protection by the developed economies depresses world rice prices by 33 to 50 percent, and for sugar and dairy products, by 20 to 40 percent.[31] When farmers in developing countries receive lower prices, they produce less and earn less for the amount they produce, reducing their total income.

TABLE 19–1 Overall Protection in Agriculture (percent tariff equivalent)

TYPE OF PROTECTION	UNITED STATES	CANADA	EUROPEAN UNION	JAPAN
Tariffs	8.8	30.4	32.6	76.4
Subsidies	10.2	16.8	10.4	3.2
Total	19.9	52.3	46.4	82.1

Source: William Cline, *Trade Policy and Global Poverty* (Washington, DC: Center for Global Development, 2004).

[31]World Bank, *Global Economic Prospects and the Developing Countries 2002.*

Large agricultural interests in the industrialized countries lobby hard to maintain their protection. Rural states and provinces often have considerable political power, far greater than their share of total population might warrant. The outcome is continued protection of agriculture even though trade barriers and production subsidies are costly to consumers and taxpayers. In 2006, according to a study by the Organization for Economic Co-Operation and Development (OECD), government policies in rich countries that supported agriculture transferred close to $300 billion worth of income to farmers, amounting to almost one third of all farm earnings. In one often-cited calculation, the average Japanese cow gets $7.50 a day from various government subsidies and other forms of protection, whereas a cow in Europe receives $2.50 a day.[32] This is at a time when over 1 billion people in the developing world live on less than $1 a day.

The combined impact of industrial country tariffs, quotas, and subsidies is significant. The elimination of industrialized country barriers would likely add more to the GDP of poor countries than they receive in foreign aid. Two studies find that the elimination of these barriers would lift an additional 300 million or more people out of poverty.[33] It is unlikely that the industrialized countries will reduce these barriers on their own. The hope has long been that they would do so in the context of multilateral trade negotiations.

MULTILATERAL TRADE NEGOTIATIONS AND THE WTO

With the outbreak of World War I, a long epoch of globalization came to an end as the industrialized countries began to erect high trade barriers, a trend reinforced by the onset of the Great Depression. With the end of World War II, this pattern began to change. Industrialized country leaders began to reduce tariffs, and they looked for ways to accelerate and consolidate the process. The result was a shift toward **multilateral trade negotiations**, involving many nations simultaneously negotiating reductions in tariffs. The idea was that, with a large international effort in which participants would pledge to reduce their tariffs if other countries did the same, each individual country would be better able to overcome narrow interest groups at home that opposed trade liberalization. These discussions led to a proposal in 1948 to establish the International Trade Organization (ITO) as a sister organization to the International Monetary Fund (IMF) and World Bank. Although the ITO was never established, in large part because of the opposition of the U.S. Congress, multilateral trade negotiations expanded and flourished in the ensuing decades through a less-formal institution known as the **General Agreement on Tariffs and Trade (GATT)**. Between 1947 and 1994, eight rounds of multilateral trade negotiations took place

[32]Nicholas Stern, "Dynamic Development: Innovation and Inclusion," Munich Lectures in Economics, Center for Economic Studies, Ludwig Maximilian University, November 19, 2002.

[33]World Bank, *Global Economic Prospects and the Developing Countries 2002*; Cline, *Trade Policy and Global Poverty*.

through the GATT, covering mostly tariffs but later including nontariff barriers and other trade related issues.

The Uruguay Round of trade negotiations (so named because its initial meeting took place in Punta del Este, Uruguay, in September 1986) was completed in 1994. It was by far the largest trade negotiation ever held, with 123 countries taking part. Developing countries had high hopes at the outset of the Uruguay Round because it was the first in which a large number of them were allowed to participate and it was the first in which industrialized countries, with great reluctance, agreed to include agricultural subsides. Discussions led to what later was called the *grand bargain* between rich countries and low- and middle-income ones. This approach represented a significant shift from the traditional GATT approach on reciprocity. Instead of each country agreeing to open its markets on certain goods if the others did the same, the Uruguay Round featured an implicit deal in which the richer countries agreed to certain steps in some areas in return for the developing countries agreeing to steps in other areas, many of which went way beyond traditional border barriers.

The industrialized countries promised (1) significant reduction in tariffs on manufactured goods, (2) the end of the Multi-Fiber Agreement that had restricted trade in textiles and garments, and (3) reductions in agricultural protection with a commitment to even larger reductions in the next round. For their part, developing countries promised (1) reductions in their own tariffs; (2) agreement on new rules on investment, trade in services, and trade-related intellectual property rights (TRIPs; intended to prevent the use of patented material and the production of generic copycat products without permission); and (3) support of a new organization, the **World Trade Organization (WTO)**, to replace the GATT. The WTO, established on January 1, 1995, would serve as a central institution for global trade negotiations aimed at establishing a system of rules for fair, open, and undistorted competition and as a forum for settling disputes among members.

Despite increased access in some industrialized country markets, many developing countries eventually were deeply disappointed in the Uruguay Round. There were three broad concerns. First, the industrialized countries' pledge to reduce subsidies for agriculture generally did not materialize. Second, the new agreements on investment, services, and TRIPs were much more complicated than simply reducing tariffs or quotas, and as a result, the developing countries had to build domestic institutional and legal expertise, which required significant time and money. These changes were a large burden on many low-income countries that had a scarcity of highly skilled legal and administrative expertise. The industrialized countries promised to provide technical assistance to help with the transition but often did not fully follow through. Third, the TRIPs agreement resulted in developing countries paying higher prices for medicines and pharmaceutical products covered by patents. For many goods and services, protecting property rights and rewarding those that invest in research and development makes sense, but paying high market prices and preventing the sale of cheaper generic brands became a particularly contentious issue

around medicines, particularly those used for treating HIV/AIDS. Only 10 years later was there the beginnings of serious discussion about how to make these drugs more affordable for low-income countries, as discussed in Chapter 9.

For many, the Uruguay Round came to be considered a bum deal rather than a grand bargain,[34] and the unhappiness contributed to the raucous demonstrations against the WTO in Seattle in 1999 and at other subsequent WTO meetings. A new round of trade negotiations under the WTO began in 2001. Called the Doha Round because the first meetings were held in Doha, Qatar, they also were referred to as "the development round" intended to serve the needs of the developing nations. Negotiations were difficult from the start and collapsed in 2008, although attempts to restart the talks continue.

The Doha Round began with an ambitious agenda, not uncommon in any round of trade negotiations. Agricultural trade protection and subsidies once again were on the table and remained one of the major reasons for the failure to reach an agreement. Under discussion was an opening up of international trade in services, everything from accounting to banking to shipping—activities that often are highly regulated and closed to foreign firms. Developed nations are particularly interested in liberalizing trade in these sectors where they often have comparative advantage. Nonagricultural market access (NAMA), covering manufactures and some primary goods, was intended to deal with problems of tariff peaks and escalation. Rich nations also wanted more discussion of trade facilitation, in essence, improved customs practices, which are an ongoing source of corruption and delay in many developing nations. More transparency in government procurement and new rules on competition policy, a form of global antitrust policy, were also part of the list of items to be discussed.

Many high-level meetings were held not only in Doha but in Cancun, Hong Kong, and often in Geneva, home to the WTO. The agenda progressively narrowed, and while progress was made, including on licenses for cheaper generic drugs, most meetings ended badly without agreement on major issues. The collapse in 2008 had much to do with where the trade round started: agricultural protection. There were divisions among the developing nations. Big agricultural exporters like Brazil had different interests than India, with its large and still relatively protected rural sector. But the real divide remained between rich nations and poor ones. The final straw was a dispute among the United States, China, and India. Developing nations rightly complain about the production and export subsidies farmers in rich nations receive, but developing nations, including China and India, also protect their farmers, relying mostly on trade barriers. WTO rules permit developing nations to impose tariffs, in this case on agricultural goods to protect their farmers, in the event of an import surge or sharp fall in crop prices. Disagreement over the use of this **special safeguard mechanism** could not be resolved. The United States wanted the trigger for when the

[34]Sylvia Ostry, "Why Has Globalization Become a Bad Word?," the Alcoa-Intalco Works Distinguished Lecture, Western Washington University, October 25, 2001.

safeguard kicked in to be set high; China and India wanted it to be set low. More fundamental, the Doha Round collapsed because many of the developing nations felt it was time for the rich nations to make large concessions, especially in agriculture, and for the poor nations to give little in return. To many economists this was a disappointing outcome because developing nations stood to gain a lot from dismantling their own trade barriers.

Earlier trade rounds collapsed only to start up again. This may happen with the Doha Round too. In the meantime, trade negotiations have not stood still. Nations have turned away from multilateral talks and entered into bilateral and regional agreements. In 2001, there were only 49 such agreements in place; by 2009 there were 167. The disadvantage of such agreements is that they do not permit the forces of *global* comparative advantage to work and can promote inefficient firms protected by a bilateral or regional agreement. Such agreements also create a confusing mix of rules and administrative requirements firms must comply with because nations sign different agreements with different trading partners. This diverts the attention of scarce managerial capacity and can impede productive investment. Bilateral and regional deals also take the pressure off of governments to compromise on multilateral agreements, like the Doha Round. To their credit, however, regional trade pacts may be better than no trade pact at all. Despite the failures to date in concluding the Doha Round, the world appears to be moving in the direction of greater market access and expanded trade, but there is no guarantee that this trend will continue.

TEMPORARY MIGRATION: ANOTHER DIMENSION OF INTERNATIONAL TRADE

The General Agreement on Trade in Services (GATS) establishes rules for WTO member nations on trade involving services such as banking, telecommunications, and tourism. Mode IV of the GATS refers to the temporary movement of people from one country to another to perform a service, which in principle could involve an actor, accountant, construction worker, or many others. The abiding principle is nondiscrimination: WTO member nations do not restrict trade, in this case of a service, based on country of origin. What is unique about Mode IV is that it involves not the movement of goods but of people, and controlling immigration is probably where nations are most protectionist. The freer movement of people across national borders is not on the agenda of most nations, especially the developed ones.

Agreements on Mode IV are highly limited but raise a critical issue: What role should increased immigration between rich and poor nations play in promoting economic development? The argument is a compelling one. There is no easier way of improving the lives of poor people than to permit them to migrate to a developed nation where wages are higher and employment opportunities better. It is far easier for a poor person to migrate to a rich nation than it is to bring economic development to where the poor person lives.

Harvard economist Lant Pritchett argues that there is no tool in the arsenal of economic development that has more potential to alleviate poverty than a significant increase in the amount of migration between developing and developed nations. The gains from further trade liberalization, or from increases in foreign aid, pale in comparison to those of even modest increases in the flow of labor between nations. One set of estimates suggests that complete trade liberalization would increase developing nation GDP by around $100 billion annually; an increase in temporary migration equal to 3 percent of the labor force of the OECD nations (about 18 million people) would raise the welfare of those moving by much more: $170 billion per year. Pritchett offers another compelling calculation, comparing the benefits of microfinance to temporary migration. The average annual microloan in Bangladesh is about $360 and the average lender receives his or her first loan at age 23. If one further assumes a generous 18 percent return to such loans, a lifetime of such borrowing yields about $700 in net present value.[35] If a low-skilled Bangladeshi man were permitted to work in the United States, he would earn this amount in only *four weeks*! Imagine the benefits if he were able to work one or more years. Think of the remittances thousands of such workers would send back home.

The benefits of migration should be obvious. The wage gap between workers in rich and poor nations is enormous. Pritchett provides a conservative estimate: Low-skilled workers across 42 low- and middle-income nations working in the United States would, on average, increase their annual earnings by $13,000 (purchasing power parity; PPP). They would be employed in activities in which many immigrants, both legal and illegal, already are employed, as laborers in construction and on farms, kitchen workers in restaurants, gardeners, providers of daycare or care for the elderly, and in other jobs that many natives often find too "dirty, dangerous, or difficult." A focus on increased global labor mobility among relatively unskilled workers also deflects concerns about brain drain and its potential impact on the migrant's country of origin.[36]

The benefits of permitting more temporary migration are clear but so is the opposition. Some of this opposition can be muted by calling for an increase in *temporary* not permanent migration. Temporary migrants would not be on a path to citizenship with all the implications that entails. They would be limited mostly to economic opportunities. Citizens of rich nations still worry that immigrants will lower native wages and take away native jobs; engage in criminal activity; and be a

[35]The calculation is from Lant Pritchett, "The Cliff at the Border," in R. Kanbur and M. Spence, eds., *Equity and Growth in a Globalizing World* (Washington, DC: World Bank, 2010). Pritchett fully develops the economic and moral argument for greater global labor mobility in *Let Their People Come: Breaking the Deadlock in International Labor Mobility* (Washington, DC: Center for Global Development, 2006). Estimates of the potential benefits of temporary migration are presented in World Bank, "Labor Mobility and the WTO: Liberalizing Temporary Movement," in *Global Economic Prospects 2004: Realizing the Development Promise of the Doha Agenda* (Washington, DC: World Bank, 2004).

[36]Brain drain itself is a controversial idea. The problem of brain drain is disputed in a case study of health workers from West Africa in Michael Clemens, "Do Visas Kill?: Health Effects of African Health Professional Emigration," Center for Global Development Working Paper 114, Washington, DC, 2007.

burden to taxpayers, contributing little in taxes while using public services such as schools, hospital emergency rooms, and state-provided welfare payments. There is an element of truth in each of these concerns but careful analysis tends to show that these costs, if they exist at all, tend to be small, especially relative to the huge gain in the well-being of the immigrant himself or herself. Unskilled immigrants tend to depress the wages of unskilled natives by only a few percentage points and do not increase the unemployment of natives with similar skills. Immigrants are less likely to commit crimes because the costs are greater; they face not only jail sentences but also deportation. Immigrants pay taxes and tend not to be a net fiscal drain in aggregate terms although they can be on local communities. Much of the opposition to increased immigration is based on misperceptions of its costs to natives and on matters of culture. Throughout the world, in rich and poor nations alike, people often have a visceral and negative response to those who are foreign. Overcoming this response, even in the name of reducing global poverty, is a formidable challenge.

SUMMARY

- There exists a continuum of approaches toward international trade that nations can adopt. At one end is autarky or economic isolation; at the other is free trade. In practice, most nations fall somewhere along this continuum, exercising some degree of inward-looking versus outward-oriented policies.
- Almost all countries have used import substitution at various times. Implementing this approach often meant using trade barriers, including tariffs and quotas, and overvalued currencies. Such interventions can prove costly to an economy. In theory, IS can allow domestic firms to learn production techniques, improve efficiency, and eventually compete on world markets. However, despite some exceptions, this strategy often leads to weak technology, low efficiency, and slower growth.
- Many developing countries began to shift to more outward-oriented trade strategies in the 1980s. Outward orientation allows firms to become more specialized, sell to larger global markets, and import leading technologies. Most countries that follow this strategy begin with labor-intensive manufacturing and agriculture, with the goal of moving to more-sophisticated products over time. Concerns that this strategy leads to a race to the bottom in wages and labor standards are generally not supported by the evidence.
- Outward orientation often is associated with policies that favor exports. This might include production subsidies such as access to cheap credit for exporters, undervalued exchange rates, or EPZs. Nations like China and Korea

that have pursued outward-looking strategies have also protected domestic firms from import competition, often conditional on export success.

- A lively debate remains over the use of both trade and industrial policy to foster rapid economic growth. Full liberalization of markets does not characterize the experience of some of the world's most successful economies, whereas many attempts at trade liberalization have delivered less than expected. At the same time, it is easy to point to both successes and failures when governments intervene to pick firms or sectors to support. Some partnership between the public and private sector is probably needed to alleviate the binding constraints on economic growth.
- Evidence suggests a strong positive relationship between trade and growth. Countries with larger amounts of trade tend to record faster growth. But the direction of causality is not certain. The evidence is even less clear relating trade policies and growth. This, in part, is due to difficulty in measuring trade policy. Some critics suggest that trade may be less important than other factors, such as strong institutions and an overall climate that supports investments, in explaining rapid growth. But few suggest that a lack of openness is conducive to growth.
- Economies that have experienced rapid growth in exports, especially of labor-intensive products, have witnessed decreases in poverty. But trade liberalization alone may have a differential impact on the poor. As import prices fall and export prices rise, some of the poor may be hurt by a more open economy, often because their mobility from contracting to expanding sectors of the economy is impeded.
- Many issues remain on the global trade agenda. Some worry about the impact of China and India, the two most populous nations in the world, on the opportunities for other economies. Most evidence suggests that there is room in the global economy for all nations.
- Protectionist policies in the industrialized countries continue to impede exports from developing countries, especially in textiles, apparel, and agriculture. It remains to be seen whether the industrialized countries will maintain these barriers or even erect new ones as competition from developing countries increases or whether multilateral negotiations will lead to reductions in these barriers and increased opportunities for developing countries.
- While multilateral negotiations surrounding trade issues continue, including the ill-fated Doha Round, nations are seeking bilateral and regional agreements to increase their access to markets. One dimension of trade negotiations that has received relatively little attention is to increase the movement of people across borders. Some economists believe that an increase in temporary migration between rich and poor nations could be one of the most effective instruments available for addressing global poverty.



Sustainable Development

The path to development followed by today's advanced economies was fueled by technologies intensive in their use of energy and natural resources and prodigious in their production of environmental pollutants. Is this the same path today's poor countries must follow? What are the potential conflicts between development and the environment? According to scientist Jared Diamond, the potential consequences of such a conflict include nothing short of societal collapse.[1] Citing both ancient and recent examples, Diamond argues that unintended ecological suicide ("*ecocide*") was a primary cause of the demise of major civilizations throughout history, ranging from the Maya to the Vikings to modern-day states like Somalia and Rwanda. Diamond elaborates,

> The processes through which past societies have undermined themselves by damaging their environments fall into eight categories, whose relative importance differs from case to case: deforestation and habitat destruction, soil problems (erosion, salinization, and soil fertility losses), water management problems, overhunting, overfishing, effects of introduced species on native species, human population growth, and increased per capita impact of people. . . . The environmental problems facing us today include the same eight that undermined past societies, plus four new ones: human-caused climate change, build up of

[1]Jared Diamond, *Collapse: How Societies Choose to Fail or Survive* (London: Penguin Books, 2005). Diamond defines *collapse* as "a drastic decrease in human population size and/or political/economic/social complexity, over a considerable area, for an extended time" (p. 3).

toxic chemicals in the environment, energy shortages, and full human utilization of the Earth's photosynthetic capacity.[2]

Concerns about the Earth's ability to support continued human development date back at least to the eighteenth century. In 1798, British economist and demographer Thomas Malthus, in his famous *Essay on the Principle of Population*, predicted that continued population growth would bring the world to disaster. Malthus based his analysis on two premises: "That food is necessary to the existence of man," to which Malthus applied the assumption that the production of food would grow arithmetically, and "That the passion between the sexes is necessary and will likely remain nearly in its present state," from which Malthus extrapolated that population would grow geometrically.[3] He concluded "The power of population is so superior to the power of the Earth to produce subsistence for man, that premature death must in some shape or other visit the human race."

Some recent thinking has echoed Malthus. In 1972, the Club of Rome published *The Limits to Growth*, a report by a team of analysts based at the Massachusetts Institute of Technology. This study predicts that "If the present growth trends in world population, industrialization, pollution, food production, and resource depletion continue unchanged, the limits to growth on this planet will be reached sometime within the next 100 years." Among other projections, the report suggests that, based on then-current reserves, the world would exhaust its supply of oil by 1992.

While these and other dire predictions have so far proved false, the question of whether today's developing countries can (or should) follow a historical development path similar to that followed by today's advanced countries, with potentially serious environmental impacts, remains valid. The depletion of both renewable and nonrenewable resources, especially fossil fuels, and the threat of environmental damage pose challenges on a scale unimaginable in the time of Malthus. Concern over the sustainability of economic development and growth has expanded to emphasize global climate change. Scientists point to increasing atmospheric concentrations of **greenhouse gases** such as carbon dioxide (CO_2), which may dangerously raise global surface temperatures by trapping solar radiation in the earth's atmosphere. As we discuss in the final section of this chapter, the potential negative consequences of global warming for development are immense. Various forecasts indicate potential reductions of grain yields of up to 50 percent in Africa, extinction of valuable plant and animal species, increased disease and water stress, and the displacement of millions of poor residents of low-lying coastal zones as the melting polar ice caps cause sea levels to rise. There are even widely expressed fears that the entire nation of the Maldives, a

[2]Diamond, *Collapse: How Societies Choose to Fail or Survive*, pp. 6–7. By "photosynthetic capacity," Diamond is referring to the limited capacity of Earth to support crop growth. He cites calculations that by the late 1980s, humankind had either used, wasted, or diverted about half of the Earth's capacity to produce crops.

[3]An arithmetic progression is 1, 2, 3, 4, 5, 6; a geometric progression is 1, 2, 4, 8, 16, 32.

collection of 1,200 coral islands in the Indian Ocean, 80 percent of which rise no more than one meter above sea level, could simply disappear by the year 2100.

WILL ECONOMIC GROWTH SAVE OR DESTROY THE ENVIRONMENT?

The relationship between economic growth and the environment is complex and dynamic. One important mediating variable is technology. When Malthus foretold of "premature death" for the human race, he excluded the possibility that technical change in agriculture would boost productivity, permitting food supplies to outpace population growth.[4] Technical change also affects the efficiency with which we use renewable and nonrenewable resources as well as the level of environmental damage caused. Increased automobile fuel efficiency and the more recent development of electric cars are only two among many examples of how technical change can help reduce environmental impacts. Our assumptions about the role and pace of technical change are critical in shaping our conclusions about whether economic growth is good or bad for the environment.

The relationship between levels of pollution and levels of income is generally thought to follow an inverted-U shape, with pollution rising as income increases from low levels but falling once income passes some intermediate level. Because this inverted-U pattern is similar to the relationship between inequality and national income famously posited by economist Simon Kuznets (introduced in Chapter 6), this empirical relationship between pollution and national income is called the **environmental Kuznets curve**. This relationship is depicted in Figure 20–1 as the "conventional EKC." Economist Susmita Dasgupta and colleagues describe the basic intuition underlying this relationship,

> In the first stage of industrialization, pollution in the . . . world grows rapidly because people are more interested in jobs and income than clean air and water, communities are too poor to pay for abatement, and environmental regulation is correspondingly weak. The balance shifts as income rises. Leading industrial sectors become cleaner, people value the environment more highly, and regulatory institutions become more effective.[5]

Statistical estimations of this relationship have suggested that pollution peaks at levels of income per capita between $5,000 and $8,000, and declines as income grows above that range.

[4]The Club of Rome's *Limits to Growth* (New York: Universe Books, 1972) study recognized the potential impact of technical change but assumed that its pace would be insufficient to overcome ensuing environmental constraints.

[5]Susmita Dasgupta, Benoit Laplante, Hua Wang, and David Wheeler, "Confronting the Environmental Kuznets Curve," *Journal of Economic Perspectives* 16, no. 1 (winter 2002), 147.

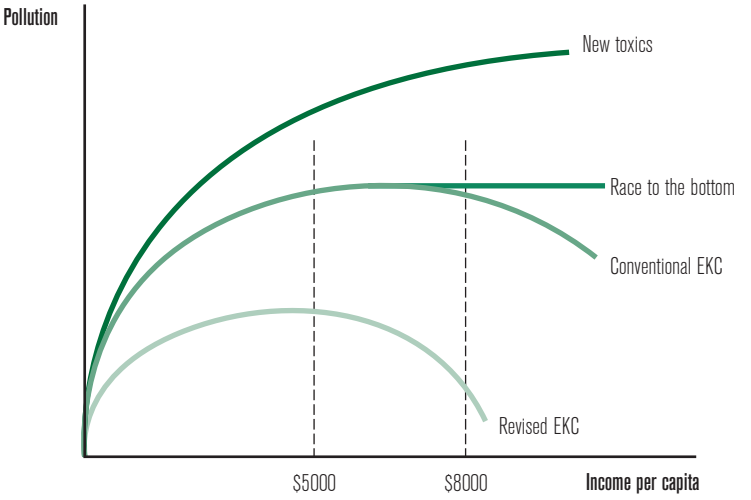


FIGURE 20–1 Potential Paths for the Environmental Kuznets Curve (EKC)

Source: "Confronting the Environmental Kuznets Curve," by Smita Dasgupta, Benoit Laplante, Hua Wang, and David Wheeler. *Journal of Economic Perspectives*, Vol. 16, No. 1, Winter 2002: 147–168, p. 148. Reprinted by permission of the American Economic Association.

Dasgupta notes, however, that estimates of the conventional environmental Kuznets curve have been criticized as misleading snapshots of a dynamic process. If, for example, free trade and globalization create a "race to the bottom", as countries compete by lowering their environmental standards, then the curve may simply flatten out (as depicted in Figure 20–1). Or if income growth and industrialization merely create new environmental hazards, it's possible that the environmental Kuznets curve could rise over time (the "new toxics" path). Conversely, Dasgupta and colleagues cite recent and more optimistic evidence that the environmental Kuznets curve has been shifting downward as growth becomes less polluting and the turning point occurs at increasingly lower levels of income (the "revised EKC"). Thus the relationship between pollution and economic growth remains a subject of debate, one with high stakes, given that current income levels in most of the developing world remain far below the turning point of even the "revised" EKC.

Moreover, we cannot simply infer that increasing levels of income *cause* pollution to first rise and then fall. Alternative explanations for the inverted-U pattern include the natural progression from clean agrarian to polluting industrial to clean service sector domination of the economy, the beneficial effects of greater institutional development and better public policies in advanced economies, and advances due to technical change. It has also been suggested that wealthier countries have reduced their pollution levels by exporting their pollution-intensive production activities to developing countries. If true, this would call into question the ability of today's poor countries to follow the inverted-U pattern, suggesting a path similar to the "race to the bottom" or even the "new toxics." If it were simply the case that economic growth

automatically reduced pollution, we would have little else to worry about in the long run, and policy makers could focus exclusively on promoting growth. This is not the case. The fact that earlier predictions of environmental catastrophe proved overly dire is no guarantee against future environmental catastrophes.

Diamond (along with many economists), skeptical about the environmental Kuznets curve, poses this choice more starkly, asserting,

> because we are rapidly advancing along this non-sustainable course, the world's environmental problems *will* get resolved, in one way or another, within the lifetimes of the children and young adults alive today. The only question is whether they will become resolved in pleasant ways of our own choice, or in unpleasant ways not of our choice, such as warfare, genocide, starvation, disease epidemics, and collapse of societies.[6]

The first step in assessing the sustainability of development lies in defining what we mean by *sustainability* and thinking about how to measure it. How will we recognize sustainable development when we see it? What is it in particular that is to be sustained?

CONCEPT AND MEASUREMENT OF SUSTAINABLE DEVELOPMENT

Among the myriad academic and popular definitions of sustainable development, the most widely cited is that of the World Commission on Environment and Development (commonly known as the Brundtland Commission, 1987), which defined it as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."[7] While highlighting the notion of intergenerational equity, the Brundtland Commission's concept of sustainable development remains vague, and in particular fails to define "needs." For instance, "needs" might be interpreted to require the maintenance of only a bare minimum standard of living, though most of us might hope for more than such a low standard for our children. The Brundtland Commission's definition is also vague with respect to the question of what it means for future generations to meet their needs. This concept could imply that the well-being of future generations remains above some (undefined) minimum standard, that future well-being be equal to present well-being, or that each generation's well-being be equal to or greater than the well-being of the previous generation. Recent discussions of sustainable development by economists

[6]Diamond, *Collapse: How Societies Choose to Fail or Survive,* 498.

[7]World Commission on Environment and Development. *Our Common Future* (Oxford: Oxford University Press, 1987), p. 43 (also known as the Brundtland report).

have tended to define a sustainable path as one along which intergenerational well-being does not decline.

Operationalizing the standard that the well-being of future generations not decline requires a clear definition of well-being and some means of measuring it. For instance, in Chapter 2 we addressed the question of happiness as a measure of well-being, with the paradoxical finding that happiness is not monotonically tied to income. In practice, economists often use income as a proxy for the level of development. For the purpose of measuring the potential well-being of future generations, however, we need to expand that concept to equate well-being with the economy's **wealth**. Economists have shown that the only way for intergenerational well-being not to decline is for per capita wealth not to decline. Non-declining wealth *per capita* requires that the growth rate of wealth be greater than or equal to the growth rate of population.[8] But what do we mean by *wealth*?

This concept of sustainable development requires a broad concept of wealth. Economics Nobel Laureate Kenneth Arrow and colleagues define wealth as "the social worth of an economy's entire productive base. Because the productive base consists of the entire range of factors that determine intergenerational well-being, we . . . refer to wealth as *comprehensive wealth*."[9] **Comprehensive wealth** (also referred to as **total wealth**) is made up of **produced capital** (buildings, roads, machinery and equipment), **natural capital** (minerals and fossil fuels, forests, agricultural land, and protected areas), and **intangible capital** (a broad category that includes human, social, and institutional capital).[10] Comprehensive wealth differs from the more familiar concept of gross domestic product (GDP) in a critical way that highlights the concept of sustainability: GDP includes, besides the production of goods and services, the value of asset liquidation as part of national output. When Chile or Zambia mine and export copper, the depletion of this natural resource counts toward each nation's GDP, but growth based on the depletion of natural resources may not be sustainable. This type of resource use diminishes comprehensive wealth.

Measuring the different categories of capital is a challenge. The value of each type of capital is a function of its price, yet we observe prices only for goods and services that are traded in markets. This is a special problem in the case of natural capital for which markets are often missing. Water is an example of a vital and limited natural resource, the majority of which is used as an input for agriculture. Often there

[8]The growth rate of any ratio is the growth rate of the numerator minus the growth rate of the denominator. Wealth per capita is the ratio (wealth to population). Thus the growth rate of wealth per capita is the growth rate of wealth minus the growth rate of population. If the population growth rate exceeds the growth rate of wealth, then wealth per capita declines.

[9]Kenneth J. Arrow, Partha Dasgupta, Lawrence H. Goulder, Kevin J. Mumford, and Kirsten Oleson, "Sustainability and the Measurement of Wealth," National Bureau of Economic Research Working Paper 16599, Cambridge, MA: NBER, December 2010, p. 2.

[10]This discussion draws on World Bank, *The Changing Wealth of Nations, Measuring Sustainable Development in the New Millennium* (Washington, DC: World Bank, 2011).

is no well-defined market for water use. Yet, depletion of water resources is a major concern, and the failure to price water to reflect its opportunity cost may exacerbate this depletion. The damage caused by air and water pollution, including the damage to human health, is also difficult to value precisely. Various ecosystem services (such as the supply of wild foods, cultural and recreational uses, and the aesthetic value of natural landscapes) also often lack market prices and are difficult to incorporate explicitly in wealth accounting. The same is true for such public goods as biodiversity and carbon storage. In part because of even more severe challenges to measurement, intangible capital is typically measured as a residual, the difference between total wealth and the sum of produced and natural capital. Total wealth is calculated as the present value of future consumption that is sustainable. In making these estimates, the World Bank applies a discount rate of 1.5 percent over a period of 25 years. This calculation of total wealth is thus independent of the value of the three forms of capital, allowing intangible capital to be calculated as the residual.

Table 20–1 provides wealth and per capita wealth data for different country groups in 1995 and 2005. Global wealth between these years increased by 34 percent. Over that same period, global population increased by 17 percent, resulting in a 17 percent increase in global wealth per capita. The largest absolute increases in per capita wealth were in the high-income nations, but every income group experienced gains. Table 20–1 also provides insights into both the composition of total wealth and changes in that composition as a function of income level. Intangible capital makes up the largest category of wealth at all levels of income and in both years. Yet there is a clear pattern of change in the composition of total wealth: low-income countries rely substantially on natural capital compared with high-income countries, with most of the difference lying in the relative proportions of intangible capital. This cross-sectional pattern is also apparent over time, as increased wealth in the low-income countries is associated with a shift in the composition of total wealth from natural capital to intangible capital. This decreasing reliance on natural capital as a component of total wealth is consistent with the structural transformation introduced in Chapter 16. Thus, notes a World Bank study, "For countries dependent on nonrenewable natural capital, transforming natural capital to other forms of wealth is the path to sustainable development."[11]

For many developing countries fortunate enough to have large endowments of natural resources, it is precisely the failure to make this transformation that threatens the sustainability of their development paths. Nigeria provides a prime example of this problem. Oil has dominated Nigeria's economy since the discovery in the late 1950s of the country's enormous reserves, accounting for nearly all of its export revenue and over one third of GDP. The government of Nigeria has financed much of its operations with oil revenue, continually drawing down this nonrenewable resource. Yet, Nigeria remains poor and underdeveloped relative to its potential because it has

[11]The World Bank, *The Changing Wealth of Nations,* p. 6.

TABLE 20-1 Wealth and Per Capita Wealth by Type of Capital and Income Group, 1995 and 2005 (constant 2005 US$)

	1995					2005				
INCOME GROUP	TOTAL WEALTH (US$ BILLIONS)	PER CAPITA WEALTH (US$)	INTANGIBLE CAPITAL (%)	PRODUCED CAPITAL (%)	NATURAL CAPITAL (%)	TOTAL WEALTH (US$ BILLIONS)	PER CAPITA WEALTH (US$)	INTANGIBLE CAPITAL (%)	PRODUCED CAPITAL (%)	NATURAL CAPITAL (%)
Low income	2,447	5,290	48	12	41	3,597	6,138	57	13	30
Lower middle income	33,950	11,330	45	21	34	58,023	16,903	51	24	25
Upper middle income	36,794	73,540	68	17	15	47,183	81,354	69	16	15
High income OECD	421,641	478,445	80	18	2	551,964	588,315	81	17	2
World	504,548	103,311	76	18	6	673,593	120,475	77	18	5

OECD, Organization for Economic Co-Operation and Development.

Source: Table 1.1 from The International Bank for Reconstruction and Development. The World Bank: *The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium*, 2011. Reprinted with permission.

failed to transform its natural capital into either manufactured or intangible capital (such as a more highly educated labor force). Indeed, there is a simple rule of thumb (known as the **Hartwick rule**) that suggests that a sustainable development path for countries that depend on nonrenewable resources requires the rents from those resources to be continually invested rather than consumed.[12] Nigeria is hardly alone among developing countries in having depleted its endowment of natural resource abundance while failing to develop accordingly. In addition to other oil exporters, such countries as the Democratic Republic of the Congo, Zambia, and Zimbabwe have all failed to develop to the potential available through their natural resource endowments.[13] In contrast, diamond-rich Botswana has been among the world's fastest-growing economies. Mexico and Peru have also succeeded in transforming natural capital into manufactured capital.

The tradeoffs are somewhat more complicated for countries in which the natural capital endowment is based on renewable resources, such as forest land. In these cases, property rights and other supporting institutions can play important roles in limiting harvests to sustainable levels. Deforestation in the Amazon exemplifies the overexploitation of a renewable natural resource. In this, and many other cases, the overutilization of renewable resources happens because the lack of market prices for such resources often leads them to be undervalued and threatened. The reliance of low-income countries on natural capital comes with the risk that the depletion of natural capital may cause permanent losses of biodiversity and other ecosystem services.

SAVING FOR A SUSTAINABLE FUTURE

Comprehensive wealth defines the current generation's well-being and constrains the well-being of future generations. A sustainable path of development (a path along which the well-being of each generation is at least equal to that of the previous generation) requires the creation of wealth at a rate at least equal to the rate of population growth. Wealth creation results from savings and investment over time. From one period to the next, the increase in comprehensive wealth equals **adjusted net saving (ANS)**, sometimes called **genuine saving**, defined as "gross national savings adjusted for the annual changes in the volume of all forms of capital."[14] We can measure countries' ANS by starting with the familiar concept of gross saving (S), which is simply total income minus total consumption. Maintaining wealth requires saving

[12]Specifically, the portion of the rents that should be invested under the Hartwick rule is equivalent to the depreciation of the capital stock, a proportion that approaches the full rent as the resource becomes progressively depleted.

[13]Chapter 18 discusses the pitfalls associated with countries' reliance on natural resource exports for their trade strategy.

[14]The World Bank, *The Changing Wealth of Nations*, p. 37. The only theoretical difference between the increase in comprehensive wealth and adjusted net savings is that the former includes capital gains (which arise from changes in the real price of assets).

sufficiently to offset depreciation of existing assets. In standard national accounting, this idea is captured by measuring **net saving (NS)**, defined as gross saving minus the depreciation of made capital (D_m):

$$NS = S - D_m \qquad [20\text{–}1]$$

But focusing on depreciation of only made capital is too limited, and in principle the concept should be expanded to include depreciation of natural capital (D_n), including the depletion of energy stocks, dwindling mineral assets, and damage from air pollution. This gives rise to the measurement of ANS:

$$ANS = S - D_m - D_n \qquad [20\text{–}2]$$

This corrected definition of net saving suggests that, if enough is saved each year to cover the depreciation of both made and natural capital, the economy can sustain its wealth and its level of consumption. In principle, this idea could be extended to include other assets as well, including human capital, knowledge, and social assets. However, measurement difficulties and lack of data make this extension difficult in practice, so we restrict our discussion here to made and natural capital. We can interpret ANS as an indicator of sustainability of an economy's development path. If ANS is declining (particularly if it is declining over consecutive years), this indicates that the economy's path is unsustainable, that it's depleting the productive base on which the welfare of future generations depends. Because some forms of capital are inevitably excluded from the calculation of ANS, its interpretation in practice requires caution and judgment. Research by economists Susana Ferreira, Kirk Hamilton, and Jeffrey Vincent has shown that ANS does predict future economic performance in a cross section of countries, but only if ANS is adjusted to account for the depletion of natural resources.[15]

Figure 20–2 illustrates ANS trends by developing region since 1970. While the data indicate substantial variability, several broad observations stand out. Both East Asia and South Asia have high and increasing rates of ANS compared with Latin America (for which the rate of ANS has been relatively low but positive over time) and sub-Saharan Africa (for which the rate of ANS has been low, trending downward, and negative after 2005). The unsustainable path reflected for sub-Saharan Africa is dominated by large oil-exporting countries, such as Nigeria, which (as noted above) fuels its economy through the gradual depletion of its oil reserves without converting that natural capital into other forms of capital. The positive trends in East and South Asia are dominated by the large and rapidly growing economies of China and India, which by comparison have relied less on the depletion of natural resources and more on the expansion of their stocks of physical and human capital. Box 20–1 extends the

[15]Susana Ferreira, Kirk Hamilton, and Jeffrey Vincent, "Comprehensive Wealth and Future Consumption: Accounting for Population Growth," *World Bank Economic Review* 22, no. 2 (2008), 233–48.



FIGURE 20–2 Adjusted Net Saving for Developing Regions, 1970–2008

Data exclude particulate emission damage.
Source: World Bank, "World Development Indicators," http://databank.worldbank.org.

analysis of ANS as an indicator of sustainability to incorporate the effects of population growth, illustrating these concepts with the case of Ghana, where ANS was positive while ANS per capita was negative.

To summarize, the concept of sustainable development has been defined in numerous ways; yet the most popular definition, that of the Brundtland Commission, highlights the requirement that consumption by today's generation must be accompanied by sufficient conservation of resources so as not to reduce the consumption potential of future generations. Knowing whether we are meeting this standard requires clear definitions and specific metrics. Economists' efforts to measure the sustainability of development have focused on net saving as an indicator of whether resource use by the current generation is expanding or contracting the potential creation of wealth by future generations. Experience to date is mixed. While some countries and regions have succeeded in rapidly expanding the potential wealth of future generations, others have been consuming rather than investing their resource base. These latter regions thus appear to be following unsustainable development trajectories. In many cases, the unsustainable depletion of natural resources and environmental damage through pollution are a result of market failure. The following section defines and illustrates this concept.

BOX 20-1 THE MALTHUSIAN EFFECT OF POPULATION GROWTH ON ADJUSTED NET SAVINGS IN GHANA

Adjusted net savings (ANS) is a concept that we can use to measure sustainability. As developed in the text, ANS tracks changes in a country's comprehensive wealth. Sustainability requires that wealth be nondecreasing from one generation to the next; yet from a welfare perspective, wealth must be measured per capita. If the population is growing, then maintaining per capita wealth at a constant level requires that comprehensive wealth grows at the same rate as population. From this perspective, it is entirely possible that development would not be sustainable in the per capita sense, even if comprehensive wealth is increasing. In other words, a negative rate of ANS reduces both total and per capita wealth; yet the converse is not necessarily true—a positive rate of ANS need not imply increases in wealth per capita. Some countries may add to their total wealth but not enough to compensate for population growth. To determine whether the rate of ANS is sufficient to increase wealth per capita, we must adjust that rate to take account of population growth. This requires subtracting from ANS the amount by which ANS must increase each year simply to keep up with population growth—the so-called **Malthusian term**. If the remaining difference is positive, then wealth per capita increases. If the remaining difference between actual ANS and the amount of ANS needed to maintain per capita wealth is negative, then wealth per capita decreases. This difference is the **adjusted net saving gap**. As a percent of gross national income, this saving gap tells us how much the rate of ANS would need to increase to keep wealth per capita constant over time.

Ghana provides an example of a country in which comprehensive wealth grew absolutely but fell per capita. Measuring saving and wealth in per capita terms requires a small modification to the accounting framework developed in the chapter. Because most of what we termed intangible capital consists of human capital (and is thus reflected in the population itself), we must exclude intangible capital from our calculation of comprehensive wealth per capita to avoid double counting (now that population enters the formula directly). For lack of data, we also exclude the value of carbon emissions per capita from our calculation of ΔANS.

The table on the following page details the calculation of changes in wealth per capita in Ghana for the year 2000, when Ghana's population grew by 1.7 percent. The left-hand column of the table decomposes tangible wealth per capita into its subcategories. The right-hand column shows the composition of adjusted net savings, subtracting the consumption of fixed capital and resource depletion

from the sum of gross national saving and educational expenditures. Ghana's ANS per capita was $16. The Malthusian term equals the population growth rate times comprehensive wealth per capita, or 0.017 × $2022 ≈ $34. Even though Ghana's ANS per capita was positive, the country's per capita wealth fell by $16 − $34 = −$18. Stated differently, the growth rate of Ghana's total wealth was $16/$2022 = 0.8 percent which was less than the 1.7 percent rate of population growth. This suggests an unsustainable long-run trajectory for Ghana.

Tangible wealth ($ per capita)		Adjusted net saving	
Subsoil assets	65	Gross national saving	40
Timber resources	290	Education expenditure	7
Non-timber forest resources	76	Consumption of fixed capital	19
Protected areas	7	Energy depletion	0
Cropland	855	Mineral depletion	4
Pastureland	43	Net forest depletion	8
Produced capital	686		
Total tangible wealth	2,022	Adjusted net saving	16
Population growth	1.7%	Δ Wealth per capita	−18

Source: Table 5.1 from The International Bank for Reconstruction and Development. The World Bank: *Where is the Wealth of Nations? Measuring Capital for the 21st Century*, 2006. Reprinted with permission.

MARKET FAILURES

The term *market failure* describes situations in which market prices deviate from scarcity values and individuals and companies make decisions that maximize their own profits but cause losses for others and society as a whole. A central theme of this chapter is that, within a single country, correcting those market failures and establishing properly working, efficient markets can be among the most powerful and effective mechanisms to promote efficient resource use, reduce environmental degradation, and generate sustainable development. At first blush, that proposition may seem counterintuitive. The point, however, is that environmental degradation often occurs because market participants do not take into account the full costs of their actions on the environment. For example, prices of goods produced in a factory may not include the costs to society of the air pollution generated by that factory. Government policies and interventions aimed at incorporating these costs into market decisions help improve environmental outcomes, make markets work better, and bring broader benefits to society.

Prominent among the market failures affecting resources are externalities—*costs* borne by the population at large but not by individual producers and *benefits* that accrue to society but cannot be captured by producers (see also Box 5–1). The most important externalities are those caused by the depletion or degradation of natural resources, including the environment. If resources are depleted at rates faster than they can be replenished or substituted by human-made capital, development will be unsustainable, either nationally or globally. If markets fail in this fundamental way, how can they promote sustainable development? To resolve this apparent conflict, we first need to analyze in greater depth the reason that markets fail to allocate natural resources efficiently.

EXTERNALITIES AND THE COMMONS

During the eighteenth century, as the Industrial Revolution began in England, cows still grazed on the commons of many villages in England and its American colonies. The essence of a village commons was **open access**, free of charge, to any member of the village. The first villagers to take advantage of open access had ample grazing for their livestock; their only cost was the time it would take to herd their animals to the commons, allow them to graze, and herd them home. But the amount of land was fixed and soil fertility and climate limited the quantity of grass. As more villagers used the commons, the grass became sparse, so the animals took longer to feed or, in the case of open rangeland, herders were forced to travel farther to find forage, so that everyone's costs rose. The rising average cost to each herder eventually discouraged grazing on the commons. But the new entrants did not have to pay compensation for the rising costs imposed on each of the previous entrants and more grazing took place than was in the interests of the village as a whole. Eventually, because no one incorporated the full cost of their grazing into their decisions, overgrazing destroyed the commons as a useful source of feed for everyone.

The dilemma of the commons is a widespread phenomenon, applicable to any limited resource to which access is unlimited. Grazing on open range, whether in the U.S. West or the African savanna, has the same outcome. Open access to timberlands or access at fees well below the social cost results in overlogging and the destruction of native forests in Brazil, Ghana, Indonesia, and many other tropical countries. Open access to fishing grounds in the North Atlantic, in Peru's Pacific waters, and in some inland lakes in Africa has already depleted fish stocks beyond their ability to regenerate. Free use of water from a stream benefits upland farmers, who have first access to the water, at the expense of downstream farmers, who get less water. Even traffic congestion in cities like Bangkok, Mexico City, and New York fits the description of a common property: City streets, to which access is free, are the common resource; each new vehicle causes worse traffic jams, forcing all previous entrants into longer, more costly commutes.

The earth's environment is composed of several common resources: air and the atmosphere, fresh water and the oceans, the earth's soils, and the diverse plant and animal species that live in this biosphere. Access to the environment typically is free. When manufacturers and farmers vent their waste into the air or water or create toxic dumps in the ground, they create health problems for the affected population, reduce the value of land in the affected area, destroy recreational potential, and generally reduce the welfare of people who value a clean environment. When lumber companies cut down a rain forest, they destroy the habitat of plant and animal species that are of value to others, including local populations that may harvest them or citizens or tourists who simply like to see them. Deforestation in the tropics significantly reduces biodiversity, perhaps eliminating potential new sources of pharmaceuticals or industrial products as well. They also may alter local climates, change patterns of water availability to surrounding farmers, and cause soil erosion. When we include the environment as a common resource, then much private activity generates external costs and market failure becomes a very general phenomenon. On a grand scale, we might think of the earth's entire atmosphere as a commons beset by the negative externality of greenhouse gas emissions, resulting in global climate change.

External costs and benefits are at the core of the common resource problem. A new producer creates higher costs for all previous entrants or all producers impose external costs on the general population. In either case, in the absence of regulation, taxes, or property rights for environmental quality, external costs are not borne by the producers who cause them and the prices of their products do not reflect the social costs of production. Thus more of these resource-depleting or -polluting goods and services are produced and consumed than would be the case if prices reflected external costs. Hence societies pollute more than their people would choose if markets reflected all social costs.

Figure 20–3 shows this process. In a market with competitive producers, the supply curve S represents private marginal costs, PMC. Market equilibrium occurs at price P_1 with output Q_1. But, if this is a polluting industry, the external costs make the social marginal cost, SMC, higher. If these costs were reflected in the market, the price would jump to P_2 and demand, and therefore output, would be reduced to Q_2. As less of the offending product is grown or manufactured, there would be less environmental degradation.

Designing responses to problems of the commons requires a closer look at what we mean by *the commons*, more formally referred to as **common-pool resources (CPRs)**. CPRs have two characteristics: It is difficult to exclude anyone from using them, and use by one person reduces the availability of that resource for use by others. Potential solutions to the types of problems noted earlier depend more specifically on the nature of property rights governing the CPR. Most of the problems of CPRs arise when property rights are defined as open access regimes, where there

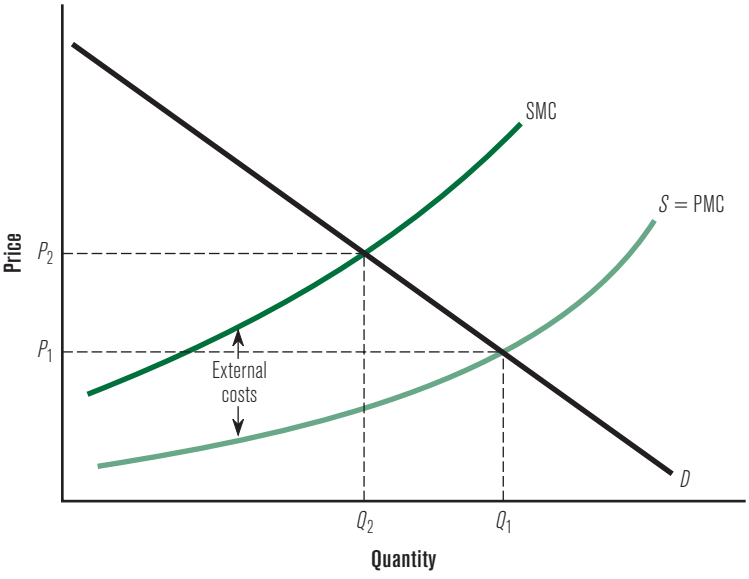


FIGURE 20-3 External Diseconomies

Polluters impose costs on others. If these external costs were reflected in the firms' costs, the social marginal cost curve (SMC) would prevail, the market price would be P_2, and output would be Q_2. But because the firms do not bear these costs, their private marginal cost curve (PMC) is lower, so more of the polluting product (Q_1) is produced and consumed.

are no property rights at all. Yet, other CPRs may be characterized by the existence of group property rights, individual (or firm) property rights, or government property rights. In such settings, tragedies of the commons are not inevitable. Economics Nobel Prize winner Elinor Ostrom shows that a wide range of institutions for governing and managing CPRs can, and have, evolved for CPRs that are not simply open access resources. Farmer-managed irrigation systems in Nepal are one among many examples of locally determined rules for conserving natural resources. Ostrom and colleagues find that particular settings are most conducive to the evolution of successful informal rules for managing CPRs. These settings are environments where the resource degradation is not yet too severe and where users have good information about conditions, situations in which users depend on the resource as a main source of their livelihoods and are interested in sustainability of the resource, and settings in which national governments facilitate local organization (rather than attempt to impose rules from outside the community).[16]

[16]Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (New York: Cambridge University Press, 1990). See also Joanna Burger, Christopher B. Field, Richard B. Norgaard, Elinor Ostrom, and David Policansky, "Revisiting the Commons: Local Lessons, Global Challenges," *Science* 284, no. 5412 (April 1999), 278.

POLICY SOLUTIONS

The market failures that lead to overexploitation of natural resources stem from external costs that are not borne by producers. Even in a well-functioning market economy, externalities require government intervention to make markets work better and to reach efficient market outcomes. Governments can bestow property rights on private users, regulate access to common resources, impose taxes (or pay subsidies) that reflect external costs (or benefits), and issue tradable access rights. There is also recent experience with informal regulation by local communities in the absence of formal government regulation. We discuss each of these options in turn.

PROPERTY RIGHTS

Common properties generate external costs because no one owns or controls the right to exploit them. For some resources, a simple solution would be to confer ownership, which economists call **property rights**, on a single individual or company. As long as the owner is a profit-maximizer and sells output in a competitive market, the socially optimal outcome is achieved without further government intervention. Property rights, to be effective, must be exclusive and well defined, leaving no doubt to the owners and possible competing claimants about what has been conferred and to whom. Rights need to be secure, so that the risk of loss through legal challenge or expropriation is reduced and enforceable through the judicial system. Ownership must be valid over a long-enough horizon that the owners have a stake in the long-term, sustainable exploitation of the resource. Longevity converts the resource into an asset for the producer, who can reap the benefits from investments in improving and sustaining its productivity. And the rights must be transferable, so the owner can realize the benefits of the resource asset by selling the property at any time.

The implications of clearly assigned property rights for market-based pollution abatement were formalized in 1960 by Nobel economist Ronald Coase. The **Coase theorem** posits that under certain circumstances (in particular, zero transaction costs), clear assignment of property rights leads bargaining in free markets to attain optimal levels of externalities. It is important that the Coase theorem also suggests that this result is independent of the specific allocation of those property rights.[17] A stylized example of the Coase theorem could be a situation in which an upstream factory dumps effluent waste into a river, which then flows downstream and destroys a fishery. When the river itself belongs to neither the factory nor the fisherman, this negative externality persists. The Coase theorem suggests that if either the factory or the fisherman is assigned property rights to the river, free bargaining between the two parties will result in the optimal level of pollution.

[17]R. H. Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (1960), 1–44.

GOVERNMENT REGULATION

As an alternative to conveying private property rights, governments themselves can act as the owners of common resources and directly regulate their use. Governments can limit the quantity of a hunter's kill, a fisher's catch, a logger's haul, a rancher's herd, or a polluter's emissions. And they can regulate the kinds of equipment that can or must be used: Some kinds of fishing nets, boats, or navigation equipment have been banned; hunters may be restricted in their choice of weapons; polluters are required to install equipment that scrubs gas emissions and treats wastewater.

Quantity regulations raise two issues. First, how do the regulators know the optimal levels of access and output? If property rights can be conveyed, efficient outcomes are approached through market forces and no government judgments are needed. But if regulation replaces the market, regulators need to estimate the characteristics of both producers' costs and users' demand for the products of a common resource. To get a sense of these information requirements, consider the regulation of air pollution. The external costs of pollution are manifest in the reduced welfare of others: poor health, unsightly environment, lower property values, fewer and more expensive recreational possibilities, and possibly reduced productivity and income. If these costs could be measured, they would be depicted by a curve such as marginal external cost (MEC) in Figure 20–4, which shows the marginal external cost of pollution (measured along the horizontal axis). Those who believe that the costs of pollution are higher than normally recognized or who put a high premium on reducing pollution argue,

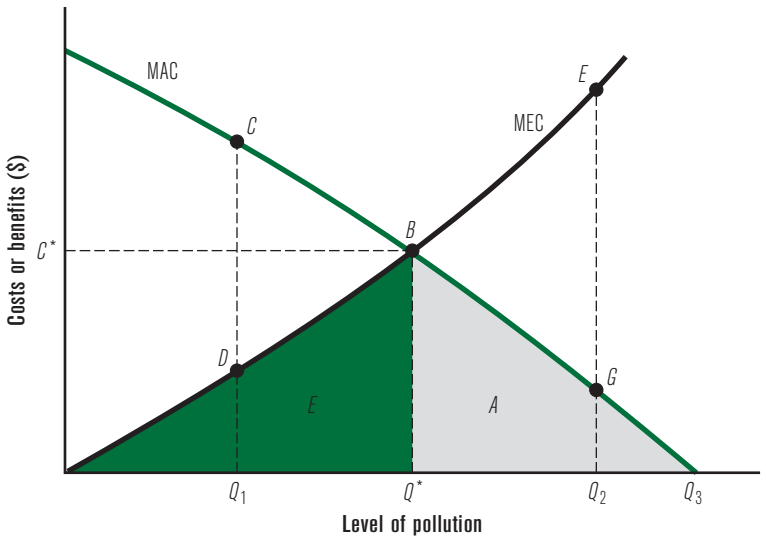


FIGURE 20–4 Optimal Level of Pollution

The marginal external cost of pollutants, borne by the population, is given by MEC; the marginal cost of abatement, borne by the firm, is MAC. The total cost to society (area $E + A$) is minimized and pollution is optimal at Q^*, where MEC = MAC.

in effect, that the MEC curve should be shifted up. Wherever it is located, any reduction in pollution (a movement to the left on the horizontal axis) means a reduction in the cost to society from pollution or, equivalently, an increase in the marginal external benefit of abatement.

However, there is a cost to abating pollution. The polluting firm, say, a petrochemical plant, can reduce its effluents either by changing its production process, installing abatement equipment such as gas scrubbers and water treatment plants, or reducing output. The marginal cost of abatement (MAC) function traces these marginal abatement costs. At any point along MAC, the cost shown is that of the lowest-cost method of abatement. Moving from right (high pollution) to left (lower pollution), MAC rises because it becomes increasingly costly to clean up air or water the stricter are the standards or the lower the level of contamination. It is important to recognize that these costs of abatement, although borne by the petrochemical firm, also are costs to society because they involve either less consumption of petrochemicals or savings spent on abatement that otherwise might have been spent on investment in other goods or services people want. These abatement costs thus have equal weight to the benefits gained by reducing pollution.

Society's (and government's) aim should be to minimize the combined costs of pollution and its abatement. This is achieved at Q^* in Figure 20–4, where MEC = MAC. At this point $Q_3 - Q^*$ of pollution has been abated (assuming for convenience that Q_3 is the maximum amount of pollution), and Q^* of pollution remains. Because these are marginal cost curves, the total external cost of pollution is the area E under MEC from 0 to Q^*. And the total cost of abatement is given by the shaded area A under the MAC curve between Q^* and Q_3. With less pollution, such as Q_1, the marginal abatement cost exceeds the marginal external cost of pollution and the total cost of additional abatement, the area of the trapezoid Q_1Q^*BC, exceeds the net gain from reduced pollution, Q_1Q^*BD. If more pollution is permitted, such as Q_2, the additional external cost, Q^*Q_2EB, exceeds the reduced cost of abatement, Q^*Q_2GB.

Therefore, society is better off with some pollution than with none, because abating the last unit of pollution is expensive relative to its benefits. Similarly, society gains from some exploitation of natural resources, even nonrenewable ones. But how do the regulators, who wish to achieve this optimal level of pollution, know what it is? To find Q^*, they have to know all the external costs of pollution, as a function of the levels of contaminants in the air, water, and soil. In addition to these estimates of direct economic costs, finding Q^* requires having a method for estimating the values people place on environmental amenities such as clean air, water, and soil. The costs in some cases are substantial. A recent collection of studies from the World Bank estimate the costs of environmental degradation in the Middle East and North Africa.[18] As summarized by country and by resource category in Table 20–2, these costs are

[18]Leila Croitoru and Maria Sarraf, eds., *The Cost of Environmental Degradation: Case Studies from the Middle East and North Africa* (Washington, DC: World Bank, 2010).

TABLE 20–2 Cost of Environmental Degradation in the Middle East and North Africa (percent GDP)

	ALGERIA	EGYPT	LEBANON	MOROCCO	SYRIA	TUNISIA
Air pollution	1.0	2.1	1.0	1.0	1.3	0.6
Lack of access to water supply and sanitation	0.8	1.0	1.1	1.2	0.9	0.6
Land degradation	1.2	1.2	0.6	0.4	1.0	0.5
Coastal zone degradation	0.6	0.3	0.7	0.5	0.1	0.3
Waste management	0.1	0.2	0.1	0.5	0.1	0.1
Subtotal	3.6	4.8	3.4	3.7	3.3	2.1
Global environment (CO_2 emissions)	1.2	0.6	0.5	0.9	1.3	0.6
Total	4.8	5.4	3.9	4.6	4.6	2.7

GPD, gross domestic product.

Source: Leila Croitoru and Maria Sarraf, eds., *The Cost of Environmental Degradation: Case Studies from the Middle East and North Africa* (Washington, DC: World Bank, 2010).

estimated to range from approximately 3 percent to 5 percent of GDP. Various studies of the cost of environmental degradation in China suggest a figure up to 13 percent of GDP. Such studies, however, confront a range of limitations (data in particular), and their results are generally not directly comparable (due to the use of differing methodologies). Their results are best interpreted as providing orders of magnitude.

In the absence of precise cost estimates, regulators must set somewhat arbitrary standards based on studies estimating the impact of pollutants on human health, animal survival, forest die back (from acid rain) and regeneration, and presumed climate changes.

TAXES, SUBSIDIES, AND PAYMENTS FOR ENVIRONMENTAL SERVICES

A third option is that, in principle, the government also could achieve optimal rates of resource use by imposing taxes that reduce the incentive for producers to use common properties or manufacture polluting products. A tax might be imposed on output that represents the external costs of production, so that the private marginal cost schedule shifts up to equal the social marginal cost schedule.[19] This might take the

[19]In Figure 20–3, the tax would shift the PMC schedule up to coincide with the SMC schedule.

form of a tax on each ton of steel or petrochemicals at a rate representing the external cost of pollution or a tax on gasoline to cover the costs of both pollution and traffic congestion. If the tax is on output or level of effort, the incentive is to reduce production of the good with external costs. If the tax can be levied on the externality itself, there is an additional incentive to invest in reducing external costs. For example, a tax on the quantity of pollutants would give petrochemical plants an incentive to abate pollution, because the tax then is reduced. In general, attempts to tax pollutants have had limited success (and even less success in high-income countries than in developing and transitional economies). Monitoring is difficult (and expensive), charge rates generally are set too low, and tax avoidance can be relatively easy. Malaysia began to implement pollution taxes in the 1970s, and China now operates the world's largest pollution tax system, though its effectiveness is disputed.[20] Box 20–2 describes efforts to tax water pollution in Colombia.

BOX 20–2 TAXING WATER POLLUTION IN COLOMBIA

Until the mid-1990s, Colombia's efforts to curb water pollution had relied exclusively on a traditional command-and-control approach in which regional environmental regulatory authorities were responsible for issuing permits for the discharge of wastewater. These permits were intended to limit the quantity of discharges and to require that they meet specific effluent standards. These regional regulatory authorities, known as CARs (which stands for *Corporaciónes Autónomas Regionales*), had issued permits to less than one third of all discharges by 2002 and were lax in their monitoring and enforcement of discharge standards.

Beginning in 1997, Colombia tried a new approach to reducing water pollution based on economic incentives, a per unit tax of pollution emitted. The implementation of this tax was specified in Decree 901, which directed that CARs first produce inventories of all facilities discharging wastewater and then set five-year pollution reduction goals for each water basin in their jurisdiction. The CARs would then set tax rates per unit of effluent based on a minimum fee determined by the Ministry of Environment, and then monitor the facilities' discharges on a regular basis.

Implementation of this water pollution tax encountered several barriers. Implementation was uneven across CARs, with only nine of Colombia's 33 CARs

[20]Jeffrey R. Vincent and Rozali Mohamed Ali, *Managing Natural Wealth: Environment and Development in Malaysia* (Washington, DC: Resources for the Future, 2005).

fulfilling all of the program requirements by 2003. In addition, less than half the polluting facilities were actually charged the tax, and collection rates from those facilities averaged only 27 percent between 1997 and 2002. Municipal sewage authorities were particularly reluctant to pay their taxes, sparking complaints of unfairness by private firms. Implementation of the pollution tax was also hindered by confusion surrounding the continued existence of the previous command-and-control system of permits.

Despite these impediments, a number of Colombia's water basins achieved significant reductions in wastewater discharges. The extent to which this achievement can be attributed to the discharge fee program is difficult to assess. The confusion lies in the fact that before 1997, nearly all of the CARs lacked the regulatory and enforcement capacity necessary to effectively implement *any* pollution reduction program. Once charged with the added responsibilities of creating inventories of polluting facilities and developing monitoring systems (with substantial technical support from the Ministry of Environment), the CARs' ability to enforce the old emissions standards was also enhanced. The discharge fee system was more transparent than the previous approach, requiring regular reporting by the CARs. In addition, the new fee system created economic incentives for better performance by the CARs themselves, which were allowed to keep fee revenues.

Source: Allen Blackman, "Incentives to Control Water Pollution in Developing Countries, How Well Has Colombia's Wastewater Discharge Fee Program Worked and Why?," *Resources* (Spring 2006) [Resources for the Future, Washington, DC].

Taxes that internalize external costs have two important advantages over regulation. First, they allow the producer to choose the method of reducing access to a common resource so that rents are not dissipated in wasteful expenditures forced by regulators. The cost savings with this flexibility can be substantial. Second, taxes can generate substantial revenues for the government. These kinds of revenues can be used to fund environmental programs or in other ways to compensate citizens for the harm caused by pollution and other environmental degradation.

The inverse of this discussion is also true. Some externalities are positive. Positive externalities exist when an activity creates social benefits that are external to the firm. Examples might include the invention of new technologies that spill over into other firms or sectors, or the provision of improved health or training for workers, or the pollination of neighboring orchards by bees from an apiary. Because these benefits are external to the firm, marginal social benefits exceed marginal private benefits, and

markets undersupply them. In such cases, subsidies (which are simply negative taxes) may be justified to encourage optimal production of positive externalities. Such programs, however, run the risk of subsidizing activities that might have happened even in the absence of the subsidy.

Payments for environmental services (PES) are an extension of the idea of subsidizing the production of positive externalities. Economist Sven Wunder has defined PES as a voluntary transaction by which a well-defined environmental service is purchased by a buyer of that service from a provider conditional on the provision of the service.[21] Such arrangements effectively create markets that had been missing for external environmental benefits. For instance, forest conservation may provide little direct benefit to farmers, whose private costs and benefits might then favor cutting down forests to expand their fields. Yet, downstream populations must then forego the benefits created by forests, such as biodiversity and carbon storage. PES systems create markets through which potential beneficiaries of environmental services pay potential providers of those services. In this example, PES could be made to farmers to induce them not to destroy forest land for agricultural expansion. The buyers of environmental services might be the actual users of the services (for instance, a producer of hydroelectric power might pay upstream land users to conserve water). In other instances, governments (or even international financial institutions such as the World Bank) might pay providers for environmental services on behalf of users of those services.[22] Examples of user-financed PES programs are payments for watershed services in Pimampiro, Ecuador; payments for biodiversity and watershed services in Los Negros, Bolivia; and payments for carbon sequestration in Ecuador. Examples of government-financed PES programs are China's Sloping Land Conservation Program, Mexico's Payments for Hydrological Environmental Services program, and South Africa's Working for Water program.[23] PES programs exemplify practical efforts to implement the Coase theorem, as do the idea of marketable permits to pollute.

MARKETABLE PERMITS

A fourth intervention is to create a property right where none exists by issuing **marketable permits**, granting the holders the right to harvest a common resource up to a given limit or giving producers a license to pollute the environment up to specified amounts. Although environmentalists sometimes scoff at the notion of a *right* to pollute or exploit resources, this idea recognizes that zero pollution is usually not optimal because of the costs involved in achieving that goal. These permits may be

[21]Cited in Stefanie Engel, Stefano Pagiola, and Sven Wunder, "Designing Payments for Environmental Services in Theory and Practice: An Overview of the Issues," *Ecological Economics* 65 (2008), 663–74.

[22]Engel et al., Designing Payments for Environmental Services in Theory and Practice," p. 666.

[23]These programs are reviewed in Sven Wunder, Stefanie Engel, and Stefano Pagiola, "Payments for Environmental Services in Developing and Developed Countries," *Ecological Economics* [Special ed.] 65, no. 4 (May 2008), 663–852.

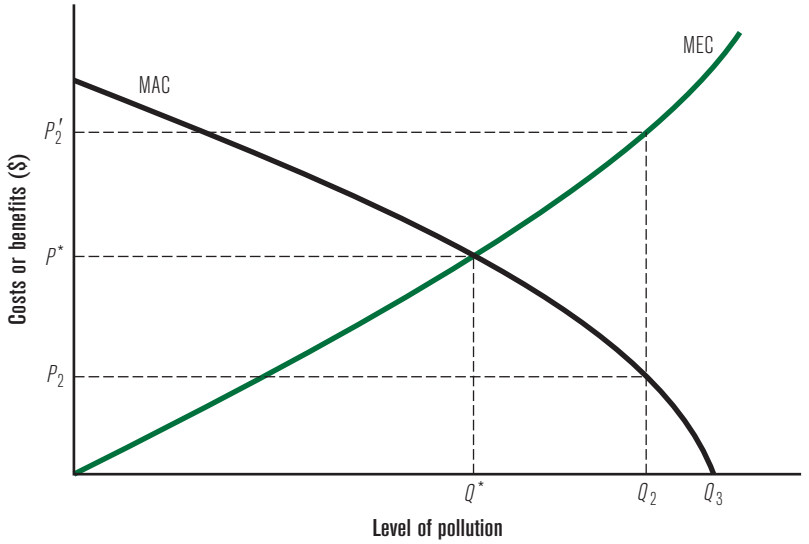


FIGURE 20–5 Marketable Permits to Pollute

The marginal cost of abatement (MAC) schedule shows the demand for pollution rights. If rights are issued to pollute up to Q_2, these are worth P_2 to polluters, but the public places a value of P_2' on reducing pollution. This could be the basis for a bargain to reduce pollution to the optimal level, Q^*.

MEC, marginal external cost.

the most efficient way to reduce pollution and resource overuse, and this approach has been a major innovation of the past decade. The U.S. sulfur dioxide reduction scheme relies on tradable rights, and both Iceland and New Zealand revived fishing stocks by assigning fishing rights at a sustainable level and allowing fishers to trade their quotas freely.

Figure 20–5 shows how a pollution permit works.[24] The MAC and MEC curves are copied from Figure 20–4. Say that the government knows the optimal level of pollution and auctions off emissions permits totaling Q^*. Any firm polluting without a permit, if detected, would be fined or shut down. (Thus permits have the same enforcement requirements as regulations or taxes.) If MAC represents the cost of pollution abatement for all firms in the market, they would bid up the price of permits to P^*. The MAC schedule (moving from right to left) shows that firms can reduce pollution from Q_3 to Q^* at costs less than P^*. Additional reductions in pollution would cost more than P^*. Therefore, the MAC schedule is the demand curve for permits. Either the government can issue a given number and observe the auction price paid for them or it can set a price and issue whatever number of permits is demanded by polluters.

[24]This figure is adapted from David W. Pearce and R. Kerry Turner, *Economics of Natural Resources and the Environment* (Baltimore: Johns Hopkins Press, 1990), 319–20.

In developing and transitional economies, where policy dependence on market forces is a more recent phenomenon, marketable permits have been less common, though are increasing in frequency. Examples are transferable quotas for fisheries in Chile (beginning in 1991) and South Africa (beginning in 1997). Chile also implemented an auctioning system for bus permits in its capital city, Santiago, starting in 1991, and a system of tradable quotas for the emission of total suspended particulates (TSP) from stationary sources in Santiago the following year. A review of Chile's TSP trading program found it to have had limited success due to several problems with its implementation.[25] One problem was that the enabling law itself was vague in regard to the way that permits would be allocated across firms. The program also initially overestimated the emissions target, which was subsequently reduced. By issuing permits totaling too high a level of emissions, effectiveness was limited. Effectiveness was further limited by the long period of time acquired for approval by the government for the sale of pollution permits between firms. These technical and institutional impediments to implementing Chile's marketable permit program indicate the challenges facing countries that choose this approach.

INFORMAL REGULATION

Informal regulation of pollution does not rely on traditional government-generated regulation. In place of formal regulation, such programs often rely on the public disclosure of information about firms' and factories' environmental performance to local stake holders (for instance people living near a factory or a firm's customers). Potential benefits of informal regulation are its low cost and its lack of reliance on formal regulatory institutions. These characteristics make informal regulation of pollution especially attractive in developing countries, which are typically short of both funds and highly evolved regulatory institutions. Public disclosure potentially empowers communities to pressure polluters to improve their environmental performance. The effects of such pressure might operate directly on the firm or indirectly if increased community awareness of the problem and its source leads to greater demand for formal intervention. In some circumstances, public disclosure can even bring pressure from capital markets for firms to alter their behavior. Public disclosure programs take one of two forms: registries of plants' emissions (without rating environmental performance), and programs that use plants' emissions data to rate environmental performance.[26] Chile and Mexico are among at least 20 countries with pollution registries, and performance ratings programs exist in China, Ghana, India, and several other developing countries.

[25]Jessica Coria and Thomas Sterner, "Tradable Permits in Developing Countries: Evidence from Air Pollution in Chile," *Journal of Environment & Development* 19, no. 2 (2010), 145–70.

[26]For an excellent survey, see Allen Blackman, "Alternative Pollution Control Policies in Developing Countries," *Review of Environmental Economics and Policy* 4, no. 2, (Summer 2010), 234–53.

The effectiveness of public disclosure types of informal environmental regulation, in practice, may depend on characteristics that may or may not be present in developing countries. For instance, in the absence of a credible threat of more stringent formal regulation, informal regulation may have less traction. In addition, capital markets may have little leverage in settings dominated by small firms with little reliance on outside investors. Further, information does not always flow freely in every country. Despite these contingencies, reviews of the effectiveness of public disclosure programs have found them to be effective in reducing pollution. These effects tend to be greater in cases in which the initial violations were the most flagrant. Yet, to date, rigorous evidence of successful ratings programs comes only from Indonesia, the Philippines, Vietnam, and China.[27]

POLICY FAILURES

Although some government intervention is necessary to correct for the market failures associated with natural resources, it is equally true that, all over the world, government policies frequently contribute to wasteful use of resources and the degradation of the environment. Too much intervention or intervention of the wrong kind can be just as costly as too little intervention. We have seen that, when production has external costs, one approach is to internalize those costs by raising production costs through taxing output or granting marketable property rights. But, instead, governments commonly subsidize or otherwise reduce the costs the production of commodities that degrade natural resources and often compromise property rights in ways that encourage rapacious exploitation. Examples are not hard to find.

Forestry policy has been especially destructive in many tropical countries. For many years Brazil subsidized ranching and other activities that encroached on the Amazon rain forest. Through at least the first five years of the twenty-first century, the area deforested each year in the Brazilian Amazon was equivalent in size to the state of New Jersey. Approximately two thirds of the deforested land in Brazil in recent decades has been used for cattle ranches, and most of the remaining third has been cleared for subsistence farming. More recently, commercial agriculture has also been an important source of deforestation in Brazil. Tax policies favorable to pasture land encouraged deforestation. Brazil has reformed some of these environmentally harmful tax incentives and stepped up efforts to regulate deforestation. At the end of 2010, the government of Brazil announced that the rate of deforestation in the Amazon for the previous year was 67 percent lower than the average rate of deforestation for the period 1996–2005. Time will tell if this improved trend proves durable. Box 20–3 describes the policy failures underlying the rapid rate of deforestation in Indonesia.

[27]Blackman, "Alternative Pollution Control Policies in Developing Countries," pp. 241–42.

BOX 20-3 POLICY FAILURES AND DEFORESTATION IN INDONESIA

Over the course of the twentieth century, Indonesia's forested area was cut by nearly 50 percent, mostly the result of commercial logging. In May 2010, the government of Indonesia introduced a moratorium on new logging contracts. Yet, for decades before 2010, Indonesia's rapid rate of deforestation was in some measure the result of government policy, more specifically, the failure of government policy. Four main categories of policy failure are to blame: policies supporting the buildup of excess capacity in the wood processing industry, conversion of forested areas into agricultural land, inefficient collection and use of rents collected from the forestry sector, and the absence of effective property rights in many forested areas.

Beginning in the 1970s, government policy in Indonesia supported the expansion of the timber processing industry by restricting the export of raw timber. This depressed domestic timber prices, thus subsidizing timber processors. The resulting overcapacity stimulated the demand for raw timber, much of it logged illegally. Government-subsidized loans to the timber processing industry heightened these incentives.

Substantial tracts of forest were also clearcut to create space for large oil palm plantations. This process accelerated during the 1990s and is cited as having contributed to the massive forest fires of 1997–98. In addition, the government was promoting migration from densely populated Java to the outer islands and cleared additional forest land to accommodate the migrants' farming activities.

The government was also inefficient in its collection of rents from the forestry sector. By law, the government was supposed to levy fees based on both forest area harvested and the number of units harvested. These fees were to be used for replanting the harvested areas. In practice, rent collection was limited, estimated to be between 24 and 36 percent of total economic rents in 1997–98, resulting in substantial windfall profits for the logging industry.

All of these problems have been made worse by the absence of secure property rights. By law, the state controls all forest areas. This means that the actual users of those lands have minimal incentive to invest in the long-term maintenance of that resource. Accelerated deforestation is one symptom of this problem.

Source: Raymond Atje and Kurnya Roesad, "Who Should Own Indonesia's Forests? Exploring the Links between Economic Incentives, Property Rights and Sustainable Forest Management," Economics Working Paper Series 76, Center for Strategic and International Studies, Jakarta, 2004.

Trade policy has been equally destructive. Ghana, Indonesia, and Malaysia, for example, placed bans on log exports as a means of promoting wood-processing industries. Export bans drive down the domestic price of tropical hardwood and so make it very profitable for sawmills and plywood mills to purchase logs and export semifinished products. But these industries are usually inefficient and consume resource rents through higher production costs. Because timber companies cannot export tropical hardwoods, such as ebony and mahogany, as logs, these valuable species are used along with low-value timber to make inexpensive products, such as plywood sent to Japan to make forms for pouring cement. The role of self-imposed log export bans in destroying rain forests should be a warning to northern countries that want to preserve tropical forests by imposing their own import bans on these logs.[28]

Energy pricing is another common policy failure. In oil-rich countries like Nigeria and Venezuela, energy has been kept cheap in an attempt to stimulate industrialization and diversification, and Indonesia subsidized kerosene ostensibly to help the rural poor. This policy has multiple adverse effects. It encourages wasteful domestic consumption and reduces the country's petroleum and gas reserves and its export-earning potential. It encourages the use of cars and minibuses and so adds to congestion. Cheap energy promotes industry ill-suited to the country's endowments. Firms and consumers have little incentive to adopt energy-saving technology. Because burning oil is an important source of air pollution, all these overuses contribute to environmental degradation. Similarly, some oil importers, such as Argentina, China, Egypt, and India, have subsidized petroleum products by as much as 50 percent of the world price and so encouraged imports they cannot afford, industries that cannot compete in world markets, and environmental degradation that market pricing would have discouraged.[29]

Infrastructure investment is a third area of widespread policy failure. In forestry and energy pricing, governments typically underprice the resource and fail to force private operators to account for external costs. In infrastructure investment, governments often create new external costs and fail to account for them in project planning. Environmental groups have focused much of their ire on investments in power dams, irrigation and flood control systems, roads, and power plants that damage their environments. Dams that flood their upstream areas displace local populations, who sometimes do not fare as well in new locations, and destroy natural habitat. Egypt's huge Aswan Dam controls the floodwaters of the Nile. But before the dam, those floodwaters had beneficial effects, replenishing soil and leaching out unwanted salts, so the dam may have reduced agricultural productivity over the long run. This

[28]On forest policies, see Robert Repetto and Malcolm Gillis, eds., *Public Policies and the Misuse of Forest Resources* (Cambridge: Cambridge University Press, 1988); Jeffrey Vincent, "The Tropical Timber Trade and Sustainable Development," *Science* 256 (1992), 1651–55.

[29]Bjorn Larsen and Anwar Shah, "World Fossil Fuel Subsidies in Global Carbon Emissions," Working Paper WPS 1002, World Bank, Washington, DC, October 1992, esp. chart 1.

does not mean that dams are necessarily bad ideas but rather the full costs are not always recognized up front and incorporated into prices.

The problem of common pool resources, policy responses to externalities, and policy failures are particularly important for the poor. The world's poor are disproportionately concentrated in fragile environments, and their livelihoods depend disproportionately on natural resources (especially open access and other forms of common property resources). The interactions between poverty and the environment are complex, yet they are central to an understanding of sustainable development.

POVERTY-ENVIRONMENT LINKAGES

Economic growth, poverty alleviation, and environmental sustainability are three primary goals for development. As Figure 20–6 suggests, these goals are intrinsically linked to one another through a complex set of interactions. In characterizing this

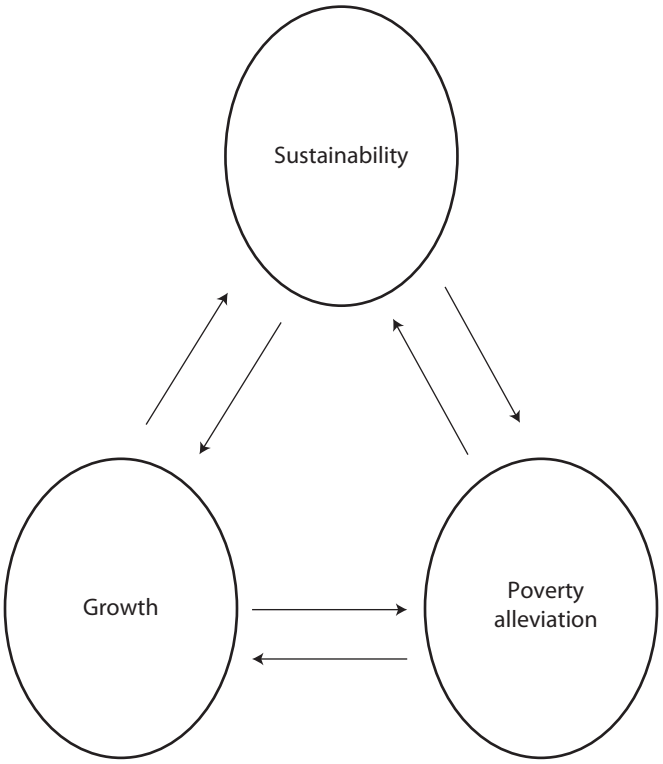


FIGURE 20–6 The Critical Triangle of Development Goals

Source: Vosti, S. A. and T. Reardon. 1997. *Sustainability, Growth, and Poverty Alleviation: A Policy And Agroecological Perspective*, Figure 1.1. Washington, DC: International Food Policy Research Institute.

"critical triangle of development goals," economists Stephen Vosti and Thomas Reardon note that these goals are highly compatible with one another *in the long run*. It is easy to see that a robust and healthy natural resource base will better support growth in agricultural productivity, which in turn can play a critical role in poverty alleviation and economic growth. However, in the short run countries may confront difficult tradeoffs among these goals. If activities of the poor contribute to the degradation of the commons, efforts to preserve the environment by limiting access to the commons might aggravate poverty. The nature and severity of these tradeoffs may vary widely across different combinations of circumstances, though institutional failures and resulting externalities of the types discussed earlier provide a unifying theme.[30] This section focuses on the bidirectional linkages between poverty alleviation and environmental sustainability.

The complex relationship between poverty and environmental degradation arises because severe poverty in the world is disproportionately rural, and rural livelihoods depend heavily on natural resources, especially land and water. The reliance of the poor on natural resources centers on agriculture but also includes livestock, fisheries, and forest resources. Degradation of these resources directly threatens livelihood strategies of many of the world's poor. The Millennium Ecosystem Assessment Board, in its 2005 report *Ecosystems and Human Well-Being*, notes that "Over the past 50 years, humans have changed ecosystems more rapidly and more extensively than in any comparable period of time in human history, largely to meet rapidly growing demands for food, fresh water, timber, fiber, and fuel. This has resulted in a substantial and largely irreversible loss of the diversity of life on earth." The report further suggests that

> The changes that have been made to ecosystems have contributed to substantial net gains in human well-being and economic development, but these gains have been achieved at growing costs in the form of the degradation of many ecosystem services . . . and the exacerbation of poverty for some groups of people. These problems, unless addressed, will substantially diminish the benefits that future generations obtain from ecosystems. The degradation of ecosystem services could grow significantly worse during the first half of this century and is a barrier to achieving the Millennium Development Goals.[31]

The problems of natural resource degradation particularly threaten the poor because the poor are concentrated in fragile environments, which are more vulner-

[30]A deeper discussion of institutional failures and externalities, including the causes and consequences of rapid population growth among the poor, as links between poverty and environmental degradation appears in Partha Dasgupta, "Population, Poverty, and the Natural Environment," in Karl-Goran Mäler and Jeffrey Vincent, eds., *Handbook of Environmental Economics*, vol. 1 (Amsterdam: North Holland, 2003), chap. 5.

[31]Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: Synthesis.* (Washington, DC: Island Press, 2005), p. 1.

able to degradation. Examples of fragile environments are areas with thin soil cover that are vulnerable to erosion, tropical forests, and coastal estuaries. About one quarter of the world's population lives on fragile land, yet the distribution of this population varies widely by region. While only 11 percent of the population of wealthy countries (Organization for Economic Co-Operation and Development; OECD) live on fragile lands, this proportion increases to about 25 percent in South and East Asia, and jumps to nearly 40 percent in sub-Saharan Africa and the Middle East. These percentages increase substantially when we consider the rural poor in particular. About 66 percent of the world's rural poor live on fragile lands, with the proportion rising to approximately 75 percent of the rural poor in sub-Saharan Africa, West Asia and North Africa.[32]

There are numerous examples of the value of environmental services in the livelihood strategies of the poor. Environmental services include such diverse functions as the supply of water and food, carbon sequestration, crop pollination, and recreation. Mangroves, coral reefs, forests and forested watersheds, and flood plains are among the key environmental features that figure prominently in the livelihood strategies of poor people around the world. Coastal mangroves, for instance, can be a valuable source of forest and fishery products. Economists Saudamini Das and Jeffrey Vincent have shown that mangroves buffered coastal villages in Orissa, India, from a super cyclone in 1999. The storm surge from that cyclone killed 10,000 people, but deaths were significantly reduced in villages protected by mangrove forests.[33]

Tropical rain forests are a prominent example of a diminishing natural resource of substantial economic value. Rain forests provide benefits that are both local and global in scope. The global environmental benefits of tropical rain forests include carbon sequestration (discussed later in the chapter), rainfall recycling, and flood control. Conversely, the loss of rain forests can impose important negative externalities, including increased buildup of carbon dioxide, increased smoke, and reduced rainfall, a particular concern in the Amazon. While these effects are quite broad, even global, the effect of tropical rain forests on the poor depend more on rain forests' local impact.

Direct harvesting of timber products is the most obvious local service and also the most problematic. The challenge with timber is to limit harvesting to sustainable yields. Overharvesting of timber resources, both legal and illegal, contributes significantly to deforestation, though deforestation for agricultural expansion may be even greater (whether by large commercial farms in Brazil and Indonesia or by small-

[32]Edward Barbier, "Poverty, Development, Environment," *Environment and Development Economics* 15 (2010), 635–60, tab. 1(b).

[33]Saudamini Das and Jeffrey Vincent, "Mangroves Protected Villages and Reduced Death Toll during Indian Super Cyclone," *Proceedings of the National Academy of Science* 106 no. 18 (2009), 7357–60.

holder farmers in Africa and mainland Southeast Asia).[34] In doing so, it also reduces the value of other local rain forest services. The World Bank estimates that nontimber forest products directly support the livelihoods of approximately 400 million people and indirectly support the livelihoods of approximately 1 billion people.[35] Nontimber forest products include food, fibers, dyes, minerals, and latex and have a total value in international trade of at least $9 billion per year (excluding medicinal products, which are valued at greater than 10 times that figure). Fuel wood is also a critical forest service, particularly for the poor; although it provides less than 7 percent of world energy use, fuel wood accounts for 40 percent of energy use in sub-Saharan Africa (and 90 percent of the use of wood in Africa).[36] Tropical forests are also critical to managing watersheds, helping regulate water flows, cleaning and filtering the water, and preventing the deterioration of soil quality. Tropical rain forests have also been an important component of tourism (ecotourism in developing countries is growing rapidly), are a vital source of biodiversity, and are thought to contain more than half of the land-based species on the planet.

The value of such services is particularly important to the world's poor. Economists have identified three different functions of forest income in rural livelihoods: (1) safety nets because forest products might help to compensate for other production shortfalls, (2) support for current consumption such as food and fuels as noted earlier, and (3) a source of income growth and potential pathway out of poverty. In a review of 51 case studies from 17 developing countries, economist Paul Vedeld and colleagues found that forest environmental income made up an average of 22 percent of total income, with an average absolute value of $678 per household. The proportion of forest income in total income was greatest among the poor, with fuel woods and wild foods together accounting for 70 percent of total forest income.[37] Yet, the ability to derive livelihoods from natural resources also depends on the quality of those resources themselves. Barbier cites evidence from India that poor households in villages near high-quality forests derived up to 40 percent of their income from forest resources, whereas similar households in villages near poor-quality forests derived closer to 10 percent of their income from forest products.[38]

While these data and examples demonstrate the broad association between poverty and the quality of environmental resources, the challenge lies in understanding the mechanisms through which poverty and the environment interact. Early studies, such as the Brundtland Commission, depicted a simple vicious circle in which "poor

[34]Kenneth M. Chomitz et. al., *At Loggerheads? Agricultural Expansion, Poverty Reduction, and Environment in the Tropical Forests* (Washington, DC: World Bank, 2007).

[35]World Bank. *Sustaining Forests: A Development Strategy* (Washington, DC: World Bank, 2004).

[36]Mandar Trevedi, Stavros Papageorgiou, and Dominic Moran, *What Are Rainforests Worth?* Forest Foresight Report Number 4, Global Canopy Program, Oxford, UK, 2008.

[37]Paul Vedeld et al., "Forest Environmental Incomes and the Rural Poor," *Forest Policy and Economics* 9 (2007), 869–79.

[38]Barbier, "Poverty, Development, Environment," p. 645.

people are forced to overuse environmental resources to survive from day to day, and their impoverishment of their environment further impoverishes them, making their survival ever more uncertain and difficult."[39] Subsequent studies suggest a more nuanced view.

Reardon and Vosti provide a useful framework for understanding poverty-environment links in rural areas.[40] Reardon and Vosti summarize their conceptual framework with the flowchart presented in Figure 20–7. Their approach relates four blocks of variables: asset categories of the rural poor, income-generating activities and land use at the household and village levels, categories of natural resources (including soil and water), and conditioning factors such as markets, prices, and infrastructure (or more broadly, institutions) that mediate interactions between the first three blocks of variables.

In this framework, each category of assets represents a potential form of poverty. Asset categories include natural resources that are both privately and commonly held (soil, water, biodiversity), human resources (health, education, labor), on-farm resources (physical and financial), off-farm resources (physical and financial), and community-owned resources (roads, common pasture lands) as well as social and political capital (political power). Households can be poor in any of these categories. Reardon and Vosti emphasize that asset-specific poverty can spill over in various ways to result in environmental degradation. If, for example, land ownership is required to obtain credit, then a household that is poor in land may be precluded from making investments to improve their farm and nonfarm capital (including soil protection). As a consequence, such households may be unable to intensify their farming activities or work away from the farm, perhaps forcing them to rely on expanding their farming activities into fragile lands. Even though a given household may not be poor in the conventional welfare sense, it may still lack sufficient resources to invest in improving the resource base.

The Reardon-Vosti conceptual framework also distinguishes between categories of environmental resources. Soil, water, biodiversity, and air may be either helped or harmed by household activities, depending in part on the value of those resources to rural households. Yet, a potential problem arises when the value of such resources to society at large differs from their value to poor rural households. Reardon and Vosti illustrate this point with the example of a large tree that society may value as a carbon sink but which a poor rural household may see as either timber for sale or a source of ash to enhance soils. As in many such cases of negative environmental externalities,

[39]World Commission on Environment and Development, *Our Common Future* (New York: Oxford University Press, 1987), p. 27 as cited by Barbier, "Poverty, Development, Environment."

[40]Thomas Reardon and Stephen A. Vosti, "Poverty-Environment Links in Rural Areas of Developing Countries," in T. Reardon and S A. Vosti, eds., *Sustainability, Growth, and Poverty Alleviation, A Policy Perspective* (Baltimore: Johns Hopkins University Press, for the International Food Policy Research Institute, 1997).

the issue is one of missing markets. In this case there may be no market for the service of the tree as a carbon sink.

Figure 20–7 illustrates how these four blocks of variables interact with one another. Households pursue their food security and livelihood goals based initially

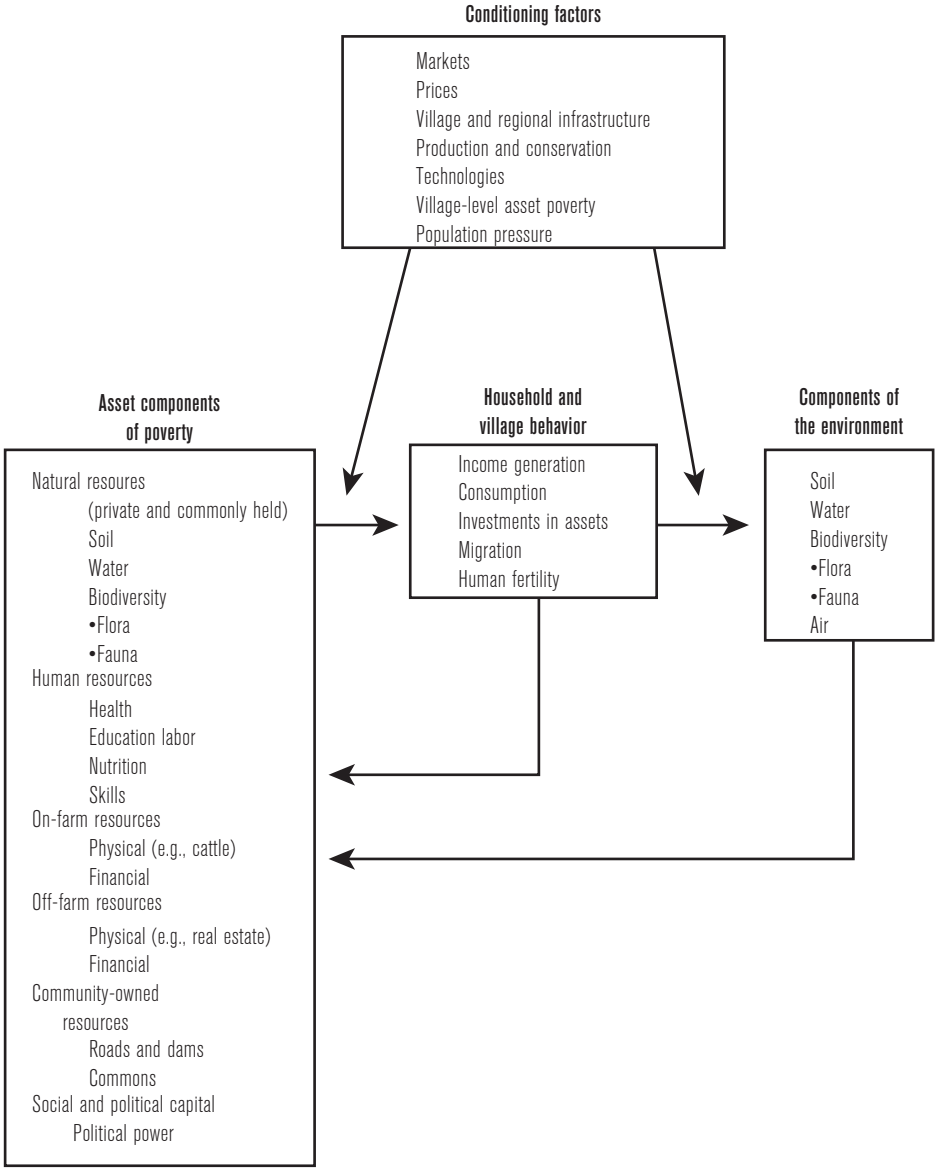


FIGURE 20–7 Conceptual Framework for Poverty–Environment Links

Source: Vosti, S. A. and T. Reardon. 1997. *Sustainability, Growth, and Poverty Alleviation: A Policy And Agroecological Perspective*, Figure 4.1. Washington, DC: International Food Policy Research Institute.

on their stock of assets. Their ability to use these assets in pursuing their objectives depends on a variety of conditioning factors, including accessibility of markets, prices, infrastructure, available technologies, and population pressure. For the poor, community assets and the availability of common pool resources may also play a large role. These factors shape household decisions regarding how to allocate their assets in pursuing their livelihood objectives. For example, the incentives for households to allocate their labor between agricultural and nonagricultural activities (such as working on others' farms) may depend on the relative prices of agricultural output versus local wage rates. Technological choices available to the rural poor may also be limited. Thus subsistence-oriented farmers may be unable to adopt modern inputs to increase yields, forcing them to expand the area cultivated onto fragile lands to increase their agricultural output. Each of these choices has environmental implications. Area expansion might require cutting trees. Working in off-farm activities might reduce pressure on the land, yet might also create incentives to reallocate investment funds away from on-farm land conservation. Diversification of farming activities to include livestock husbandry also has environmental implications, particularly because grazing is often done on public land, where overgrazing is common and can lead to soil erosion and degradation. It does not necessarily follow, however, that it is the poor who contribute the most to this type of environmental degradation. Reardon and Vosti note that it is typically wealthier households that own larger herds and potentially cause greater harm to the environment. The sum total of these environmental effects of household income-earning and investment activities, both positive and negative effects, in turn define the natural resource and asset base available to households in the future.

Reardon and Vosti provide several examples of the complex interactions between poverty and the environment, stressing in particular the dependence of these interactions on the specific type of asset poverty and the specific environmental context. Rwanda is one of the most densely populated countries in Africa. Most of the poor are farmers with little land and few animals. Their primary asset is their own labor. Lacking capital and credit, they are often prevented from investing in soil conservation and unable to intensify their farming practices by purchasing chemical fertilizers and other modern inputs. In such circumstances farmers' main strategy for increasing agricultural output may be simply to increase their own labor input. Yet, in the absence of complementary inputs, potential yield increases are limited, with the likely consequence that the limited land resources of poor farmers in Rwanda quickly become depleted. The result may be a vicious circle in which resource-constrained farmers have little choice but to exhaust their own land, thus increasing their own poverty.

Reardon and Vosti provide a quite different example from the Brazilian rain forest, where the poor may be rich in biodiversity but poor in terms of soil quality and access to capital and labor. Once again, the problem is one of missing markets, in this case a market for biodiversity. The broad range of asset poverty confronting poor

farmers in the Brazilian rain forest may constrain them to cut down trees to expand their farmland and to burn those trees to generate nutrients for the poor-quality soil. Farmers thus substitute biophysical processes for markets to meet their needs. Breaking this cycle may require new agricultural technologies to increase the productivity of existing farmland, improved off-farm opportunities, and the creation of markets for forest cover and biodiversity. Reardon and Vosti note that much of the alarming deforestation of the Brazilian Amazon has been not the result of poverty but rather the response by large landowners to fiscal and other incentives.

The links between poverty and the environment are thus strongly conditional on given circumstances, in particular the specific nature of poverty; the range of available environmental resources; and the institutional, technological, and market conditions governing the choices and activities of households, both poor and rich. Reducing poverty may protect the environment when poverty is driving unsustainable mining of resources. Yet poverty alleviation might contribute to other sources of environmental degradation, such as overuse of agricultural chemicals. At the same time, environmental improvements might reduce poverty that results from the dependence of poor farmers on a depleted resource base. Yet environmental improvement achieved by barring the poor from common pool resources could increase poverty. The conceptual framework presented in this section suggests that a nuanced understanding of the categories of asset poverty and local conditioning variables is necessary to simultaneously reduce poverty and protect the environment. These challenges are writ large in the case of global climate change.

GLOBAL CLIMATE CHANGE

Global climate change may represent the greatest single challenge to sustainable development. While its causes are a subject of debate and predictions of its consequences are uncertain, there is broad scientific consensus regarding the existence of global climate change. As summarized by the Intergovernmental Panel on Climate Change (IPCC) in its fourth assessment report, "Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice and rising global average sea level."[41] The IPCC report notes that the last years of the twentieth century and the first years of the twenty-first century include 11 of the 12 warmest years since 1850 (when systematic recording of global surface temperatures began). Changes in sea level and snow cover are consistent with the evidence of increasing global surface temperatures. Figure 20–8 illustrates long-term changes in global mean surface air temperatures, underscoring the rapid temperature increases since the mid-1970s.

[41]IPCC, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, p. 104.

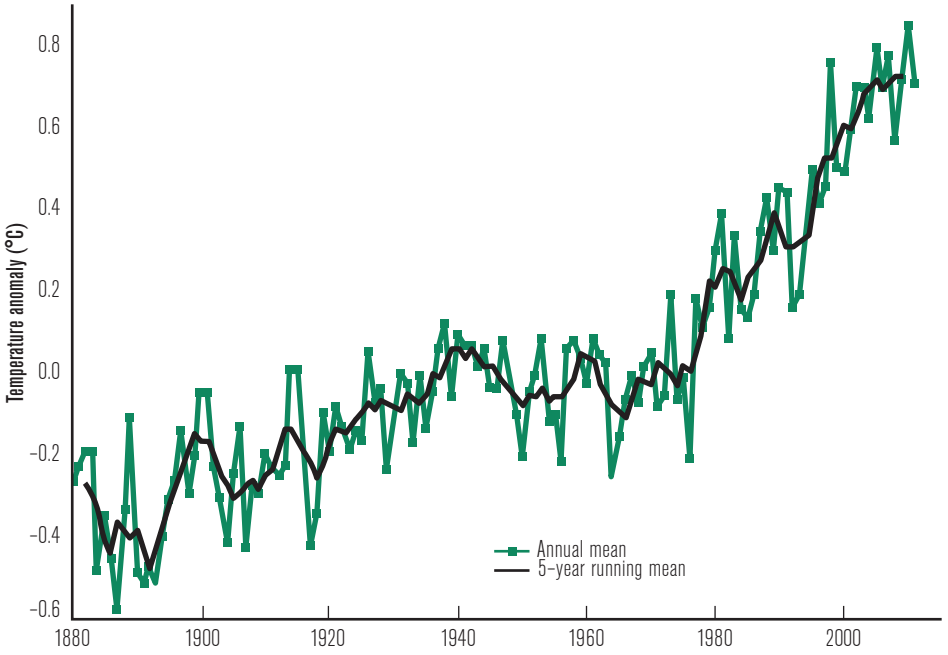


FIGURE 20–8 Changes in Global Annual Mean Surface Air Temperature

Source: National Aeronautics and Space Administration, Goddard Institute for Space Studies, available at http://data.giss.nasa.gov/gistemp/graphs_v3/Fig.A.gif.

The current effects of climate change are apparent from observations of a broad range of natural systems. The IPCC reports major impacts on hydrological systems, such as increased runoff from glacier- and snow-fed rivers and warming of lakes and rivers. They also cite the impact of climate change on biological systems—for example, earlier timing of spring events such as bird migration and reproduction and geographic (poleward) shifts in the ranges of plants and animals in addition to the effects changing aquatic biological systems have on fish migrations and changes in algal, plankton, and fish abundance as well as the health of coral reefs.

Global climate change is driven primarily by the accumulation of greenhouse gases (GHGs) in the earth's atmosphere, where they trap heat. The main GHGs are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and halocarbons. Global concentrations of GHGs have increased substantially since 1750, a result of human activities, according to the IPCC. Agriculture and fossil fuels use have been the leading contributors.[42] Climate change is a central concern for sustainable development owing to the role of human economic activity in the emission of GHGs and the ensuing physical and economic implications of potential changes in the earth's natural resource base.

[42]Intergovernmental Panel on Climate Change, *Climate Change 2007*, p. 37.

The developed world has been the major source of GHG emissions. Average per capita carbon emissions in developing countries are approximately one third the level produced by developed countries. Figure 20–9 illustrates per capita levels of CO_2 emissions as a function of countries' relative population sizes. The height of each bar in Figure 20–9 indicates per capita emissions, and the width indicates population (so the area of each country's bar indicates total emissions). The top 10 countries in terms of CO_2 emissions account for over two thirds of the global total. It is interesting to compare emissions in the United States and China. As of 2006, China overtook the United States as the world's leading source of CO_2 emissions, contributing just over 22 percent of the global total compared with just under 20 percent from the United States. Per capita emissions in the U.S. are approximately four times greater than those in China, yet China compensates for this difference by having a population just over four times greater. International arguments about how to address the problem often start with this observation. There is no simple answer to the question of how efforts to limit climate change should be divided between lifestyle changes in the heavily carbon-emitting rich countries versus creating and paying for the means for poor countries to grow and develop without replicating the carbon dependence of the development path followed by the rich countries. Carbon emissions clearly

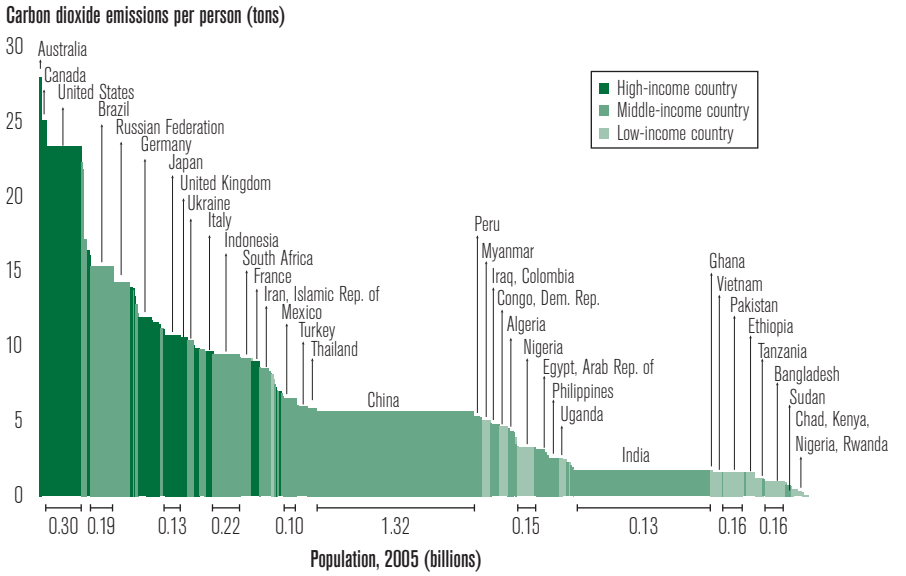


FIGURE 20–9 Per Capita Carbon Dioxide Emissions As a Function of Population Size

The width of each column depicts population and the height depicts per capita emissions, so the area represents total emissions. Per capita emissions of Qatar (55.5 tons of carbon dioxide equivalent per capita), UAW (38.8), and Bahrain (25.4)—greater than the heights of the y-axis—are not shown. Among the larger countries, Brazil, Indonesia, the Democratic Republic of Congo, and Nigeria have low energy-related emissions.
Source: World Bank, *World Development Report 2010: Development and Climate Change.* (Washington, DC: World Bank, 2009), fig. 1.1.

increase with national income. While technological change has the potential to reduce this association, continued income growth around the world seems likely to result in substantial momentum in global emission trends.

While there is growing evidence that some negative impacts of climate change have already occurred, much of the concern about global climate change and its threats to development and human welfare lies in the possible future effects of climate change.[43] Unfortunately, as noted by sportsman Yogi Berra, "It's tough to make predictions, especially about the future." Predictions regarding the future course of global climate change and its effects on human activity are highly uncertain, and scenarios vary widely. The IPCC itself presents six alternative scenarios for temperature change (in °C at 2090–99 relative to 1980–99) ranging from 1.8 to 4.0.

The severity of these impacts varies widely by region as well as within regions. As a general matter, the poor are the most vulnerable; their lives are more exposed to nature and their livelihoods more dependent on it. Africa is particularly at risk. According to the IPCC, global warming of only 1.5°C could increase the number of Africans exposed to water stress by 75 to 250 million; a 2.5°C increase could add 350 to 600 million Africans exposed to water stress and reduce yields from rain-fed agriculture by up to 50 percent. Effects within Africa will vary. For instance, the IPCC predicts that by 2080, arid and semiarid land in Africa will increase between 5 and 8 percent. In Asia, the IPCC projects decreased availability of fresh water in central, south, east, and southeast regions. A 1.5°C increase in average temperatures could decrease wheat and maize production in India by 2 to 5 percent, while an increase of 3°C could decrease China's rice production by 5 to 12 percent. The densely populated river delta regions of South, East, and Southeast Asia are predicted to suffer increased flooding, which will then increase morbidity and mortality from waterborne diseases. In Latin America, global warming greater than about 2.5°C could increase the number of people suffering water stress by 80 to 180 million. Tropical areas of Latin America face the loss of significant biodiversity; indeed, the IPCC suggests that global warming on the order of 4°C could lead to the extinction of up to 45 percent of Amazonian tree species. Increases in temperature and decreases in soil water are projected to transform tropical forest areas in eastern Amazonia into savannas. While soybean yields in the temperate zones of Latin America may increase with global warming, the IPCC predicts increased food insecurity and decreased water availability throughout the region.[44] Global climate change thus poses a direct and fundamental challenge to development. This challenge is complicated by the overarching need

[43]For instance, evidence that climate change resulting from greenhouse gases and other forms of air pollution have significantly reduced the growth rates of rice harvests in India since the mid-1980s is presented in Maximilian Auffhammer, V. Ramanathan, and Jeffrey Vincent, "From the Cover: Integrated Model Shows That Atmospheric Brown Clouds and Greenhouse Gases Have Reduced Rice Harvests in India," *Proceedings of the National Academy of Sciences* 103 (2006), 19668–72.

[44]Auffhammer et al., "From the Cover," p. 50.

to reduce poverty and hunger while avoiding doing so at the expense of continued environmental degradation that might ultimately undermine development itself.

There is great momentum in climate change. The current stocks of GHG emissions in the atmosphere are sufficient to drive continued climate change for decades, even if all future emissions were to cease immediately. According to the World Bank,

> Immediate action is needed to keep warming as close as possible to 2°C. That amount of warming is not desirable, but it is likely to be the best we can do. . . . From the perspective of development, warming much above 2°C is simply unacceptable. But stabilizing at 2°C will require major shifts in lifestyle, a veritable energy revolution, and a transformation in how we manage land and forests.[45]

While estimates vary in their magnitude, there is no doubt that substantial reductions in global carbon emissions relative to the steep trajectory based on business as usual will be required to limit global warming to 2°C.

From an economic perspective, the emission of GHGs into the atmosphere is an example of a negative externality, "the biggest market failure the world has seen," in the words of economist Nicholas Stern.[46] The key distinctions, in Stern's view, between more common externalities and GHG emissions are that GHG emissions are global in both impact and origin, some of the effects may be quite long term, the scientific understanding of the process of global warming and its long-term future impacts is uncertain, and the impacts are potentially immense and irreversible.

Even without agreeing on the details, most analysts agree that responding to the potential threats from global climate change requires simultaneous action on two fronts: **mitigation** and **adaptation**. Mitigation refers to the actions taken to reduce the extent of future climate change, and adaptation refers to the actions required to adjust human activities to the environmental changes that occur. Greater expenditures on mitigation tend to be associated with lower expenditures required for adaptation, and vice versa. Evaluating the tradeoffs between economic growth and environmental protection is a complex and highly controversial exercise. Issues to be resolved with respect to mitigation include choice of a stabilization target for surface temperature increases, when to begin changing behaviors and economic activities to attain that target, whose behaviors and activities to change, the cost of mitigation, and critically, who pays those costs.

Estimates by economists that weigh the costs against the benefits of mitigating climate change vary substantially and have a big impact on conclusions regarding the optimal timing and degree of mitigation. Such calculations are sensitive to a

[45]World Bank, *World Development Report 2010: Development and Climate Change* (Washington, DC: World Bank, 2009),p. 3.

[46]Nicholas Stern, "The Economics of Climate Change [Richard T. Ely Lecture]," *American Economic Review*: *Papers and Proceedings* 98, no. 2 (2008), p. 1. Stern is also the lead author of a major international study of the economics of climate change, *The Economics of Climate Change: The Stern Review* (Cambridge: Cambridge University Press, 2007).

variety of necessary assumptions and choices by analysts. Key parameters are the time horizon covered by the analysis, the sensitivity of global climate increases to given changes in CO_2 concentration, the costs of mitigation, likely damages from climate change, and the choice of the discount rate (important for comparing present versus future costs or benefits). Strong assumptions are required for all of these parameters in order to calculate the **optimal target** for CO_2 concentrations. *Optimal* in this context refers to the concentration level that would result in the least reduction in the present value of global consumption.

Cost–benefit analysis is an essential tool in efforts to design policy responses to global climate change, yet cost–benefit analysis alone cannot resolve the debate. Even if all scientists agreed on the sensitivity of global climate increases to given changes in CO_2 and all economists agreed on the costs of mitigating the associated damages, cost–benefit analysis would still be limited by its inability to assign values to nonmarket environmental goods and services. In addition, choosing the appropriate discount rate to apply to future costs and benefits is a particularly critical component of cost–benefit analysis. Policy prescriptions arising from cost–benefit analyses are highly sensitive to the choice of discount rate, yet there has been little agreement about what rate to apply or even how to choose one. The higher the discount rate, the lower the weight given to the future in comparison with the present. The sensitivity of cost–benefit analysis of climate change mitigation to the choice of discount rate arises because the costs of mitigation occur in the short run while the benefits occur primarily in the long run. One complicating factor is that if incomes grow over time, then future generations would be better able to pay the costs of climate damage, potentially justifying the choice of a lower discount rate today. Ethical concerns are even more complicated, as the discussion is essentially about tradeoffs between the welfare of present and future generations. Is it right to place less importance on the welfare of future generations (as discounting implies)? Conversely, what level of mitigation cost is ethically assigned to the current generation when the benefits are both uncertain and in the distant future?

Two prominent economic studies of climate change have reached quite different conclusions in regard to the optimal timing and extent of mitigation actions, largely owing to their divergent choices of discount rates. The *Stern Review*, which assumes a low discount rate (along with a high sensitivity of climate change to CO_2 emissions and low mitigation costs), concludes that large-scale and immediate interventions to mitigate climate change are essential.[47] In contrast, a study by economist William Nordhaus, which assumes a higher discount rate (along with a lower sensitivity of climate change to CO_2 emissions and higher mitigation costs), argues for more limited and more gradual efforts to mitigate climate change.[48] We can interpret these differing

[47] *The Stern Review.*

[48] William Nordhaus, *A Question of Balance: Waiting the Options on Global Warming Policies*. (New Haven, CT: Yale University Press, 2008).

approaches in terms of their targets for future global climate change. While the more aggressive efforts at mitigation tend to target a 2°C increase in average global temperature, the more gradual approach implies a target closer to 3°C increase. The incremental cost of achieving the lower target (a form of **climate insurance**) is estimated to be only about 0.5 percent of global GDP.[49] Part of the reason the incremental cost of achieving the lower target is not greater is that the lower mitigation cost associated with the higher limit on climate change is partially offset by the greater future costs of adaptation required to address the greater future damage resulting from increased temperatures.

Much of this debate over the urgency and appropriate extent of mitigation hinges on the analysts' choice of discount rates. Critics have assailed the *Stern Review* as having stacked the deck in favor of urgent and substantial mitigation by its choice of a low discount rate. In reviewing this debate, economist Martin Weitzman suggests that the *Stern Review* might be "right for the wrong reasons." For Weitzman, the primary argument in favor of more aggressive mitigation of climate change is not about consumption smoothing and discount rates. He argues that our fundamental uncertainty regarding future impacts of climate change is paramount and that even if utter catastrophe is a low-probability event, mitigation provides an important form of insurance.[50]

Adaptation to climate change refers to deliberate actions taken to reduce the impacts of climate change. In its fourth assessment report, the IPCC summarizes planned adaptation measures by sector. In response to expected water shortages resulting from climate change, potential adaptation measures include expanded rainwater harvesting, improved water storage and conservation techniques, and desalination. The primary constraint to these types of adaptations is financial. Agricultural systems might adapt to climate change by adjusting planting dates, altering crop varieties and the mix of crops produced, and by improving the control of soil erosion by planting more trees. These changes will require both training and technological support for farmers. As discussed in Chapter 17, institutional support for farmers in the form of improved land tenure arrangements may also be necessary to create incentives for improved soil management. Protecting coastal zones from the effects of rising sea levels is another important dimension of adaptation to climate change. The construction of seawalls and storm surge barriers, and the creation of marshlands and wetlands as coastal buffer zones may be critical to protect vulnerable populations in low-lying coastal areas. In this case, land availability and financing may be the primary constraints. Additional measures, such as improved energy

[49]This difference, as summarized in World Bank, *World Development Report 2010*, p. 8, is calculated as the difference between each model's predicted net present value of global GDP versus the net present value of global GDP, assuming no climate change.

[50]Martin Weitzman, "The Stern Review of the Economics of Climate Change [Review]," *Journal of Economic Literature* 45, no. 3 (2007), 703–24.

efficiency and increased use of renewable energy resources are also important planned forms of adaptation to climate change.

The costs and benefits of adaptation at the global level are difficult to estimate precisely, in part owing to the uncertainties noted in this chapter. Two dimensions of the cost are clear. The capacity for adaptation depends significantly on a country's level of development. Adaptive capacity is thus highly unequally distributed around the world. It is also clear that lesser efforts at mitigation in the short run will result in higher costs of adaptation in the long run.

The wide-ranging assumptions (especially regarding the ceiling for global temperature increase) required to estimate the cost of either mitigation or adaptation produce widely ranging results. The estimates for annual mitigation costs in the year 2030 range from $139 to $175 billion (in 2005 US$).[51] The financing requirements can be substantially greater because of the high up-front capital costs of mitigation projects, a potential barrier to participation by developing countries. The range estimated for annual adaptation costs is wider. Estimates for the short term (2010–15) range from $4 billion to greater than $100 billion, and estimates for the medium term (2030) range from $15 billion to $100 billion. At the high end, this funding requirement is less than 0.2 percent of the world's GDP. As of 2010, donor funding commitments for adaptation and mitigation totaled only $9 billion, far short of the estimated requirements.

Global progress in efforts to mitigate and adapt to climate change requires international cooperation. Fears of reduced rates of economic growth and questions of fairness between countries and of how to share the substantial costs have impeded international cooperation to date. Since 1992, international climate negotiations have taken place within the United Nations Framework Convention on Climate Change (UNFCCC), which sets targets for stabilizing atmospheric concentrations of GHGs. The Kyoto Protocol, adopted in 1997 and implemented in 2005, supplemented the original convention. Difficult negotiations failed to secure U.S. participation. Subsequent efforts, such as the Bali Action Plan, adopted in 2007 through the UNFCCC, the 2009 United Nations Climate Change Conference held in Copenhagen (the 15th conference of parties to the Kyoto Protocol), the 2010 United Nations Climate Change Conference held in Cancun, Mexico (the 16th conference of parties to the Kyoto Protocol), and the 2011 United Nations Climate Change Conference held in Bonn, Germany, reflect the continuing difficulties associated with organizing international cooperation in this vital area.

The issues at stake in international climate negotiations are central to the topic of sustainable development. This chapter began by raising the question of whether today's developing countries have the option of following the same energy- and resource-intensive development path previously established by today's wealthy

[51]These estimates are drawn from World Bank, *World Development Report 2010*, table 6.2.

economies. The argument that this path is not sustainable is manifest largely in its implications for the continued atmospheric accumulation of GHGs and climate change. In the context of global climate negotiations, the specific questions revolve around who should pay to fix this problem. Specifically, the Kyoto Protocol set targets for reducing global GHG emissions. These targets are binding only on the relatively wealthy countries (the so-called Annex I countries under the UNFCCC). Unfortunately, 13 years after adoption of the Kyoto Protocol (and 5 years after it came into force), official pledges of emissions reductions offered at Cancun amounted to only 60 percent of the reductions required to limit the rise in global temperature to 2°C.

The position taken by developing countries (ironically led by China, the world's leading emitter of CO_2) is that they should continue to be exempt from commitments to reduce emissions. For the developing countries it is a matter of fairness. After all, they argue, it was the developed countries that largely created the problem over a period of many decades, while the developing countries are both most at risk and least capable of bearing the costs of mitigation and adaptation. On their side, developed countries (the United States in particular) continued to push back, contending that developing countries too should be bound by emissions reduction targets. Despite its signature by President Clinton, the U.S. Senate never ratified the Kyoto Protocol (a legal requirement for U.S. accession to the treaty). This refusal to ratify was based on "the disparity of treatment between the Annex I Parties and Developing Countries," and the fear that a legally binding commitment to reduce GHG emissions would undermine the U.S. economy's international competitiveness. China in particular, with its large trade surplus with the United States, became a focus of U.S. dissatisfaction with the Kyoto Protocol. In essence, the U.S. position rests on the lack of requirement for China to reduce its total emissions, despite the much greater level of per capita emissions in the United States. Ultimately, these issues are highly politicized within both the developed and developing countries, and therein lies a central and continuing impediment to international negotiations on climate change.

SUMMARY

- Natural resource constraints and the cumulative effect of negative environmental externalities raise the possibility that the energy- and resource-intensive path toward development followed by today's advanced economies may no longer be available for today's developing countries. Long-term development must be environmentally sustainable.
- There is a general consensus, based on historical data, that economic growth from a low starting point is associated with increasing levels of pollution; yet beyond some intermediate level of income, economic growth is associated with decreasing levels of pollution. This inverted-U pattern is

known as the environmental Kuznets curve. (The association may not be causal.)

- Sustainable development has been defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."
- Comprehensive wealth (as distinct from GDP) is an accounting approach that incorporates the use of both natural resources and human capital. Changes over time in comprehensive wealth are measured by adjusted net saving. One economic indicator of sustainability is that comprehensive wealth not decline from one generation to the next. This requires that adjusted net saving grows more rapidly than population.
- Much of the negative environmental effect of economic activity results from externalities, social costs not included in private firms' profit-maximizing decisions. Such externalities are particularly prevalent in settings with common property resources.
- Categories of policy solutions to the problem of negative externalities include the creation of property rights where none previously existed, government regulation, taxation of the polluting activity, the creation of marketable permits to pollute, and informal regulation.
- The relationship between poverty and environmental degradation is complex. Most people living in fragile environments are poor, and poor people more frequently rely on the environment for their livelihood. A nuanced understanding of the categories of asset poverty and local conditioning variables is necessary to simultaneously reduce poverty and protect the environment.
- Global climate change may represent the greatest single challenge to sustainable development. We can think of the accumulation of greenhouse gases as a negative externality, and the earth's atmosphere as a global commons.
- Though details of the potential consequences of global climate change are shrouded in uncertainty, it seems clear that concerted international efforts to mitigate climate change are urgent if future economic development is to be sustainable.

Index

Note: Figures are noted with *f*; tables are noted with *t*; boxed topics are noted with *b*.

Abdullah, Rachmina, 3–5, 8, 10
Abdul Latif Jameel Poverty Action Lab (J-PAL), MIT, 286, 451*b,* 543
ability to pay, measuring, 411
abortion, sex-selective, 251*b*
Abramovitz, Moses, 70
absolute advantage, 675
absolute disadvantage, 675
absolute poverty, 181, 182, 187, 213, 214, 215, 218
  in Bangladesh, 185*f*
  determining extent of, 185, 186
  dissenting opinions on, 191
  reducing, 190
  by region, 1981-2005, 190*t*
absorption, 552–53, 558, 560, 561, 562*f,* 564, 568, 570, 577
absorptive capacity, 527
access to world markets, 367
accumulation, population and, 237–38
Acemoglu, Daron, 59*n,* 82*n*
achievement distribution, 623
acute inflation, 428, 447
  among developing countries, 1948-2008, 429–30*t*
adaptation, global climate change and, 796, 798–99
Adelman, Irma, 74*n*
adjustable peg, 434, 434*f,* 435
adjusted net saving gap, 768*b*
adjusted net savings (ANS), 765–67,767*f,* 801
  defined, 765
  for developing regions, 767*f*
  in Ghana, 768–69*b*
ad valorem tariff, 713, 717
advance market commitment (AMC), 331*b*
"advantages of backwardness," 38, 317
adverse selection problem, 134*b*
Afghanistan, 26, 241*n,* 338, 401, 506, 512, 516, 640
Africa, 12, 20, 55, 137, 144, 157, 242, 402, 435, 441, 567, 685, 686, 693
  adjusted net savings in, 766
  agriculture in, 586, 608, 612, 618, 619, 621, 629, 630, 632, 633, 634, 635, 638, 650, 651, 653, 663, 791
  average rank in doing business, Sub-Saharan Africa, 157*t*
  average TFP growth in, 71, 72*t,* 73
  Berg Report and, 146
  cereal yields and poverty in, 612*f*
  child mortality rate in, 48
  civil wars in, 705
  climate change and, 795
  economic growth in, 36, 59–60*b,* 118–19*b,* 135
  education in, 261, 262*f,* 276–77, 284, 285, 287
  emerging countries and income growth in, 1996-2008, 156*t*
  famines in, 311
  foreign aid in, 500, 512, 514, 543
  foreign debt in, 473
  foreign direct investment in, 365
  fragile environments in, 787
  guinea worm control in, 335
  head-count index in, 192
  health in, 310, 321
  HIV/AIDS in, 6, 9, 48, 56, 58, 59*b,* 60*b,* 245, 308, 311*t,* 320, 321, 325, 326, 327, 328, 342, 344
  incomes in, 15
  inequality in, 173
  inflation in, 430, 432
  informal credit markets in, 449
  levels of GDP per capita in, 1820-2008, 35*f*
  life expectancy in, 308, 344
  malaria in, 329
  market reforms in, 155–57
  micro credit in, 449
  population of, 224, 227, 231
  poverty in, 189, 190, 200, 213
  rural *vs.* urban poor in, 611
  state owned enterprises in, 396
  sustainable development and, 770
  total fertility rate in, 225*b,* 226*b*
  trade and, 669, 673, 726, 735, 742
  transport costs in, 86

Africa *(continued)*
under-five mortality rate in, 302
wars in, 76, 77*f*
Washington Consensus and, 148, 160, 161
African Development Bank, 474, 478, 479*b,* 504, 511
African National Congress, 658
age dependency ratio, gross domestic saving and, 379*f*
age-earnings profiles, 267
age pyramids, 299, 300*f*
aggregate capital flows, 495
aggregate production functions, 91–92, 94
Agreement on Textiles and Clothing (ATC), 742, 748
agricultural production function, 592, 592*f*
agriculture, agricultural sector:
agricultural inputs, by region, 639*f*
cell phones and development of, 651–52*b*
climate change and, 798
constraints to development of, 622–27
decollectivization of, 82, 189, 655–56
early, 218
economic ceiling and, 623, 624, 624*f,* 625, 637–45, 647–51
economic growth and, 34–35*b,* 82, 585, 587, 591, 601, 604–10
in Ethiopia, 5, 9*n*
exports and, 726
food crisis of 2005-08, 659–63
food production analysis, 638–45, 647–51
foreign aid and, 521, 535
global climate change and, 793
health and, 310, 321
HIV/AIDS and, 327
industry and, 591–97, 601
inequality and, 174, 178
institutions for development of, 653–56
labor in, 591–97, 642, 643
in Malaysia, 4
model of induced technical change in, 635–37*b*
modern product of labor in, 594*f*
percent of households participating in, deriving income from, 616*f*
population and, 219, 241–42
poverty alleviation and, 610–17
production functions of, 585–86, 592, 592*f,* 596*f,* 600*f,* 643
productivity trends in, 632–35
returns to schooling and technological change in, 279*b*
settled, 621, 622, 663
slash and burn, 621
structural transformation in, 587–90, 588*f,* 605, 617
sustainable development and, 759, 782, 791
systems of, 621–22
technical ceiling and, 623, 624, 624*f,* 625, 627–35
technology and, 586, 627, 630, 633, 635–37*b,* 651–52*b,* 663
in Thailand, 101*b*
trade in, 605, 678*f,* 683
trade protection in, 749–50, 749*t*
traditional, 619, 620–22, 663
in tropical countries, 85
uniqueness of, 584–87, 663
see also land
Ahluwalia, Montek S., 138*n,* 175*n*
Ahmed, Akhter U., 197*n*
Aitken, Brian, 369*n*
Aker, Jerry, 652*b*
Alam, Mahmudul, 266*n*
Albania, 710
albendazole, 290
Alcoa, 685, 685*n*
Alesina, Alberto, 180*n,* 516*n,* 518
Algeria, tax rates in, 412
Ali, Rozali Mohamed, 777*n*
Allende, Salvadore, 559*b*
Alwang, Jeffrey, 620*n,* 621
American Economic Association, 172, 191, 257
American International Group (AIG), 391
Amin, Idi, 479*b*
Amsden, Alice, 729*n*
Anaconda, 685, 685*n*
Angola, 45, 412, 430, 432, 687, 705
distinction between GNP and GDP in, 25, 26
education in, 282
annual compound growth, formula for, 32*n*
Ansu, Yaw, 428*n*
anti-natalists, 237
antiretroviral (ARV) therapy, 16, 324*b,* 326, 328, 342
Apartheid regime, collapse of, 658
Argentina, 84, 122*f,* 124*n,* 135, 432, 435, 556, 573, 574, 665, 711, 784
average farm size in, 653
currency board arrangement in, 154
education in, 181
financial crisis in, 397, 457, 480, 481, 488, 491, 493
financial policy in, 447

Gini coefficients for, 172
inflation in, 428, 430, 558
Washington Consensus and, 148
Aristotle, 39, 40
arithmetic mean, HDI and, 42
Armelagos, George J., 309*n*
Armendariz, Beatriz, 453
Arrow, Kenneth J., 762, 762*n*
Aruba, 12
ASEAN Free Trade Area, 720
Asia, 12, 157, 174, 326
adjusted net savings in, 766
agriculture in, 612, 619, 628, 629, 631, 633, 635, 651, 653, 655
climate change and, 795
customs tariffs in, 414
economic growth in, 15, 118–19*b,* 210
education in, 261, 263–64
financial crises in, 9*n,* 101*b,* 204, 386, 391, 427, 442, 457, 480–81, 480*t,* 488, 489, 491, 496
financial policy in, 441
fragile environments in, 787
head-count index in, 192
HIV/AIDS in, 344
inflation in, 432
levels of GDP per capita in, 1820-2008, 35*f*
micro credit in, 449
multinational corporations in, 364
rural *vs.* urban poor in, 611
total fertility rate in, 225*b*
trade and, 668, 669, 670, 671
see also East Asian tigers
Asian Development Bank, 474, 511
Atje, Raymond, 783*n*
Auffhammer, Maximilian, 795*n*
Australia, 226, 516, 660, 685
top marginal tax rate cut in, 408
Australian model, 546, 546*n,* 547, 549, 552, 553, 564–65, 567–71, 573–75, 576
equilibrium under, 547–48, 548*f*
Greek debt crisis and, 572–73*b*
Austria, 86, 673
autarky, 710, 755
autarky prices, 675, 677, 691
average product, of farm labor, 593
Azerbaijan, 703*n,* 706
Aztec Empire, 34

backward linkages, 134, 686
Baird, Sarah, 291*n*
Baker, James, 472
Baker Plan, 472, 473
balance of payments:
balance of payments identity, 579
financial policy and, 442
import substitution and, 725
national income accounting and, 577–79
savings and, 578–79
balance of payments deficit, 579
balance of trade, 549, 568
Bali Action Plan, 799
Balk, Deborah, 227*n*
Ballinger, Jeff, 746*b*
banana producers, cell phone use and, 652*b*
Banco Bradesco (Brazil), 380
Banerjee, Abhijit, 96, 196*n,* 317, 319, 319*n,* 452*n*
Bangladesh, 12, 138, 180, 181, 504, 524, 620, 748
absolute poverty in, 165, 166, 184–85, 185*f*
distinction between GNP and GDP in, 25, 26
education in, 266, 288
family planning programs in, 248
food riots in, 583
foreign direct investment in, 366
health in, 335
import duties in, 401
income distribution in, 167, 168*f*
microfinance in, 450, 451, 754
oral rehydration therapy in, 340, 341, 342
population of, 231
poverty lines in, 165*n,* 182, 183–84*b,* 184–85, 187, 201
shape of frequency distribution for, 169
size distributions of consumption or income in, 170–71, 170*t*
under-five mortality rate in, 302
Bank Rakyat Indonesia (BRI), 450–51, 453
banks, banking systems, 18
foreign aid and, 511
foreign debt in, 459, 468, 469, 472, 473, 474. 482, 483, 486, 488–89, 491–93, 494, 496
see also savings, saving rates; World Bank
Barbier, Edward, 787*n,* 788*n*
Barder, Owen, 332*n*
Barrett, Ronald, 309*n*
Barro, Robert, 74, 74*n,* 76*n,* 383, 383*n*
barter economy, 423
base, of logarithm, 43*b*
base *e,* 43*b*
basic growth model, 91–94
Bauer, Peter, 500, 500*n*
Bazzi, Samuel, 521*n*
Bear Stearns, 391
Becker, Gary S., 235, 235*n,* 308*n*
bed nets, malaria control and, 317, 330
Behrman, Jere, 278, 278*n*

Belgium, 516
Belize, 112*b*
Bell, Clive, 138*n,* 607*n,* 608*n*
bell curve, 169
benefit-cost ratio. *see* cost-benefit analysis
Bennell, Paul, 277*n*
Berg, Eliot, 146
Berg Report, 146
Berlin Wall, fall of, 33, 148, 149
Bernanke, Ben, 730, 738*n*
Berra, Yogi, 795
Bhagwati, Jagdish, 10
Bhavnani, Rikhil, 521*n*
Bhutan, 86
Bienen, Henry, 701*n*
big bang approach to reform, 151, 158
Big Mac index, 30*n*
big push theory, 132, 134–35
Bihar, 260
bilateral aid, 503–4, 517, 535
Bill & Melinda Gates Foundation, 343, 502
biodiversity, 242, 763, 765, 771, 791, 795
biofuels, food crisis and demand for, 659, 660
biological systems, climate change and, 793
Bird, Richard M., 406*b*
Birdsall, Nancy, 148*n,* 236*n,* 238*n,* 466*n,* 477*n*
birth control, 244
birth rate
  decline in, 233–36
  methods for reduction of, 233
  population and, 220, 221, 222, 245
births, unwanted, 247
black death (bubonic plague), 219
Blackman, Alan, 781*n*
Blalock, Garrick, 368*n*
bliss point, 556
Block, Steven, 608, 608*n*
Blomstrom, Magnus, 368*n,* 369*n*
blood diamonds, 705
Bloom, David, 241*n*
Bluír, Ales, 532*n*
Bolivia, 397, 401, 511, 573–74
  inflation in, 428, 430, 573
  land reform in, 657
  registering property in, 361
  tax administration in, 405, 406*b*
  tax reform in, 408*n,* 409
  trade and, 726
Bolsa Familia Program (Brazil), 207, 215, 288, 290
bonded warehouses, 733, 734
bond markets, 448, 454
Bongaarts, John, 228*n,* 247
Boone, Peter, 533*n*
border price, 713
Borensztein, E., 369*n*
Borlaug, Norman, 628, 632
Boserup, Ester, 242*n*
Boskin, Michael, 418*n*
Bosti, Stephen, 786
Bosworth, Barry, 71, 71*n,* 73
Botswana, 58, 82, 122*f,* 123, 124*n,* 325, 326, 328, 402, 500, 517, 520, 673, 687, 700, 703, 705, 765
  economic growth in, 15, 39, 58, 59–60*b,* 86, 87
  education in, 277
  natural resources in, 370
Bottom Billion, The (Collier), 704
Brady, Nicholas, 472
Brady bonds, 473, 497
Brady Plan, 472, 473
Braga de Macedo, J., 464*n*
Brainard, L., 502*n*
brain drain, 754*n*
Brazil, 84, 135, 342, 397, 434, 473, 546, 556, 560, 561, 573, 589, 709, 711
  agriculture in, 620
  chronic inflation in, 428
  deficits in state enterprises in, 396
  deforestation in, 770, 782, 791–92
  education in, 259, 259*f,* 260, 285, 288
  financial crisis in, 457, 480, 488, 491, 493
  financial policy in, 448
  fiscal policy in, 441
  foreign direct investment in, 365
  health in, 320
  inequality in, 209
  inflation in, 558
  land distribution in, 653
  multinational corporations in, 364, 380
  poverty in, 194
  reining in fiscal decentralization in, 398–99*b*
  trade and, 742, 752
  Washington Consensus and, 148, 155
bribery, 362, 406*b,* 412, 446, 483, 725
British East India Company, 129
British Fabian socialists, 153
British Labour government, 132
British Petroleum, 685
broad money, 424, 425, 425*t*
Brown, Drusilla, 744*n*
Brown, Lester, 242*n*
Brumberg, R., 377*n*
Brundtland Commission, 761, 767, 788
Brunei, 12, 360, 435

Bruno, Michael, 175*n,* 560*n*
Bulgaria, 435
Burger, Joanna, 772*n*
Burger, Kees, 650*n*
Burkina Faso, 87, 112*b,* 583, 589, 608, 683
Burma. *see* Myanmar (Burma)
Burnside, Craig, 534, 534*n*
Burundi, 241*n*
business, barriers to, 359–63
business environment, global, 360*t*
buybacks, 471
Byerlee, Derek, 647*b*

Cairo International Conference on Population and Development, 249
caloric intake
  population and, 241, 241*n*
  see also food; nutrition
Calvo, G., 464*n*
Cambodia, 26, 328
  education in, 282, 288
  quota system and, 748*n*
  total fertility rates by mother's education in, 246*f*
Cameroon, 583
Canada, 226, 415, 685
  NAFTA and, 710, 740
  trade and, 667
Canadian International Development Agency (CIDA), 503
Canning, David, 241*n*
"capability deprivation," alleviating sources of, 40
capital:
  agriculture and, 592, 606, 642
  diminishing marginal product of, 105
  economic growth and, 63–64, 90, 131
  intangible, 762, 764*t,* 765
  mobilization of, in developing countries, 71
  natural, 762, 763, 764*t,* 765, 766
  ownership of, 167
  produced, 762, 764*t*
  shadow prices and, 355
  trade and, 679, 683, 685, 696
capital account, 577, 579, 579*n*
capital accumulation, 88
  economic growth and, 66, 69, 71
  growth models and, 64
  life expectancy and, 77
capital-augmenting technological change, 114*n*
capital deepening, 107, 237, 237*n*
capital flows, 31, 546
capital gains, 415
capital gains taxes, 417
capital-intensive products, 97
capital investment expenditures, 394
capitalism, 132
capital-labor ratio, 95
capital-output ratio, 96
  Harrod-Domar framework and, 96–98, 99
capital per effective worker, 115
capital per worker, 67, 67*f,* 104, 105, 106
capital stock, 94
  changes in, 93, 97
  investment and, 65
capital-to-labor ratios, 99, 100
capital-to-output ratios, 99, 100, 104
capital widening, 107, 237, 237*n*
carbon dioxide emissions, 793, 794–95, 794*f,* 800
carbon sequestration, 787
Cardoso, Jaime, 495*n*
Care, 502
CARs, in Colombia, 776–79
Carson, Rachel, 333
Casanegra de Jantscher, Milka, 406*b*
Cashin, Paul, 691*n*
cash standards, 745
Caves, Richard E., 546*n*
cell phones, 10, 684
  trade and, 681–82
Center for Global Development, 523
Central African Republic, 361, 362
Central America, 685
Central Asia, 36, 174
  mean levels of life satisfaction in, 53
  rate of GDP per population growth (percent/year), 37
central banks, 421
central planning, 152
cereals, 623, 632
  growth rates of yields, 634*f*
  trends in nominal prices of, 659*f*
  yields by region, 633*f*
ceteris paribus, convergence debate and, 121
Ceylon, 172
Chad, 87, 360, 703*n*
  education in, 282
Chávez, Hugo, 665
checking deposits, 424
Chen, Shaohua, 176*n,* 187*n,* 188, 189*n*
Chenery, Hollis, 131, 138*n,* 378
childbearing, as economic decision, 234
child labor, 745
child mortality, income and, 313, 313*f*

children
death rates of, 299, 300, 302, 329, 499, 524
education of, 260–61, 263–64, 267, 287–90
population and, 234
underweight, 301, 344
Chile, 84, 135, 397, 413, 432, 564, 573, 684, 685, 700, 703
chronic inflation in, 428
credit ceilings in, 439
economic growth in, 15, 39
financial policy in, 447
foreign debt and, 495
Gini coefficient for, 175, 176*f*
interest rates in, 445
pension system in, 254
personal income taxes in, 412
pollution registry in, 781
sales taxes in, 417
stabilization in, 559–60*t*
tax reform in, 406, 408*n*
TPS trading program in, 781
trade and, 726
China, 12, 144, 153, 157, 159, 176*n,* 187, 188, 209, 214, 247, 326, 405, 501, 510, 555, 589, 613, 659, 660, 666, 707, 709, 735, 752, 766, 781, 782, 784
agriculture in, 82, 189, 203, 655–56
carbon dioxide emissions and, 794, 800
climate change and, 795
Communist revolution in, 141
corruption in, 362
cost of environmental degradation in, 776
credit ceilings in, 439, 440
economic growth in, 12*n,* 15, 20, 32, 33, 50, 58, 59*b,* 87, 193, 202, 205, 211, 212, 392, 546
education in, 259
epidemiologic transition in, 310
EPZs in, 371
exchange rate policy in, 729–30, 731–32*b*
financial policy in, 441
foreign aid and, 515–16*b*
foreign debt in, 496
foreign direct investment in, 365
foreign exchange reserves in, 387
global competition and rise of, 741–43, 756
head-count index in, 189, 192
health in, 317
income growth in, 211
land reform in, 206
life expectancy in, 307
market reforms in, 152, 160
multinational corporations in, 364, 380
nonfarm rural economy in, 615, 617
one child campaign in, 245, 250–51, 253
one child policy in, 379
PES program in, 779
"planned births" campaign in, 249–50
pollution taxes in, 777
population of, 224, 231
poverty gap lowered in, 190
price system in, 145, 146
reining in fiscal decentralization in, 398–99*b*
reserve requirements in, 439
sales taxes in, 403
savings rate in, 92, 374
son preference in, 251–52*b*
sources of growth in, 71, 73
Soviet model and, 135, 137
state owned enterprises in, 396
subsidies in, 396
surplus labor in, 598–99*b*
trade and, 666, 667, 671, 676, 681, 725, 726, 738, 739, 748
tuberculosis control in, 336–37
undervalued exchange rates and, 724
China National Petroleum, 364
cholera, 312, 340, 341
Chollet, D., 502*n*
Chomitz, Kenneth M., 788*n*
Christophe-Kolm, Serge, 539*n*
chronic inflation, 428
among developing countries, 1948-2008, 429–30*t*
CIMMYT (International Maize and Wheat Improvement Center), 628
civil war, GDP per capita before and after, 76, 77*f*
Clemens, Michael, 50*n,* 517, 521*n,* 533*n,* 534*n,* 754*n*
climate, 85
climate change, 21, 758, 792–800, 801
climate insurance, 798
Cline, William, 467, 739*n*
Clinton, Bill, 665, 800
closed economies, savings in, 92
Club of Rome, 758, 759*n*
Coale, Ansley J., 237, 237*n,* 238, 377*n*
Coase, Ronald H., 773, 773*n*
Coase theorem, 773
cocoa, 139*b,* 146, 402
coffee, 402
cold war, 9, 151, 514, 516, 518
Collier, Paul, 76, 76*n,* 534*n,* 540, 540*n,* 704, 705, 706, 706*n*
Collins, Susan, 71, 71*n,* 73, 467

Colombia, 135, 402, 434, 711, 724
education in, 266, 285
financial policy in, 448
personal income taxes in, 412
taxing water pollution in, 777–78*b*
tax reform in, 408*n,* 409
total fertility rates by mother's education in, 246*f*
trade and, 740, 741
colonialism, 132
Columbus, Christopher, 10, 129
commercial banking system, 421, 422
commercial bank lending, 386, 445, 484
commercial evaluation, 388
commercial project appraisal, 356
Commitment to Development Index, 507–8*b*
commodity prices, 703*n*
nonfuel trends, 1960-2009, 692*f*
trends in, 691–93
commodity taxes, excess burden of, 419*f*
common-pool resources (CPRs), 771–72
common property resources, 655
commons, externalities and, 770–72
Communist revolution, 384
comparative advantage, 674–83, 680, 690, 707, 726, 735
core insight of, 677, 681, 682
production costs and, 675*t*
understanding, 676
comparative trade theory, 683
competition:
global, and rise of China and India, 741–43
market failure and, 133*b*
markets and, 149–50
resource traps and, 705
trade and, 680, 707
completion point, 478
compound growth, formula for, 61*b*
comprehensive wealth, 762, 765, 801
concessional assistance, 502
concessional terms, 386
conditional cash transfers (CCTs), 206–7, 215, 288, 288*b,* 289–90
conditional convergence, 123–25, 128
conditionality, 537–40, 543
Congo, Democratic Republic of, 45, 75–76, 77*f,* 156, 241*n,* 360, 430, 432, 475, 500, 527, 653, 684, 705, 765
natural resources in, 370
population of, 231
constant prices, 352
constant returns to scale, 96, 126
constant-returns-to-scale production function, 96*n*
Consultative Group on International Agricultural Research (CGIAR), 628, 632
consumer price index (CPI), 25*n*
consumers, consumption:
food prices and, 611–12
foreign aid and, 529–30, 529*f*
government saving and, 382
household, measuring, 166
household saving and consumption, 376–79
savings and, 65–66
trade and, 669–71, 680–81, 683–86, 694, 710, 726, 740, 742
consumer subsidies, 395
consumer surplus, 714
consumption cost of the tariff, 714
consumption per capita, 186
continuous compounding, basic equation for, 62*b*
contraception, 248
failure in market for, 244
family planning and, 246–49
convergence
conditional, 123–25, 128
of income, 121–25, 209
unconditional, 121, 125
Cooper, Richard, 495*n*
Corbo, Vittorio, 560*n*
Corden, W. Max, 546*n,* 564, 564*n,* 694*n,* 699*n,* 716*b*
Corla, Jessica, 781*n*
corporate income taxes:
fiscal policy and, 411
incidence problem and, 414–16
rate increases and, 404
corporate saving, 379–80
corruption, 82, 140, 142, 145, 155, 156, 159, 207, 358, 361–62, 388, 527, 539, 706, 723, 725
Cortés, Hernán, 34
Costa Rica, 396, 473, 727
mean levels of life satisfaction in, 53
cost-benefit analysis, 351, 388
effects of shadow pricing on, 357*t*
crawling peg exchange rates, 434, 434*f,* 435, 560–61
credit ceilings, 439–40, 454
credit default swaps, 422
creditor panics, 486–87*b,* 488–89
Croatia, 328
Croitoru, Leila, 775*n*
crude birth rate, 220, 222, 224, 233
crude death rate, 220, 222, 224

Cuba, 76, 326, 328, 514, 546
  tax rates in, 412
Cull, Robert, 453
curb (or kerb) market, 449
currency, 424
currency boards, 154, 434–35, 495, 495*n*
current account, 577, 580
current account deficit, financing, 579
current account surplus, in China, 731*b*
Cutler, David, 316, 316*n*

Dalgaard, Carl-Hohan, 522*n*
dams, 784–85
Das, Saudamini, 787, 787*n*
Dasgupta, Partha, 762*n,* 786*n*
Dasgupta, Susmita, 759, 759*n,* 760
Das Kapital (Marx), 130
Datt, Gaurav, 205*n*
Daud, Fatimah, 3*n*
Dawe, David, 610, 610*n*
DDT, 333
deadweight loss, 714, 721
Deardorff, Alan, 744*n*
deaths, death rates:
  diarrheal disease and, 340–42
  health and, 299, 300, 300*f,* 302, 304, 306, 310
  malaria and, 329
  population and, 220, 221, 222, 224
Deaton, Angus, 31*n,* 51*n,* 191, 191*n,* 241*n,* 316, 316*n,* 378
debt crises, 395, 423, 576
  in Greece, 2010-2012, 571–73*b*
  see also foreign debt
debt crisis of 1980s, 466–73, 497
  causes of, 467–69
  escape from, for some countries, 470–73
  impact of, on borrowers, 469–70
debt-equity swaps, 472
debt/export ratio, 462
debt/exports, 461, 462
debt forgiveness, 17
debt/GDP ratio, 460, 462, 463
debt indicators, 460–63
debt overhang, 464
debt ratios, 471–72, 471*t*
debt repayment crisis, 567–68, 569*f*
debt restructuring, 493–94
debt/revenue (or NPV debt/revenue), 461
debt service, 460
debt service/exports, 461, 462
debt service/revenue, 461, 462
debt sustainability, 459–60
decision point, 478
decline in the terms of trade, 567–68, 569*f*
deep financial strategy, 447–48
Deese, Brian, 466*n,* 477*n*
defaults, 463–73
defensive lending, 477
deficits, 203, 388, 394–95, 396, 437
deforestation, 765, 771, 791–92
DeGregorio, J., 369*n*
Deininger, Klaus, 176*n,* 655, 655*n,* 657, 657*n*
de Janvry, Alain, 611, 611*n,* 613*n,* 615
de la Torre, Augusto, 148*n*
de Lesseps, Ferdinand, 322*b*
Delgado, C., 607*n,* 608*n*
demand deposits, 424
demand elasticities, price, 403
Demirguc-Kunt, Asli, 453
democracy, foreign aid and, 518
demographic dividend (or bonus), 238, 241, 241*n*
demographic gift, 112*b*
demographic transition, 220–24, 232, 238, 241, 254, 389
  for less-developed regions, 223*f*
  stages in, 377–78
Denison, Edward, 70
Denmark, 505, 507*b*
  health in, 299–302, 300*f,* 301*t,* 306, 310
Department for International Development (DFID), 503
dependency ratio, 238, 378
depreciation:
  changes in capital stock and, 93
  investment and, 65
  reduction of capital per worker, 106
deregulation, Washington Consensus and, 147
de Soto, Hernando, 83, 83*n*
devaluations, 571, 695, 701, 702*b*
  trade and, 547
Devarajan, Shantayanan, 91*n,* 546*n*
developed countries. *see* industrialized countries
developing countries, 3–13, 20, 70, 71
  adjusted net savings for, 767*f*
  agriculture in, 622
  capital flows to, 386
  capital inflows to, 387
  climate change and, 800
  debt flows to, 456–57, 457*t*
  debt ratios in, 471–72, 471*t*
  diminishing marginal product of capital and, 120–21
  disease in, 323
  economic recovery programs in, 139*b*

fertility decisions of families in, 236
financial flows from and to, 387*t*
fiscal policy and, 392
foreign debt in, 463–64
foreign direct investment in, 365
global trade and, 742
health in, 317
import duties in, 401
inflation in, 429–30*t,* 432
long-term debt in, 467*t*
malaria in, 329–20
multinational corporations in, 364, 380
new growth models and, 127
open economies in, 546
population of, 220, 224, 227, 228
scarcity of education in, 259–60, 259*f*
sustainable development in, 758
tax systems in, 416
technological innovation and, 113
technology transfers and, 373
trade in, 83, 84, 135–36
development, economic, 20, 54, 545, 685
approaches to, 16–17, 40
commodity exports and, 704
continuum of, 13–14
critical triangle of goals in, 785*f,* 786
debate over role of government in, 137–40
definition of, 14, 39
diversity in, 15–16
economic growth *vs.,* 13–14, 23–24, 38–39
education and, 257, 293
fertility outcomes and, 246
foreign aid and, 518–35, 543
globalization and, 8–10
government and, 129–30, 142, 735
industrialization and, 11
measuring, 40–44
natural resource exports and, 707
population and, 236–38, 240–41, 255
post-World War II, 130–32, 134–37
poverty and, 202*b*
study of, 15–16, 18–19
temporary migration and, 753–55, 756
two-sector models of, 590–602
what is meant by, 38–40
see also growth, economic
Development Assistance Committee (DAC), 502, 508*b*
development expenditures, 394
Diamond, Jared, 34–35*b,* 35*n,* 757, 757*n,* 758, 758*n,* 761
diamonds, 402
diarrheal disease, 340–42, 345, 524
diet, health and, 323. *see also* food; nutrition
digital provide, 652*b*
diminishing marginal product of capital (MPK), 117, 120–21, 120*f*
diminishing marginal returns, 643
diminishing returns
agricultural production and, 591
production function and, 117–21
diminishing returns to capital, Solow's model and, 105, 522
directly observed treatment, short course (DOTS), 334, 336, 337, 343
direct taxes, 399
Dirksen, Everett, 499
disability, 306, 309
discounting, discount rates, 352
shadow discount rates and, 356
sustainable development and, 797–98
discounting future benefits, of education, 270, 270*n*
disease (morbidity), 309, 310, 312, 344
double burden of, 344
population growth and control of, 220
disequilibrium, 562*f,* 576
equilibrium and, 556–60
four combinations of conditions tied to, 556*f,* 557–60, 576
in open economies, 556–60
disguised unemployment, 591
disposable income, 561
dividends, 415
Diwan, Ishac, 467
Djibouti, 77*f,* 435, 673
Doha Round, 752–53, 756
Dollar, David, 83, 84*n,* 198*n,* 516*n,* 518, 534, 534*n,* 540, 540*n*
dollarization, 435, 495
Domar, Evsey, 96, 96*n,* 103, 103*n*
see also Harrod-Domar growth model
domestic content requirements, 373
domestic prices, 715*b*
domestic savings, 374–75, 376, 389
Dominica, 209
Dominican Republic, 371, 396, 734
Dornbusch, Rudiger, 462*n,* 470*n*
double burden of disease, 344
doubling time, 222
calculating, 62–63*b*
drought, macroeconomic reflection of, 574, 575*f*
Due, John F., 403*n,* 411*n*
Duflo, Esther, 96, 196*n,* 277, 278, 278*n,* 294*b,* 317, 319, 319*n,* 452*n,* 525, 525*n*

Duloy, John H., 138*n*
Dutch disease, 527, 527*n,* 564–65, 565*f,* 567, 693–703, 704, 708
  avoiding in Indonesia, 702–3*b*
  geometric presentation about, 697–99*b*
  in Nigeria, 700–701*b*
  preventing, 701–3
duty exemption systems, 733
Duvalier, Jean-Claude, 476*b,* 520

earnings. *see* wages
earthquakes, macroeconomic reflection of, 574
East Asia, 36, 55, 58, 160, 711
  economic growth in, 117
  education in, 262, 262*f*
  land reforms in, 205–6
  outward-oriented trade policies and, 83
  population of, 224
  rate of GDP per population growth (percent/year), 37*n*
  sources of growth in, 72*t*–73*t*
  total fertility rate in, 226–27
  Washington Consensus and, 148
East Asian tigers, 153, 741
  government and development of, 142
  influence of, 144
  rise of, 141
  trade and, 726
  see also Hong Kong; Singapore; South Korea; Taiwan
Easterlin, Richard, 51, 318*b*
Easterlin paradox, 51
Easterly, William, 45*n,* 102, 102*n,* 244*n,* 280*n,* 500*n,* 533*n,* 539*n*
Eastern Europe, 149, 150, 151, 160, 173, 326, 546
  economic growth in, 33
  inequality in, 177, 178
  levels of GDP per capita in, 1820-2008, 35*f*
  poverty in, 189, 194
Eastern Europe and Central Asia (ECA), poverty in, 189
ecocide, 757
economic ceiling:
  in Mosher framework, 623, 624, 624*f,* 625
  raising, 637–45, 647–51
Economic Commission for Latin America (ECLA), 136
economic efficiency, 639
  budget and, 417–20
economic evaluation, 388
economic globalization, defined, 10
economic growth. *see* growth, economic
economic institutions, 80*b*
economic isolationsim, 710
economic miracle, 37
economic opportunity costs, 355
economic project appraisal, 356
economics, important questions in study of, 55
Ecosystems and Human Well-Being, 786
ecotourism, 788
Ecuador, 396, 473, 687
  financial crisis in, 480, 493
  sales taxes in, 417
education, 257–98
  absenteeism and, 288*b,* 291, 293–94, 294–96*b,* 298
  accountability in, 293–94, 296–97, 298
  attainment in, adult population and, 259*f,* 262*f*
  benefits of, 257–58, 267
  class size and, 287, 291–92, 292*n*
  controversies of, 278–80
  costs related to, 270, 277, 282, 287, 291
  deworming and, 290–91
  discounting future benefits of, 270, 270*n*
  dropout rates and, 287
  economic growth and, 42, 56, 257, 258, 267–69, 274–76, 280, 298
  enrollment rates and, 259
  in Equatorial Guinea, 14
  estimated rates of return for, 272–73, 298
  in Ethiopia, 5
  externalities of, 271
  first-generation estimates in, 272–76
  flows in, 259
  foreign aid and, 522–26, 525–26, 543
  gender distribution in, 263–64, 263*t,* 276, 289*b*
  for girls, fertility and, 245
  government and, 296–97
  gross enrollment rates and, 260–61, 260*t*
  health and, 312, 320, 345
  in Hungary, 14
  improvement of, 286
  inefficient use of resources for, 290–93
  internal rates of return and, 271, 276
  investment in, 77–79, 88, 90, 117, 126, 127, 257, 267–69
  learning outcomes, 264–67, 265*f*
  in low-income countries, 15
  macroeconomic puzzles and, 280
  in Malaysia, 4
  microeconomic puzzles and, 278–79
  misallocation of resources in, 282, 284–85
  poverty and, 205, 206

productivity and, 281, 298
public spending on, 282, 283*f*
rate of return in, 269–71, 272, 298
reducing costs related to, 287–90
resources for, 288–89*b,* 290
returns to schooling and income opportunities, 279–80*b*
returns to schooling by level and country income group, 273*t*
schooling *vs.,* 264–67
second-generation estimates and, 276–78
social returns of, 271, 272*f,* 273*t*
sources of growth in East Asia and other regions, 1960-2000, 72*t*
stocks in, 259
student retention, 287, 288
teacher absences and, 293–94, 294–96*b,* 298
textbooks in, 292–93
total fertility rates and, 246*f*
trends and patterns in, 258–59
in Ukraine, 6–7
underinvestment in, 282
Educo program (El Salvador), 297
Edwards, Alejandra Cox, 254*n*
Edwards, Sebastian, 378, 384, 495*n,* 695*n*
effective rates of protection (ERP), 715–16*b*
effective taxes, 412
effective units of labor, 114
efficiency, 133*b,* 138
productivity and, 64
efficient tax system, 418–19
Egypt, 187, 396, 564, 784
economic growth in, 58, 89
education in, 181
food riots in, 583
oral rehydration therapy in, 340, 341, 342
polio in, 338
total fertility rates by mother's education in, 246*f*
trade and, 733, 734
Eichengreen, Barry, 496*n*
elasticities:
of demand, 441
demand, taxation and efficiency, 419*f*
fiscal policy and, 418–419, 419*f*
of isoquants, 636*n*
perfect, 594, 594*n*
rice production and, 649
"elementary functionings," 39
Elliot, Kimberley Ann, 745*n,* 747*n*
el Rafie, M., 341*n*
El Salvador, 77*f*
civil war in, 297
education in, 42, 297
HDI of, 44
life expectancy in, 41–42
emergency relief, foreign aid and, 526
emerging economies, 11
subsidies and, 721
emerging market financial crises, 480–96
creditor panic and, 486–89
domestic economic weaknesses and, 482–84
five Asian economies: private external financing before/after, 480*t*
foreign debt and, 480–82
GDP growth before/after, 481*t*
international financing for selected crisis countries, 492*t*
lessons from, 494–96
short-term capital flows, 484–86
short-term foreign debt and reserves, 485*t*
stopping panics, 489–91, 493–94
employment, 204
generation of, FDI and
shallow finance and, 447
trade and, 740
see also labor; unemployment
Employment Guarantee Scheme (EGS), 208, 215
Encina, Francisco, 411
End of Poverty, The (Sachs), 213
endogeneity, class size and, 292*n*
endogenous growth models, 126–27, 128
endogenous to the model, nontradable prices and, 547
energy pricing, policy failure and, 784
Engel, Ernst, 587, 689
Engel, Stefanie, 779*n*
Engel's law, 587, 617, 689
England, 172
Enos, John, 533, 533*n*
enterprise funds, 504
environment, 16
common resources and, 771
dams and, 784–85
economic growth and, 759–61
foreign aid and, 522–26
fragile lands and, 787, 789
global climate change and, 792–800
marketable permits and, 779–81
poverty and, 785–92, 790*f,* 801
sustainable development and, 757–59, 775–76, 782–84
taxes, subsidies and, 776–79
technology and, 759
environmental Kuznets curve (EKC), 759–61, 760*f*

epidemics, pandemics *vs.,* 323*n*
epidemiologic transitions, 309–10, 344
Equatorial Guinea, 14, 687, 703*n*
equilibrium, 576
  disequilibrium and, 556–60
  in open economies, 546–63, 556–60, 562, 565, 567
equity, fiscal policy and, 411, 419, 420
Eritrea, 5, 241*n,* 519, 672
Essay on the Principle of Population (Malthus), 232, 758
estimated rates of return, for education, 272–73
Estonia, 435
  economic growth in, 15, 39
Eswaran, Mukesh, 601, 601*n*
Ethiopia, 8, 9*n,* 71, 87, 325, 509, 525, 574, 651, 672–73
  agricultural growth multiplier in, 608
  average farm size in, 653
  land reform in, 657
  population of, 231
  poverty in, 5–6
  total fertility rates by mother's education in, 246*f*
ethnicity, poverty and, 194
Eurasians, historical dominance of, 34–35*b*
Europe, 36, 84, 151, 226, 387
  education in, 262*f*
  financial crisis of 2008 and, 422, 494, 545
  foreign borrowing in, 458
  Marshall Plan and, 131
  mean levels of life satisfaction in, 53
  rate of GDP per population growth (percent/year) in, 37, 37*n*
  recession of 2008 and, 391
  total fertility rate in, 225*b*
European Bank for Reconstruction and Development, 511*n*
European Commission, 508
European Union (EU), 402, 511, 646*b,* 681, 720, 747
  financial crisis in, 455
  free trade in, 710
  Greek debt crisis and, 571*b*
Evenson, Robert, 628
excess burden, 403–4, 417, 418
exchange-rate anchor, 559, 559*b,* 573–74
exchange-rate regimes, 560
exchange rates, 453
  in China, 731–32*b*
  continuum of prototypes with, 434*f*
  conversion problems and, 27–32
  determining, 28
  domestic currency and, 694
  management of, 722–24
  monetary policy and, 433–35
  nominal, *see* nominal exchange rates
  overvalued and undervalued, 722, 723, 723*f,* 724
  real, *see* real exchange rates
  trade and, 738
  undervaluation of, 557
  see also foreign exchange
excise taxes, 402–3
exogenous to the model, world market prices and, 547
Expanded Program on Immunization, 337
expenditures, government. *see* government expenditures
explicit costs, children and, 234
export credits, 386
export orientation, 711, 725–30, 733–34
export pessimism, 688–90, 708
export-processing zones (EPZs), 4, 371, 730, 733–34, 755
export promotion, 725
  see also outward orientation
exports, 667, 668, 669, 670, 671, 674, 676, 679, 685, 686, 687, 691, 694, 695, 703, 711, 724
  favoring, 728–30
  manufactured, importance by region, 670*f*
  removing bias against, 727–28
  world, growth by region, 1974-2008, 668*f*
export taxes, 402
external balance (EB), 548, 555*f,* 556*f,* 557*f,* 558, 561, 562, 570, 576
external diseconomies, 772*f*
externalities, 778, 779, 801
  commons and, 770–72
  of education, 271
external transfer problems, 460
Extractive Industries Transparency Initiative (EITI), 705, 706, 708

factor accumulation, 87, 116, 117, 127
  definition of, 64
  economic growth and, 66, 67–68, 67*f,* 74, 87
Falcon, Walter, 584, 584*n,* 623*n*
Fallon, Peter, 532*n*
family planning, 246–49, 255
famine, 310, 311
Fan, S., 659*n,* 660, 662
farm-gate price, 650
farming, mixed, 622
farming systems, 621. *see also* agriculture

favelas, 194*n*
FDI. *see* foreign direct investment
Feder, Gershon, 655, 655*n* 657, 657*n*
Federal Reserve System, 18, 141, 422, 438*n*
Feed the Future (USAID), 541
Feenstra, Robert C., 577*n*
Fei, John, 592, 592*n*
Fei-Ranis model, 597, 603*b,* 605, 618
Feldstein, Martin, 491*n,* 496*n*
feminization of poverty, 195
Fenwick, A., 319*n*
Ferreira, Susana, 766, 766*n*
fertility, 308
  authoritarian approaches and, 249–51, 253, 255
  microeconomic model of, 243
  modern theories related to, 235–36
  negative relation to income, 235
  population and, 224–27, 227, 233, 234, 240–41, 245–46, 255
  total rates of, by mother's education, 246*f*
fertilizer subsidies, in Malawi, 645–47*b*
feudalism, 14
Feyzioglu, Tarhan, 531*n*
fiat money, 423
Field, Christopher B., 772*n*
Fields, Gary, 172*n*
Figenschou, B. M., 319*n*
Fiji, 705
Finance and Development, 202*b*
financial account, 577, 579, 579*n*
financial assets, 424
financial crises, Washington Consensus and, 148
financial crisis of 1997-1998, 391
financial crisis of 2008, 386, 391, 421, 422, 427, 442, 455, 494, 666, 692
  foreign reserves in China and, 730
  India and, 153
  poverty and, 193
financial deepening, 443, 454
financial development, 442–53
financial intermediation, 423, 426
financial panics, 454
financial policy, 422, 423
  exchange rates in, 722–24
  foreign debt in, 458–64
  inflation in, 435–38, 442–51
  prices in, 546–49, 552–56, 558–59
financial system, 421
  financial intermediation, 426
  functions of, 423–27
  money and money supply in, 423–26
  stabilization and, 427
  transformation and distribution of risk in, 426–27
Findlay, R., 464*n*
Finland, 224*n,* 233
  demographic transition in, 220–22, 221*f*
  education in, 264
Firestone, 685, 685*n*
fiscal austerity, 561, 562, 573, 574
fiscal equity, 411
fiscal linkage, 686, 699
fiscal policy, 18, 391–420
  changes in tax administration, 405
  corporate income taxes, 404, 414–16
  economic efficiency and the budget, 417–20
  fundamental tax reform, 405–11
  government expenditures, 393–97, 399–401
  income distribution in, 411–16
  international trade, taxes and, 401–2
  luxury consumption taxes, 413–14
  macroeconomic equilibrium and, 576
  new sources of tax revenues, 404–5
  in open economies, 549, 561, 567, 663
  personal income taxes, 404, 412
  property taxes, 414–16
  purpose of, 392
  sales and excise taxes, 402–4
  stabilization and, 560, 561
  taxes in, 399–414
Fischer, Hannah, 516*n*
Fischer, Stanley, 491*n*
fisheries, 652*b,* 770, 781
Fishlow, Albert, 462*n*
Fiszbein, Ariel, 207*n*
fixed-coefficient production function, 94–96, 106
fixed-coefficient production technology, isoquants for, 95*f*
fixed exchange rates, 434*f,* 435, 436, 438, 453, 495, 497, 560, 568
fixed-exchange-rate system, 433
fixed nominal exchange rate, 559
fixed rent schemes, 655
fixed-rent tenancy, 655
flexible-exchange rates, 438
floating exchange rates, 434*f,* 436–37, 454, 495, 560, 568
flows, in education, 259
flypaper effect, 530
Fogel, Robert, 241, 320, 320*n,* 608–9*b*
food, 232, 517, 759
  contribution of modern varieties of, 631*t*
  exemption from sales taxes, 413, 420

food *(continued)*
income elasticity of demand for, 689
population and, 220, 241*n,* 242–43
poverty and, 182, 183–84*b,* 196, 196*n*
projecting future demand for, 622*n*
trade in, 683
see also agriculture, agricultural sector; nutrition
food aid, food production and, 527, 528*b*
Food and Agricultural Organization (FAO), 524*b,* 584, 662
food crisis of 2005-08, 659–63
causes of, 659–61
consequences of, 661–63
Food for Schooling program (Bangladesh), 288
food insecurity, climate change and, 795–96
food prices, 605, 611, 612, 663–64
farmers' incentives and, 620, 663
increases in, 583–84
industrial labor supply and, 597
national investment and, 609–10
food production analysis, 638–45, 647–51
factor-factor decision, 641–43, 641*f*
factor-product decision, 643–45, 643*f,* 647–50
market access, 651
product-product decision, 638–40, 639*f*
food riots of 2008, 583, 609
food stamps, 395
Forbes, Kristin, 180*n*
Ford, Henry, 126
Ford Foundation, 136
Ford Motor Company, 126
foreclosures, 391
foreign aid, 11*n,* 499–544
agriculture and, 521, 535
commercial ties and, 517–18
Commitment to Development Index and, 507–8*b*
conditionality in, 537–40, 544
coordination in, 541–42
country selectivity in, 534, 540–41
country size and, 517, 543
democracy and, 518, 543
dependency and, 531–33
description of, 501–2
diversity of opinions on, 499–500
donor relationships and, 535–43, 544
donors of, 503–4, 508–12
Dutch disease and, 694
economic growth and, 518–35, 519*f,* 523*f,* 543
foreign policy objectives and political alliances, 514, 516
fungible, 520–31, 543
harmonization in, 541–42
health and, 341, 343, 522–26, 543
historical role of, 503
income levels, poverty and, 516–17
major official development aid recipients, 2009, 513*t*
motivations for, 514, 516–18
net official development assistance, 2009, 505*f*
objectives of, 518
official development aid receipts by region and income, 514*t*
recipients of, 512–13, 541, 544
results-based management and, 542–43, 544
savings and, 374, 529–31, 543
tradables and, 553
types of, 534–35
foreign assets, sovereign wealth funds, 705–6
foreign assistance, 501
foreign debt, 457
advantages and disadvantages of, 458–59
default in, 463–73
emerging market financial crises and, 480–82
1980s debt crisis and, 141, 466–73
short-term, 485, 485*t*
sustainability of, 459–60
foreign direct investment (FDI), 363–74, 369*n,* 374, 384, 386, 388, 458
access to world markets and, 367
attracting, policies aimed at, 371–73
benefits and drawbacks of, 365–69
categories of, 365
controversy tied to, 364
defined, 363
in electronics industry in, 367
generation of employment and, 366
growth and, 369–70
patterns and products relative to, 364–65
policies toward, 370–74
requirements and restrictions on, 373–74
transfer of technology and, 367
Washington Consensus and, 147
foreign exchange, 694
Dutch disease and, 695–96
gap in, 131
reserves, 387–88
shadow prices and, 355
windfalls, 564, 565*n, 567*
foreign income tax credit, 410
foreign investment
in China, 152

tradables and, 553
trade and, 685–86
foreign policy, foreign aid and, 514, 516
foreign savings, 374, 384–88, 385, 389, 462, 579
forestry, forests:
climate change and, 795
mangroves, 787
policy failure in, 782–84
sustainable development and, 765, 770, 771, 788
tropical rainforests, 787–88
forward linkages, 686
fossil fuels, 758, 793
Foster, Andrew, 279*b*
Fox, Vicente, 665
France, 12, 113, 372, 510, 514, 516, 709, 748
education in, 259, 284
foreign aid through, 504–5
Frankel, Jeffrey A., 83, 83*n,* 546*n,* 736, 736*n,* 737
Frankenberger, Tim, 197*n*
freely floating exchange rates, 434*f*
Freeman, Richard B., 745*n,* 747*n*
free trade, 681, 710
Free Trade Area of the Americas, 665
frequency distribution, 167, 169
Friedman, Milton, 142*n,* 376*n,* 500
fundamental tax reform, 405–6, 408–11
fungible aid, 520–31
Furman, Jason, 491*n*
futures markets, 150

Gabon, 378, 412
gains from exchange, 679, 680
gains from specialization, 680
gains from trade, 679
Gallup, John, 85*n,* 198*n,* 332
Gambia, 401, 402, 686
Gandhi, Indira, 249
Gang, Ira N., 406*b*
Gardner, B., 608*n*
Gardner, Gary, 242*n*
Gates, Bill, 499, 499*n*
GDP deflator, 25*n*
Geary-Khamis (GK) dollars, 33*n*
Gelb, Alan, 701*n*
gender, poverty and, 194–95, 214
gendercide, 252*b*
General Agreement on Tariffs and Trade (GATT), 136, 510*n,* 750–51
General Agreement on Trade in Services (GATS), 753
genuine saving, 765
geography:
agricultural production and, 586
economic growth and, 84–87, 88, 90,117,119*b*
GDP and, 86*f*
income levels and, 85*f*
trade and, 671–74, 737
geometric mean, HDI and, 42, 44–45
Georgia, Republic of, 45, 77*f,* 92, 360, 430
education in, 282
Germany, 113, 135, 172, 222, 510, 514, 555
fall of Berlin wall, 148, 149
foreign aid through, 505
total fertility rate in, 226
trade and, 667
Gerschenkron, Alexander, 38, 129*n,* 317
Gersovitz, M., 378
Gertler, Paul, 248, 248*n,* 368*n*
Getachew, 5–6, 8, 10
Ghana, 137, 138, 375, 402, 432, 448, 478, 500, 511, 527, 558, 564, 589, 656, 781
acute inflation in, 428
deforestation in, 770
education in, 259, 286
independence in, 139–40*b*
informal credit markets in, 449
log export bans in, 784
Malthusian effect of population growth on ANS in, 768–69*b*
recovering from mismanagement in, 566*b*
Ghana Cocoa Board, 146
Gillis, Malcolm, 402*n,* 403*n,* 408*n,* 784*n*
Gini, Corrado, 171
Gini coefficient, 171–72, 171*n,* 175, 176, 181, 209, 214, 215, 653
for Chile and Taiwan, 175, 176*f*
by country and region, 173*f*
global, estimates of, 211
for South Africa, 177
girl's education, fertility and, 245, 255
Glennerster, Rachel, 332*n,* 452*n*
Glewwe, Paul, 261*n,* 281*n,* 292
Global Agreement on Tariffs and Trade (GATT), 711
Global Alliance for Vaccines and Immunisations, 332*b*
global annual mean surface air temperature, changes in, 793*f*
global climate change, 792–800, 801
cost-benefit analysis of, 797
Global Fund to Fight AIDS, Tuberculosis, and Malaria, 343, 541
global inequality, 215
end of poverty and, 208–14

globalism, globalization, 116, 681, 741
backlash against, 153
controversy over, 665–66
defined, 10
development and, 8–10
industrial policy and, 734–36
global official development assistance, 1975-2008, 509*f*
Global Polio Eradication Initiative, 338
global trade, rapid growth in, 9
global trade agenda:
expanding market access, 747–50
multilateral trade negotiations and WTO, 750–53
rise of China and India, 741–43
sweatshops and, 743–45
temporary migration and, 753–55
global warming, 16, 133*b,* 758
Goldstein, Morris, 496*n,* 731*n*
Gollin, Douglas, 589, 589*n,* 628
Gomez-Granger, Julissa, 516*n*
goods and services:
nontradable, *see* nontradable goods and services
tradable, *see* tradable goods and services
Gorbachev, Mikhail, 149
Gorgas, William, 322*b*
Goswami, Omkar, 406*b*
Goulder, Lawrence H., 762*n*
governance:
economic growth and, 79, 80*b*–81*b,* 82, 88, 90, 117, 138
investments and, 358
primary product exports and, 704–5
government, 686
agriculture and, 645
default by, 465*b*
disaster response and, 574
economic development and, 142
economic growth and, 39
economic progress and, debate over, 129–30, 140
education and, 284, 296–97
expenditures of, *see* government expenditures
hard, 158
health and, 321
import licenses and, 717
investment and, 350, 354, 356, 371
Korea, declining effectiveness of intervention by, 143–44*b*
revenue, consumption, and saving, 1990–2002, 382*f*
self-discovery and, 735–36
soft, 158
stopping panics and, 489–90
sustainable development and, 774–76, 782
trade and, 136, 673–74, 718
welfare weights and, 356–57
government expenditures, 393–411, 568
categories of, 394–97, 399
commodity booms and, 699–700
equity and, 420
in fiscal policy, 392–93
foreign aid and, 531, 532
intergovernmental transfers and, 400–401
preventing Dutch disease and, 701
recurrent, 394–95, 397
SOEs and, 396, 399
taxes and, 399–401
government failure, 134*b*
Washington Consensus and, 148
government policy, inequality and, 178
government revenues. *see* taxation, taxes
government savings, 380–84, 385, 389
grade survival rates, 261
Graham, Edward, 368*n,* 369*n*
Grameen Bank, 449, 450, 451*b,* 504
grand bargain, 751
grants, 386
Greaney, Vincent, 266*n*
Great Depression, 132, 136, 159, 391, 456, 465*b,* 490, 710, 711, 750
Great Recession (2008), 8, 9
Greece:
debt crisis of 2010-12, 571–73*b*
doing business in, 360
financial crisis in, 455, 458, 464, 511
greenhouse gases (GHGs), 133*b,* 758, 771, 793–94, 796, 800, 801
Green Revolution, 500, 521, 619, 627, 628–32, 631*t,* 635*b,* 663
Gresser, Edward, 748*n*
Gribble, James, 248*n*
Griffen, Keith, 533, 533*n*
Griliches, Zvi, 70
gross domestic expenditure, 552
gross domestic product (GDP), 32, 62*b,* 131, 280
agriculture and, 584, 587, 604, 618, 633
in Angola, 25
in Bangladesh, 25
broad money as percentage of, 425–26, 425*t*
central government expenditure as share of, 393*t*
comparing levels in two economies, 28–30, 29*t*

comparing per capita using market exchange rates and PPP in 2009 (US$), 31*t*
debt sustainability and, 460
deficits in state enterprises and, 396
defined, 24, 577
financial crises and, 481*t*
flaws and benefits of, 26–27
foreign debt and, 470
geography and, 86*f*
government saving and, 381
growth, conditional on population growth and, 124*f*
index number problem and, 32
investments and, 357
levels per capita by region: 1820-2008, 35*f*
in open economies, 552
per capita, 36, 38, 53, 77*f*
population growth and, 237
poverty and, 199
primary product exports and, 687
production functions and, 91
rate of, per population growth, 37*t*
share of agriculture in, and rural population share, 585*t*
sustainable development and, 762
trade and, 83, 667–69, 737
unconditional growth and, 122*f*
gross domestic savings, 65
age dependency ratio and, 379*f*
gross enrollment rates, 260–61
gross national expenditure, 577
gross national income (GNI), 12, 13, 13*t,* 42, 42*n,* 209*n,* 501, 505, 512, 513, 517, 543, 578
education and, 284
net official development assistance as percentage of donor GNI, 2009, 506*f*
per capita by country, HDI *vs.,* 46*f*
gross national product (GNP), 11–12
in Angola, 25
in Bangladesh, 25
defined, 24
gross savings, 765, 766
growth, economic, 8, 12, 15, 20, 54
in Africa, 36
agriculture and, 585, 587, 591, 601, 604–10
capital and, 131
in China, 12*n,* 32, 33
climate change and, 799
deep finance and, 448
desirability of, 50–51, 53, 54
divergent patterns of, 56–60
domesticated agriculture and, 34–35*b*
as dynamic process, 682–83
in Eastern Europe, 33
economic development *vs.,* 13–14, 23–24, 38–39
education and, 42, 56, 257, 258, 267–69, 274–76, 280, 298
effective governance and institutions and, 79–82
environment and, 759–61
factor accumulation and, 60, 64, 66, 87
favorable environment for private enterprise and, 82–83, 88, 90, 117
foreign aid and, 518–35, 519*f,* 523*f,* 543
foreign debt and, 469–70
foreign direct investment and, 369–70
geography and, 84–87, 88, 90, 117
health and, 319–20
in India, 12*n,* 33
inequality and, 174, 180
inflation and, 76, 76*n*
investment and, 65, 66, 90, 349, 388
investment in health and education and, 77–79, 88, 90, 117
labor and, 90
landlocked nations and, 673
in Latin America, 33, 36, 154*t*
macroeconomic stability and, 75–76
MDGs and, 50
measuring, 24–25
modern, 14
natural resource abundance and, 687, 688*f*
new approaches to, 125–27, 128
1970-2010, 36–38
nutrition linkage to, 608–9*b*
openness and, 83–84
political stability and, 75–76, 88, 90, 117
population growth and, 112*b*
poverty and, 166, 197–98, 197–201, 203
primary exports and, 687–88
problems with, 198
productivity and, 60, 64, 66, 68, 87–88, 90
savings and, 65–66, 90, 126
Solow model and, *see* Solow growth model
sources of analysis of, 66–74, 67*f*
strong governance and institutions in, 88, 90, 117
sustainable, 66, 74
sustainable development and, 758
theories of, 89–128
trade and, 83–84, 707, 736–39, 756
in various countries, 1960-2009, 57*t*
world-wide, 32–33, 36–38
growth accounting, 68
growth multipliers, 607

growth rate of output, 97–98
growth rates
  of any variable, 237*n*
  calculating, 61*b*–62*b*
  explaining differences in, 118*b*–119*b*
  Solow growth model and, 111, 113
growth theory, 89–128
Guan Zhong, 257, 258
Guatemala, 375
Guerin, Bill, 401In
Gugerty, Mary Kay, 198*n*
Guinea-Bissau, 92, 191
Guriev, Sergei, 53*n*
Guyana, 328, 671

Haberger, Arnold C., 428*n*
hacienda systems, 657
Hae, Lee Kyang, 665
Haiti, 12, 244, 500, 527, 574, 583, 681
Hall, Robert, 85*n*
halocarbons, 793
Halweil, Brian, 242*n*
Hamann, Javier, 532*n*
Hamilton, Kirk, 766, 766*n*
Hansen, Henrik, 522*n*
Hanson, Gordon, 738*n*
Hanushek, Eric, 292, 292*n*
happiness, incomes and, 51–53, 762
Haq, Mahbub ul, 40, 539, 539*n*
Harberger, Arnold C., 418*n,* 435*n*
hard governments, 158
Hardin, Garret, 243*n*
hard loans, 386
Harrison, Ann E., 369*n,* 736, 736*n,* 737, 738*n,* 740, 746*b,* 747*n*
Harrod, Roy, 96, 96*n*
Harrod-Domar growth model, 94–100, 102–3, 107, 109, 110, 116, 127, 128, 130, 159, 457, 520, 520*n*
  basis of, 96
  capital-output ratio and, 96–98
  economic growth in Thailand and, 102*b*
  fixed-coefficient production function and, 94–96, 102, 106
  rigid production function in, 99, 103
  shallow financial strategies and, 447
  strengths and weaknesses of, 98–100, 102–3
Hartwick rule, 765
Harvard Development Advisory Service, 136
Harvard Institute for International Development, 136
Hayami, Yujiro, 635*b* 637*n*
Hazell, Peter, 607*n,* 608*n*
head-count index, 185, 186, 189
Headey, Derek, 659*n,* 660, 662
Heady, Derek, 157*n*
health, health conditions, 299–345
  affordability and, 342–43
  causes of disease and death, 309, 311*t*
  challenges in, 343–44
  critical diseases and, 321–34
  definitions of, 302–3
  determinants of, 310–12
  economic growth and, 316–17, 319–20
  education and, 312, 320, 345
  epidemiologic transitions and, 308–10, 344
  foreign aid and, 341, 343, 522–26, 543
  global, transitions in, 307–8
  income and, 314–17, 319, 344
  investment in, 77–79, 88, 90, 117, 127, 320–21
  leadership and, 342
  lessons learned for, 342–43
  life expectancy and, 304, 305–6*b*
  in low-income countries, 15
  medical services and, 301–2, 315
  poverty and, 205, 206, 313–14, 344
  productivity and, 319–20
  successes in, 335–43
  in Ukraine, 7
health-adjusted life expectancy (HALE), 306
Heavily Indebted Poor Countries (HIPC) Initiative, 477, 479*b*
Heckscher, Eli, 677, 682
Heckscher-Ohlin model, 677
hedge funds, 421
Hee, Park Chung, 143*b*
Heller, Peter, 531*n*
Helms, Jesse, 499, 499*n*
Hersch, Joni, 747*n*
Hicks, Joan Hamory, 291*n*
Hicks-neutral technological change, 114*n*
high-income countries, 11, 12, 13*t,* 20
  central government expenditure as share of GDP in, 393*t*
  composition of tax systems in, 400*t*
  consumer subsidies in, 395
  consumption habits in, 212
  doing business in, 360
  education in, 262*f,* 264, 265
  government saving in, 381
  health in, 309
  population of, 224
  tertiary education enrollment rates in, 259
  trade and, 671
high-powered money, 437
high-yielding crop varieties, 628

Hill, Ruth Vargas, 197*n*
Hinkle, Lawrence, 552, 552*n*
Hirschman, Albert O., 132, 134, 134*n,* 605, 686, 686*n*
HIV/AIDS, 9, 16, 48, 205, 212, 276, 304, 309, 311*t,* 323, 324–25*b,* 327*f,* 330*b,* 332, 344, 345, 474, 752
  in Africa, 6, 9, 48, 56, 58, 59*b,* 60*b,* 245, 308, 320, 321, 325, 326, 327, 328, 342, 344
  preventing in Thailand, 336
  in tropical countries, 85
Hobbes, Thomas, 411
home country, 677, 679
Honduras, education in, 277
H1N1 swine flu, 309
Hong Kong, 141, 160, 435, 711
  financial policy in, 448
  foreign direct investment in, 365
  free trade in, 710
  government's role in development of, 142
  liberalization and, 728, 728*n*
  semiconductor industry in, 4
  trade and, 668, 726, 741
  see also East Asian tigers
Hoover, Edgar M., 237, 237*n,* 238, 377*n*
Hoover Institution, 292
Hopkins, J., 607*n,* 608*n*
horizontal spillovers, 368
household consumption, measuring, 166
Hoxha, Enver, 710
human capital, 114
human capital investments, 257, 258
human development, defined, 41*b*
Human Development Index (HDI), 40–42, 42*n,* 43*b,* 44–46, 46*f,* 54, 79, 181
Human Development Report, The, 41, 197*n,* 198
humanitarian aid, 526, 534
Hungary, life expectancy in, 14
hunger-gatherers, 218
hurricanes, 574
Husain, Ishrat, 467
Hussein, Saddam, 476*b*
Hutterites, 233
hyperinflation (runaway inflation), 154, 427, 428, 430, 573
  in Peru, 430, 431*b*

identification, empirical studies and, 273, 277
Iglesias, Augusto, 254*n*
illiquid borrowers, 461, 486*b*
illiquid economies, 484
immigration, population growth attributable to, 224, 226, 231
immiserizing growth, 200
implicit costs, children and, 234
import controls, 436*n*
import duties, 401
import quotas, 714, 717–18
imports, 667, 668, 669, 671, 673, 674, 676, 681, 686, 691, 724, 725, 727, 740
import substituting industrialization, 134, 135, 140
import substitution (IS), 17, 414, 674, 711–12, 723, 726, 735, 755
  definition for, 674
  ERPs and, 715–16*b*
  exchange rates in, 722–24
  import quotas and, 712, 714, 717–18
  outcomes of, 724–25
  subsidies and, 720–22
inactivated polio vaccine (IPV), 339
incidence problem, corporate income and property taxes, 414–16
income distribution, 14, 199, 214, 416
  in Bangladesh, 167, 168*f*
  Gini coefficient and, 171–72
  Lorenz curve and, 170–71, 170*f*
  luxury consumption taxes, 413–14
  measurement of, 166–67, 169–72
  in Mexico, 166, 167, 168*f*
  taxes and, 411–16
  in United States, 167, 168*f*
income growth:
  per capita, 61*b*
  shallow finance and, 447
income redistribution, progressive income taxes and, 412
incomes, 15
  in agriculture, 588, 615
  child mortality and, 313, 313*f*
  converting into "capability to live minimally acceptable life," 40
  differential rates of economic growth and divergence in, 36
  economic growth and, 39, 55, 58, 800
  education and, 278, 279*b*
  fertility negatively related to, 235
  foreign aid and, 516–17
  GNI and, 45
  happiness and, 51–53, 762
  health and, 312, 314–17, 319, 344
  HIV/AIDS and, 327
  inequality and, 166–72
  per capita, 43*b*
  pollution and, 759–60
  sources of, 616*f*

incomes *(continued)*
tax collection and, 400–401
transfers of, 206–7
war and, 76, 77*f*
see also wages
incomes, per capita, 14, 24
exchange rates and, 27
expressing in logarithms, 33*n*
growth in, 61*b*
worldwide, 32–33, 36–38
income taxes:
corporate, *see* corporate income taxes
countries reducing highest rates of, 1984-2009, 409*t*
fiscal policy and, 411–16
personal, 412
increasing returns to scale, 126
incremental capital-output ratio (ICOR), 97, 98, 100, 102, 127
India, 12, 27, 89, 129, 159, 172, 176, 187, 188, 211, 326, 342, 360, 509, 510, 659, 660, 673, 687, 752, 766, 781, 784
agricultural growth multiplier in, 608
climate change and, 795
economic growth in, 12*n,* 15, 20, 33, 58, 193
education in, 259, 260, 293–94, 294–96*b,* 297
foreign aid in, 512
foreign direct investment in, 365
forest resources in, 788
global competition and rise of, 741–43, 756
health in, 317
impact of schooling on farm profitability in, 279*b*
income growth in, 211
informal credit markets in, 449
land reform in, 657
market exchange rate *vs.* purchasing power parity methods of converting GDP in, 28–30, 29*t*
market reforms in, 153, 160
multinational corporations in, 364, 380
national family program in, 247
nonfarm rural economy in, 615
oral rehydration therapy in, 340
polio in, 338
population of, 224, 231
poverty in, 189, 194, 196
poverty line in, 191
sales taxes in, 403
sex-selective abortions in, 252*b*
smallpox in, 338
sources of growth in, 71, 73
sterilization campaign in, 249
tax administration in, 405, 406*b*
tax reform in, 408*n*
trade and, 681, 725, 726, 739, 740, 741
under-five mortality rate in, 302
Indian Congress Party, 132
indirect taxes, 399
Indonesia, 20, 82, 137, 138, 158, 187, 189, 396, 397, 401, 405, 413, 432, 434, 468, 500, 564, 567, 568, 589, 687, 700, 703, 705, 782
credit ceilings in, 439, 440
deforestation in, 770, 782
Dutch disease in, 702–3*b*
economic growth in, 15, 89
education in, 259*f,* 266, 277, 294
financial crisis in, 386, 457, 480, 481, 488, 489, 491, 494
financial policy in, 440, 448
ICOR in, 98, 100
inflation in, 445
labor activists and labor outcomes in, 746–47*b*
log export bans in, 784
micro credit in, 451
natural resources in, 370
personal income taxes in, 412
policy failures and reforestation in, 783*b*
population of, 218
reserve requirements in, 439
rice price stabilization in, 610
tax reform in, 406, 407–8*b,* 408*n,* 409
trade and, 726, 729, 739, 742, 748
under-five mortality rate in, 302
Industrial and Commercial Bank of China, 380
industrialization:
development and, 11
environment and, 759, 760
health and, 323
inequality and, 174
Soviet approach to, 135
through import substitution, 17
industrialized countries, 11, 20, 70, 71
agricultural interests in, 749–50
climate change and, 800
greenhouse gas emissions and, 794
multilateral trade agreements and, 751, 756
multinational corporations in, 364
industrial labor, supply and demand for, 595*f*
industrial policy, trade strategy and, 734–36
Industrial Revolution, 232, 770
population and, 219, 220
industry, industrial sector
agriculture and, 591–97, 601

backward linkages and, 134
inequality and, 174
infant, *see* infant industries
linkages and, 132, 134
population and, 220
sustainable development and, 759, 779–81
trade and, 689, 720–22, 726, 728, 734–36
inefficiencies, sources of, 417–18
inelastic demand, taxation and efficiency and, 419*f*
inequality, 138
economic growth and, 166, 174–77, 180
importance of, 178–80
international, 209, 210*f*
intrahousehold, 167
Lorenz curve and, 170–71, 170*f*
measuring, 166–67, 167*n,* 169–72
patterns of, 172–74
poverty and, 180, 203
infant industries
trade and, 712
trade protection and, 718
infant mortality rate, 304, 308, 317
in Botswana, 59*b*
economic growth and, 56
in Ethiopia, 6
gender and, 251*b*
in low-income countries, 15
in Malaysia, 4
poverty and, 192*b*
in Thailand, 101*b,* 102*b*
infectious disease, 16, 34, 318*b,* 321, 323, 344, 345
inflation, 145, 150, 392, 423, 427–32, 453, 555, 558
credit ceilings and, 439–40
deep finance and, 448
economic growth and, 76, 76*n*
episodes of, 428, 429–30*t,* 430–32
foreign aid and, 527
in Latin America, 154*t*
macroeconomic stabilization and, 563
monetary policy and, 141
nonmonetary causes and, 438
poverty and, 203–4
rate of, 12*n*
reserve requirements and, 439
sources of, 435–38
stabilization and, 569, 570–71, 570*f,* 573–74
transition to market economy and, 151
Washington Consensus and, 442
informal credit markets, 449, 454
information, to investors, 371
information asymmetries, market failures and, 133*b,* 134*b*
information revolution, 16
infrastructure:
capital investments and, 394–95
corruption and, 361, 362
foreign aid and, 520, 535
investment, policy failure and, 784–85
liberalization strategy and, 728
market access and, 650, 651
opportunity costs and, 355
population and, 242
poverty reduction and, 205
infrastructure linkages, 686
innovation possibility curve, 636*b*
input prices, effect of tariffs on, 715*b*
insolvent borrowers, 461, 486*b*
insolvent debtors, 461–62
institutionally fixed wages, 593
institutions, 19
definition for, 653
economic growth and, 88, 90, 117
effective governance, institutions, and role of, 79, 80–81*b,* 82
investments and, 358
insurance, 422
insurance markets, information asymmetries and, 134*b*
intangible capital, 762, 764*t,* 765
Intel, 367
intellectual property rights, 751
intensive annual crops, 622
Inter-American Development Bank, 508, 511
intercropping, 638
interest, interest rates, 145
crisis economies and, 491
deficits and, 394–95
in financial policy, 444, 445
foreign debt and, 462, 463, 468
inflation and, 141
recessions and, 442
soft loans and, 386
Washington Consensus and, 147
see also discounting, discount rates
intergenerational equity, 761–62
Intergovernmental Panel on Climate Change (IPCC), 792, 793, 795, 798
intermediate goods, 24
internal balance (IB), 548, 555*f,* 556*f,* 557*f,* 558, 561, 562, 570, 576
internal rates of return (IRR), 271, 352
internal transfer problems, 460
international agricultural research (IARCs), 628

International Bank for Reconstruction and Development (IBRD), 11*n,* 508, 510
see also World Bank
International Centre for Settlement of Investment Disputes (ICSID), 508
International Commission on the Measurement of Economic Performance and Social Progress, 26*n*
International Comparison Program (ICP), 30, 31, 32, 187
international debt, combating recessions and, 441–42
international debt crises, first wave of, 1980s, 36
International Development Association (IDA), 459, 508, 509, 515*b*
International Finance Corporation (IFC), 508
International Food Policy Research Institute (IFPRI), 196, 632
international inequality, 209, 210
population-weighted measure of, 211
trends in, 1961-2008, 210*f*
International Institute for Applied Systems Analysis (Austria), 227*n*
International Labour Organization (ILO), 745
International Maize and Wheat Improvement Center (CIMMYT), 628
International Monetary Fund (IMF), 25, 30, 139*b,* 146, 202*b,* 206, 351, 374, 395, 474, 477, 491, 503, 508, 511, 512, 526, 537, 538, 540, 541, 545, 566*b,* 646*b,* 750
credit ceilings and, 440
establishment of, 490
foreign debt and, 472, 479*b*
founding of, 510
in Greece, 2010-2012, 571–73*b*
stabilization program, 570, 571
international prices, deriving, 30
international reserves, 433, 436
International Rice Institute (IRRI), 628
international trade, 83–84, 88
taxes on, 401–2
terms of, 677
International Trade Organization (ITO), 510*n,* 750
Internet, 10, 16
intersectoral linkages, 605, 606, 618
intestinal worms, school children and, 290–91
inverse elasticity rule, 419
investment, investing
barriers to doing business and, 359–63
deep finance and, 448
domestic savings and, 385
economic growth and, 65, 66, 90, 350, 388
in education, 77–79, 88, 90, 127, 257, 267–69
foreign, *see* foreign investment
foreign aid and, 529–30, 529*f*
foreign direct investment, 363–74
governance, institutions and, 82
government and, 354, 356, 371
in Harrod-Domar framework, 99
in health, 77–79, 88, 90, 127, 320–21
inequality and, 179
opportunity costs and, 355
present value and, 351–54
preventing Dutch disease and, 701
price inflation and, 204
productive public and private, barriers to, 357–59
productivity of, 351–57
in research and development, 126
savings and, 374–88, 389
shadow prices and, 355–56
stability of food prices and, 609–10
sustainable economic growth and, 74
technology adoption and, 116
trade and, 9, 666, 701, 745
inward-oriented trade policies, slow growth and, 83
Iran, 137, 360, 657, 685
Iraq, 475, 506, 516
Ireland, financial crisis in, 458, 511
Irwin, Douglas, 681*n*
Isham, Jonathan, 534*n*
Islamic Development Bank, 511
Isle of Man, 12
isocost line, 637*b,* 642
isolationism, economic, 710
isoquants, 95, 96, 100, 103, 104*f*
for fixed-coefficient production technology, 95*f*
output in, 94
isorevenue line, 638, 639, 640
Israel, 564
Italy, 222
doing business in, 360
financial crisis in, 455
total fertility rate in, 226
Ivanic, Maros, 662, 662*n*
Ivory Coast, 195, 401, 589

Jakarta, 440
Jamaica, 76, 413, 685, 687
education in, 292
tax reform in, 408*n,* 409
James, Estelle, 254*n*

Japan, 12, 141, 160, 177, 226, 384, 387, 491, 508*b,* 510, 555, 666, 709, 710, 720, 747
convergence debate and, 121–22
economic growth in, 131
foreign aid and, 504, 517
land reform in, 657
life expectancy in, 41
population of, 231
sales taxes in, 417
semiconductor industry in, 4
total fertility rate in, 226
trade and, 671, 681
Japan Fund for Poverty Reduction, 288
Japan International Cooperation Agency (JICA), 503
Java, 85
jawbone control, 441
Jayachandran, Seema, 477*b*
Jensen, Robert, 652*b*
Johnson, Rachel, 404*n*
Johnson, Simon, 59*n,* 82*n*
Johnston-Mellor linkages, 606–7, 608, 618
joint ventures, 373
Jolly, Richard, 138*n*
Jones, Charles I., 85*n,* 103*n*
Jones, Christine, 406*b*
Jones, Ronald W., 546*n*
Jordan, 222, 516
Jorgenson, Dale, 70, 599, 599*n*
jute, 402

Kagame, Paul, 651–52*b*
Kanbur, Ravi, 539, 539*n,* 542*n,* 754*n*
Kapur, Ishan, 566*n*
Kauffmann, Dani, 81*b*
Kaufmann, Daniel, 82*n,* 534, 534*n*
Kazakhstan, 366, 703*n*
Keefer, Philip, 82*n*
Kelley, Allen, 45*n,* 236*n*
Kelly, V., 607*n,* 608*n*
Kenny, Charles, 50*n,* 308*n*
Kenya, 76, 187, 404, 426, 448, 650, 673, 744
doing business in, 361
education in, 259, 287, 291, 292, 297, 320
Gini coefficients for, 172
savings rate in, 107
textile industry in, 724
trade and, 733, 734
Khandker, Shahidur, 266*n*
Kharas, Homi, 5*n,* 502*n*
Kinder, Molly, 335*n,* 524*n*
King, Elizabeth, 276*n*
Kinnan, Cynthia, 452*n*
Klasen, Stephan, 252*n*
Klein, Michael W., 495*n*
Knack, Stephen, 82*n*
knife-edge problem, 99–100
Koch, Robert, 312
Korea, 176*n,* 362, 405, 500, 568, 687, 711
economic growth in, 15, 39, 59*b,* 205
education in, 264
EPZs in, 371
financial crisis in, 386, 457, 480, 481, 488, 563
land reform in, 657
mean levels of life satisfaction in, 53
revisionist school on, 729*n*
tax reform in, 405
trade and, 668, 671, 726, 728–29, 730, 738, 741
Kotwal, Ashok, 601, 601*n*
Kouri, P., 464*n*
Kraay, Aart, 81*b,* 82*n,* 198*n*
Kraft, Joseph, 467
Kremer, Michael, 261*n,* 281*n,* 287, 291*n,* 292, 293*n,* 294, 332*n,* 477*b*
Krueger, Anne O., 434*n,* 726*n*
Krugman, Paul, 464*n,* 676, 676*n*
Kuala Lumpur, 78
Kurihara, K., 377*n*
Kuvyenhoven, Arie, 650*n*
Kuzawa, Christopher W., 309*n*
Kuznets, Simon, 14, 172, 172*n,* 174, 175, 236, 236*n,* 759
Kuznets curve, 175, 176, 177, 177*n*
Kyoto Protocol, 799, 800
Kyrgyz Republic, education in, 266

labor, labor sector, 88
activism in Indonesia, 746–47*b*
agriculture and, 588–89, 589*n,* 591–97, 606, 642, 643
in Cambodia, 26
change in, 93
economic growth and, 63–64, 90
effective units of, 114
FDI and, 366
health and, 319–20, 321
inequality and, 174
land reform and, 657
poverty and, 196
productivity of, 115–16, 119
in rural areas, 585, 588, 591, 593
in semiconductor industry, 4–5
shadow prices and, 355
taxes and, 18
temporary migration and, 753–55, 756

labor, labor sector *(continued)*
trade and, 676, 683, 685, 686, 696, 742–47, 753–55
see also employment; marginal product of labor; wages
labor accumulation, economic growth and, 66, 69
labor augmenting technology, 114
labor-intensive products, 97
labor-output ratio, 96
labor-saving technologies, 635*b*
labor supply, 94
labor surplus, 591
labor surplus model, 591–97
labor-to-output ratios, 99, 100
land, 683, 685
economic growth and, 90
markets and, 614–15, 614*t*
ownership of, 167
shadow prices and, 355
sustainable development and, 787, 789, 798
see also agriculture, agricultural sector; property rights
Landberg, Hans, 241*n*
land distribution, equity of, 653
land holdings, farm size and equity of, in selected countries, 654*t*
landlocked nations, 707
trade and, 671–73, 673*f*
land ownership, Gini coefficient for, 653
land reforms, 177, 664
in East Asia, 205–6
forms of, 656–57
redistributive, 657–58
land rights, forms of, 654–55
land-saving technology, 635*b*
land titling, 656
land to the tiller with compensation, 656–57
land to the tiller without compensation, 657
land valuations, taxes and, 415
Lane, Timothy, 491*n*
Laos, 58, 360, 681, 718
Laplante, Benoit, 759*n*
Lardy, Nicholas, 731*n*
Larraín, Felipe P., 495*n*
Larsen, Bjorn, 784*n*
latifundia, 653
Latin America, 137, 144, 157, 441
adjusted net savings in, 766
agriculture in, 588, 619, 621, 628, 629, 631, 663
average TFP growth in, 71, 72*t*
climate change and, 795
customs tariffs in, 414
debt crisis in, 141, 474
defaults in, 464, 465*b*
economic growth and inflation in, 154, 154*t*
economic growth in, 33, 36, 118–19*b*
education in, 261, 262*f,* 263, 284, 285, 289*b*
favelas in, 194*n*
financial crisis in, 497
financial policy in, 440
head-count index in, 189, 192
health in, 310
immiserizing growth in, 200
import substitution in, 724
inequality in, 173, 177, 178
inflation in, 154, 154*t,* 428, 432, 445, 558
land distribution in, 653
land reform in, 657
levels of GDP per capita in, 1820-2008, 35*f*
life expectancy in, 307
"lost decade" in, 36, 37*n,* 148, 152, 160, 210
micro credit in, 449
polio elimination in, 338–40
population of, 227
poverty in, 200, 201
primary export markets in, 711
stamp taxes in, 417
total fertility rate in, 225*b,* 226*b*
trade and, 670, 735
Washington Consensus and, 146, 154–55, 160, 161
Latvia, 360
Laurens, Bernard, 495*n*
law of one price, 29*n*
lean season, 621
least-developed countries, 11
least-squares regression, estimating average growth rates by, 62*b*
Lee, Dwight R., 410*n*
Lee, Jong-Wha, 118*n,* 369*n,* 378
Lehman Brothers, 391
Leith, Clark, 59*n*
Leland, Anne, 516*n*
Lensink, Robert, 522*n*
Lesotho, 58, 401, 742, 748
less-developed countries (LDCs), 11
Leuthold, Steven C., 444n
level of output per unit of land, 663
Levine, Ross, 75*n,* 533*n*
Levine, Ruth, 332*n,* 335, 335*n,* 524*n*
Lewis, Jeffrey D., 321*n*
Lewis, W. Arthur, 53, 53*n,* 174, 174*n,* 175, 592, 605, 618
liabilities, 424
liberalization, 728, 735, 738, 756

Liberia, 112*b,* 402, 504, 658
education in, 277
Liechtenstein, 12
life-cycle model, of household savings, 377–79
life expectancy, 14, 344
in Botswana, 59*b*
computing, 304, 305–6*b*
economic development and, 41–42
economic growth and, 77–78, 78*f,* 319–20
in Equatorial Guinea, 14
in Ethiopia, 6
health-adjusted, 306
health and, 304
HIV-affected countries and, 327*f*
in Hungary, 14
income and, 312, 315, 316*f,* 317
increases in, 307, 307*t,* 310, 318*b*
in Malaysia, 4
poverty and, 314, 314*f*
in Thailand, 56, 101*b,* 102*b*
in Ukraine, 7
in Zambia, 56
life satisfaction
happiness *vs.,* 51
as measure of well-being, 52*f*
Ligon, Ethan, 613
limited markets, 724–25
Limits to Growth, The (Club of Rome), 758, 759*n*
Lindahl-Kiessling, Kerstin, 241*n*
Lindauer, David L., 75*n,* 532*n,* 733*n,* 745*n*
linkages, 132
backward, *see* backward linkages
fiscal, 686, 699
forward, 686
intersectoral, 605, 606, 618
poverty-environment, 785–92
primary products and, 686
Lipsey, Robert, 369*n*
liquid financial assets, 424, 425, 437
liquidity crisis, 486*b*
Lister, Joseph, 312
literacy, 4
in Thailand, 102*b*
in Ukraine, 8
Lithuania, 360, 435
livestock systems, 622
Living Standards Measurement Study (LSMS), 195
Lleras-Muney, Adrianna, 316, 316n
Loayza, Norman, 378, 384
logarithms:
natural, 43*b*
use of, 43–44*b*
log export bans, 784
lognormal distribution, 169
Lomborg, Bjorn, 242*n*
long run, 109
long-term capital flows, 495
Lorenz, Max, 170
Lorenz curve, 170–71, 171*f,* 171*n,* 172, 214
lower-middle-income countries, 11, 12, 13*t*
central government expenditure as share of GDP in, 393*t*
composition of tax systems in, 400*t*
education in, 283*f*
see also middle-income countries
low-income countries, 11, 12, 13*t,* 20
central government expenditure as share of GDP in, 393*t*
composition of tax systems in, 400*t*
debt crisis in, 473–80
debt reduction in, 474–75
economic development achievements in, 15
education in, 260, 264, 282, 283*f,* 298
government saving in, 381
health in, 315
life expectancy in, 307, 307*t,* 312, 344
population in, 218, 224
private capital flows in, 387, 389
savings rate and, 375
tertiary education enrollment rates in, 259
working conditions in, 745
Lsotho, 325
Lucas, Robert, 89, 89*n,* 125*n*
Lum, Thomas, 516*n*
Luxembourg, 89, 505
luxury consumption taxes, 413–14

macroeconomic equilibrium, defined, 553
macroeconomic instability, 545
investments and, 358
macroeconomic policy, objective of, 556
macroeconomics, 18
macroeconomic stability, 149, 203, 204
economic growth and, 75–76
Madagascar, 56, 87, 122*f,* 123, 124*n,* 222
Maddison, Angus, 32–33, 32*n,* 36, 36*n,* 54, 58, 218*n*
Magaziner, Ira, 729*n*
Magellan, Ferdinand, 10
Mahatma Gandhi National Rural Employment Guarantee Act, 208
maize, 633
trends in prices of, 659, 659*f*
yield gap in Africa, 634*f*

Majeed, Hamdan, 5*n*
Maki, Sara, 244*n*
malaria, 9, 78, 78*n,* 85, 180, 205, 306, 317, 321, 323, 324*b,* 329–30, 330*b,* 332–33, 342, 345
  Panama Canal and, 322*b*
Malawi, 511, 613, 650
  fertilizer subsidies in, 645–47*b*
  tax reform in, 408*n*
Malaysia, 8, 9*n,* 137, 176*n,* 189, 397, 402, 686, 687, 700, 724
  agricultural growth multiplier in, 608
  business regulations in, 359
  credit ceilings in, 439, 441
  doing business in, 360
  economic growth in, 15
  electronics industry in, 367
  EPZs in, 371
  financial crisis in, 457, 481, 488, 489
  life expectancy in, 304, 305–6*b*
  log export bans in, 784
  malaria control in, 78
  natural resources in, 370
  pollution taxes in, 777
  reserve requirements in, 439
  semiconductor industry in, 3–4
  trade and, 726, 729, 733, 738
Mäler, Karl-Goran, 786*n*
Mali, 87
malnutrition, 241*n,* 311
Malthus, Thomas R., 232–33, 235, 241, 242, 758, 759
Malthusian term, 768*b*
managed float, 434, 434*f*
Mandela, Nelson, 342
mangrove forests, 787
manufacturers:
  economic growth and, 83
  foreign direct investment and, 365
  trade and, 670, 689
maquiladoras, 681
March of Dimes, 339
Marco Polo, 10
Marcos, Ferdinand, 362, 476*b,* 520, 527
marginal cost of abatement (MAC), 775, 780*f*
marginal external cost (MEC), 774–75, 780*f*
marginal product, of fertilizer, 644
marginal product of labor (MPL), 593, 595
marginal rate of technical substitution, 638
marginal revenue product, 355
marginal social benefits, 133*b*
marginal social costs, 133*b*
marketable permits, sustainable development and, 779–81
market access:
  expanding, 747–50
  for farmers, 651
market-creating institutions, 80*b*
market equilibrium, 133*b*
market exchange rates:
  comparing GDP per capita using PPP, in 2009 (US$), 31*t*
  purchasing power parity methods of converting GDP *vs.,* 29*t*
market failures, 133–34*b,* 143*b*
  defined, 243
  infectious disease and, 318*b*
  population and, 243–44
  prevalence of, 132
  sustainable development and, 769–72
marketing margin, 650
market-legitimizing institutions, 80*b*–81*b*
market-regulating institutions, 80*b*
markets, market economies
  agriculture and, 614–15, 614*t,* 658–63
  competition and, 149–50
  Dutch disease and, 564–65, 567
  export processing zones and, 730, 733–34
  functioning, elements in, 149–50, 152
  globalized trade and, 741–55
  import substitution and, 724–25
  investment and, 355, 356, 359–70, 374
  macroeconomic stability and, 149
  in open economies, 558
  sustainable development and, 769–72
  trade and, 553–56, 554*f,* 555, 666–67, 673, 674–76, 679–81, 726–27, 750–53, 755–56
  transition from Soviet command model to, 148–53
market-stabilizing institutions, 80*b*
Marshall, George, 503*b,* 504*b*
Marshall Plan, 131, 131*n,* 499, 503–4*b,* 518, 520
Martens, Bertin, 536, 536*n,* 537
Martin, Will, 662, 662*n*
Marx, Karl, 135, 159
Mason, Edward S., 137*n*
Massachusetts Institute of Technology, 758
Masson, Paul R., 378, 384
Masters, William A., 620*n,* 621
Mauritius, 15, 58, 360, 368, 371, 670, 671, 687, 726, 733, 748
Mbeki, Thabo, 342
McDade, Thomas, 309*n*
McDermont, C. John, 691*n*
McEwan, Patrick, 277*n*
McGillvary, Mark, 531*n*

McKay, Andrew, 650*n*
McKenzie, Richard B., 410*n*
McLure, Charles E., Jr., 418*n*
McNamara, Robert, 217, 218, 219, 224, 236, 253, 342
measles, 34
medical services, 522
  in Denmark *vs.* in Sierra Leone, 302
Medicare, 399, 400, 404
medium of exchange, 423
Meijer, Siet, 622*n*
merchant capitalism, 14
Merck & Co., 524*b*
Mercosur, 402
methane, 793
Mexico, 180, 181, 426, 473, 564, 589, 693, 711, 765
  agriculture in, 628
  debt crisis in (1982), 469–70, 470*t*
  education in, 284, 285, 288–89*b*
  financial crisis in, 455–57, 467, 480, 481, 491, 493, 567
  financial policy in, 448
  income distribution in, 167, 168*f*
  maquiladoras in, 371
  NAFTA and, 710, 740
  PES program in, 779
  pollution registry in, 781
  poverty in, 166, 201
  poverty lines in, 183–84*b*
  size distributions of consumption or income in, 170–71, 170*t*
  trade and, 674, 675, 676, 677, 681–82, 719–20*b,* 744
  Washington Consensus and, 148
micro credit, 449–53, 454, 478
  poverty and, 451–52*b,* 452, 454
micro-credit institutions, 422, 449
microeconomics, 18
microfinance, 449, 453
  temporary migration *vs.,* 754
Middle East, 174, 516, 543, 651, 671
  average TFP growth in, 71, 73*t*
  fragile environments in, 787
  population of, 227
  rate of GDP per population growth (percent/year), 37, 37*n*
  total fertility rate in, 226*b*
Middle East and North Africa (MENA)
  cost of environmental degradation in, 775, 776*t*
  education in, 262, 262*f*
  poverty in, 189
middle-income countries, 20
  doing business in, 360
  education in, 260, 264, 265, 266, 282, 298
  government saving in, 381
  health in, 309
  life expectancy in, 307, 307*t,* 312, 344
  population in, 218, 224
  private capital flows in, 387, 389
  tertiary education enrollment rates in, 259
migrants, migration
  agricultural growth and, 613, 617, 618
  temporary, international trade and, 753–55
Miguel, Edward, 291*n*
Mikesell, Raymond, 403*n*
Milanovic, Branko, 210*n*
military expenditures, 399
Mill, John Stuart, 411
Millennium Challenge Corporation (MCC), 503–4, 540, 541
Millennium Development Goals (MDGs), 46–50, 54, 181, 201, 218
  economic growth and, 50
  for education, 260, 261, 263, 276, 287
  environment and, 786
  improving maternal and child health, 249, 343
  targets of, 47–48*b*
Millennium Ecosystem Assessment Board, 786
Millions Saved: Proven Successes in Global Health (Levine), 335, 342
Minah Karan: The Truth about Malaysian Factory Girls (Daud), 3*n*
mineral rights, verifiable international auctions and, 706
mineral wealth, 178
  civil war and, 705
minifundia, 653
minimum wages, minimum wage laws, 593
Mishel, Lawrence, 747*n*
mitigation, global climate change and, 796, 797–98, 799
modern crop varieties (MVs), 628, 629, 629*f,* 631, 631*t,* 632
modern economic growth, 14
modern growth theory, 90
Modigliani, Franco, 377, 377*n*
Moldova, 26, 27
Molyneaux, John, 248, 248*n*
monetary policy, 18, 392, 422, 423, 437
  credit ceilings, 439–40
  exchange-rate regimes and, 433–35
  inflation controlled through, 438–39
  interest-rate regulation and, 440–41
  macroeconomic equilibrium and, 576

monetary policy *(continued)*
moral suasion and, 440–41
open economies and, 561, 563, 567
price stability and, 432–42
recession of 2008 and, 391
reserve requirements and, 439
stabilization and, 560, 561
see also financial policy
money, money supply and, 423–26
money multiplier, 439*n*
money supply, 423–26, 427, 453
defined, 424
domestic and international component in, 435–36
markets and, 141
in open economies, 558, 564
recession of 2008 and, 391
Mongolia, 683
monopolies, 133*b*
price distortions and, 149
monopoly power, 445
Montezuma, 34
Montiel, Peter, 552, 552*n*
moral hazard, 493
moral suasion, 439
Moran, Dominic, 788*n*
Moran, Theodore H., 368*n,* 369*n,* 370, 370*n,* 744*n*
morbidity, 304
see also disease (morbidity)
Morduch, Jonathan, 453
Morocco, 187
Morris, Cynthia Taft, 74, 74*n*
Morrisey, Oliver, 531*n,* 650*n*
mortality, 309, 310, 310*n*
defined, 304
leading causes of, 311*t*
modern history of, 318*b*
see also deaths, death rates
Moses, Lawrence O., 650*n*
Mosher, Arthur, 623, 623*n*
Mosher framework, 623–25, 624*f,* 626*f,* 627, 630, 637
Mosley, Paul, 533*n*
Moss, Todd, 50*n*
Moyo, Dambisa, 500*n*
Mozambique, 45, 92, 500, 520, 532, 583, 658, 681
Mubarak, Hosni, 532
Mugabe, Robert, 206, 658
multidimensional poverty index (MPI), 182*n*
multidrug-resistant TB (MDR-TB), 334
Multi-Fiber Agreement (MFA), 748, 751
multilateral aid, 506–7, 517, 535
Multilateral Investment Guarantee Agency (MIGA), 508
multilateral trade negotiations, 750–53, 756
multinational corporations (MNCs), 364, 367, 368, 369, 370, 371, 372, 373, 380
Mumford, Kevin J., 762*n*
Museveni, Yoweri, 479*b*
Mutharika, Bingu wa, 645*b*
Muto, Megumi, 652*b*
Myanmar (Burma), 15, 58, 546, 687, 710
Mycobacterium tuberculosis, 312, 324*b*
Myint, Hla, 685, 685*n*
Myrdal, Gunnar, 140, 140*n*
Mystery of Capital, The (de Soto), 83, 83*n*

Namibia, 326, 328
narrow money, 424
National Autonomous University of Mexico (UNAM), 285
National Foundation for Infantile Paralysis, 339
national income, 166
gross national product and, 24
nontraded goods and services and, 27–28
reporting on, 30
national income accountiing identity, 577
national income accounting, balance of payments in context of, 577–79
National Tuberculosis Project Office, 337
national vaccine days (NVDs), 339
natural capital, 762, 763, 764*t,* 765, 766
natural disasters, 574–75
natural experiment, education and, 277–78
natural experiments, 286
natural increase, in population, 220
natural logarithms, 43*b*
natural resources, 683–84, 685
economic growth and, 59–60*b,* 63, 687, 688*f*
foreign direct investment and, 365, 370
sustainable development and, 763, 765, 766, 775, 782, 786, 790
taxes on, 402
natural resource trap, 703–5
Neary, S. Peter, 564, 564*n,* 694*n*
negative real interest rates, 445, 446, 447, 454
negative spillovers, 371
Nehru, Jawaharlal, 132
neoclassical growth model. *see* Solow growth model
neoclassical production function, 103, 117
neoclassical production technology, isoquants for, 104*f*
neoclassical two-sector model, 599–602, 600*f*
neo-Malthusians, 242, 242*n,* 243

Nepal, 187, 326, 375, 613, 673, 772
net barter terms of trade, 692, 693*f*
net cash flow, 351, *353*
net enrollment rates, 261
Netherlands, 505
net present value (NPV), 352, 353–54, 354*t,* 356, 460
net saving (NS), 766
net unilateral transfers, 578
neutral tax, 418, 419–20
nevirapine, 326
new household economics, 235, 236
Ne Win, 710
New Zealand, 226
Nicaragua, 77*f,* 432, 683
  education in, 267, 268*f*
Niger, 45, 87, 112*b,* 608, 656
  cell phones and grain traders in, 652*b*
  polio in, 338
Nigeria, 71, 326, 509, 687, 693, 702, 703, 705, 709
  Dutch disease in, 700–701*b*
  natural resources in, 370
  oil boom in, 696
  oil revenues and, 763, 766
  polio in, 338
  population growth, age structure, and dependency ratios in, 239–40*b*
  population of, 231
Nike, 746–47*b*
nitrous oxide, 793
Nixon, Richard M., 468
Nkrumah, Kwame, 135, 138, 139*b*
Noland, Marcus, 729*n*
nominal exchange rates, 547, 549, 694, 695
  fixed, 559
  in open economies, 550–52*b*
  real exchange rates *vs.,* 550–52*b*
nominal GDP, 25*n*
nominal rate of interest, 447, 454
nominal rates of protection, 713, 713*f*
nonagricultural market access (NAMA), 752
nonconcessional assistance, 502
nonexcludability, 393, 394
nonfinancial assets, 424
nongovernmental organizations (NGOs), 138, 502, 646*b*
nonprice constraints, 649
nonrenewable resources, 758, 765, 775
nonrival consumption, 393, 394
nontradable goods and services, 27–28, 547, 548
  disequilibrium in, 576
  Dutch disease and, 697–99*b*
  exchange-rate management and, 722
  markets and, 554*f,* 558
  natural disasters and, 574
  open economies and, 553, 561
  prices for, 694, 695, 696
Nordhaus, William, 797, 797*n*
Norgaard, Richard B., 772*n*
North, Douglass, 79, 79*n,* 80*b,* 80*n,* 653, 653*n*
North Africa, rate of GDP per population growth (percent/year), 37, 37*n*
North American Free Trade Agreement (NAFTA), 204, 710, 725*n,* 740
North Korea, 15, 514, 546, 710
North *vs.* South, 11
Norton, George W., 620*n,* 621
Norway, 505
Notzon, Francis, 308*n*
NPV debt/export ratios, 477, 478
NPV debt/exports, 461
NPV debt/GDP, 460–61
NPV debt/revenue, 461
Nunberg, B., 532*n*
Nunns, Jim, 404*n*
nutrition:
  caloric intake, population and, 241, 241*n*
  health and, 323
  linkage to economic growth, 608–9*b*
  see also food
NUTs. *See* net unilateral transfers

Odell, Kathleen, 452*n*
odious debt, 466, 475, 476–77*b*
official development assistance (ODA), 502–3, 506, 511
official foreign saving, 385–86
Ohlin, Bertil, 677, 682
oil, as resource trap, 703
oil prices, 36, 568, 692
  food crisis of 2005-08 and, 661
  inflation and, 432, 468
  nominal, trends in, 659*f*
  role of government in development and, 140–41
  in Venezuela, 155
old age support, 254, 255
Oleson, Kirsten, 762*n*
oligopoly power, 445
Omran, A. R., 309*n*
one child campaign, in China, 245, 250–51, 250*n,* 253
one child policy, in China, 379
One Hundred Countries, Two Billion People (McNamara), 217

O'Neill, Brian, 227*n*
on-the-job training, 257–58
open access, 770
open economies, 435, 564, 575
  disequilibrium in, 556–60
  equilibrium in, 546–49, 548*f,* 552–63
  poverty and, 203, 204
  stabilization in, 559–60*t,* 560–63
open-market operations, 422, 438, 438*n,* 439, 454
open-mouth operations, 441
openness, 725, 738
Operation Mop-Up, 339
Oportunidades (Mexico), 207, 215, 288–89*b,* 290, 395
opportunity costs, 355
optimal target, for CO_2 concentrations, 797
oral polio vaccine (OPV), 339
oral rehydration therapy (ORT), 317, 340, 341, 342, 345, 500
Orazem, Peter, 276*n*
Organization for Economic Cooperation and Development (OECD), 173–74, 264, 266, 417, 501, 750
Organization of Petroleum Exporting Countries (OPEC), 140, 141, 407*b,* 468, 511, 700*b*
Orlando, Maria Beatriz, 525*n*
Ostrom, Elinor, 772, 772*n*
Ostry, Sylvia, 752*n*
output:
  growth rate of, 97–98
  in isoquants, 94
  price and, 648*f,* 649
output per effective worker, 115
output per worker, 68, 104, 105
  sources of growth in East Asia and other regions, 1960–2000, 72*t*
  technological progress and, 113
output prices, tariffs and, 715*b*
outsourcing, 670
outward orientation, 674, 725, 728, 734, 741, 755
  core standards of, 745
  economic growth and, 83, 84
  sweatshops and, 743–47
  wages and, 744
Oxfam International, 502
Oxford Poverty & Human Development Initiative, 182*n*

Pack, Howard, 368*n,* 531*n,* 729*n*
Pack, Janet Rothenberg, 531*n*
Pagett, Richard, 408*n*
Pagiola, Stafano, 779*n*
Paisner, Michael, 622*n*
Pakistan, 122*f,* 123, 137, 138, 402, 516, 525, 589, 660
  financial policy in, 448
  population of, 222, 231
Palestinian Territories, polio in, 338
Panama Canal, 321
  malaria, yellow fever and, 322*b*
PanAmerican Health Organization, 339
pandemics, epidemics *vs.,* 323*n*
panics:
  creditor, 486–89
  rational, 488
  stopping, 489–91, 493–94
Papageorgiou, Stavros, 788*n*
Papua New Guinea, 15, 500
  education in, 294
Paraguay, 27, 653, 656
parasite presence, in poor children, 329*f*
parent teacher associations (PTAs), 296
Pareto, Vilfredo, 133*b*
Pareto efficiency, 133*b*
Pareto optimality, 138
Paris Club, 474, 475, 478
Paris Declaration on Aid Effectiveness and the Accra Agenda for Action, 540
Pasteur, Louis, 312
pastoralists, 194
pastoral nomadism, 621–22, 663
Patinkin, Mark, 729*n*
Patrinos, Harry, 273, 273*n,* 275, 276*n,* 277, 280
Paxson, Christina, 241*n*
payments for environmental services (PES), 779
Peace Corps, 504
peanuts, 402
Pearce, David W., 780*n*
Pearson, Scott, 584, 584*n,* 623*n*
Pechman, Joseph A., 408*n*
pegged exchange rates, 433, 434*f*
PEMEX (Mexico), 364, 456
Penang, Malaysia, semiconductor industry in, 3–5
pension systems, population and, 254
People's Bank of China, 387–88
per capita income. *see* incomes, per capita
per capita wealth, 764*t*
Pereira da Silva, Luiz, 532*n*
perennial crops, 622
perfectly elastic, 594
Perkins, Dwight H., 408*n*
Perry, Matthew, 710

Persian Gulf, 687
"persistence of the past," inequality and, 178
personal income taxes, 404, 412
  see also income taxes
Peru, 77*f,* 195, 339, 432, 473, 765
  education in, 259*f,* 266, 282, 294
  fishmeal industry in, 686
  hyperinflation in, 430, 431*b*
Pew Research Center, 709
phase diagram, 553–56, 555*f,* 561
Philippines, 58, 247, 362, 396, 401, 473, 511, 527, 782
  agriculture in, 628
  education in, 292
  financial crisis in, 480–81, 488, 489
  impact of schooling on farm profitability in, 279*b*
Philipson, Tomas J., 308*n*
physical capital per worker, sources of growth in East Asia and other regions, 1960-2000, 72*t*
Pinochet, Augusto, 559*b*
Plasmodium falciparum, 324*b*
Plato, 212
Pleskovic, Boris, 85*n*
Poland, 473, 504, 726
  education in, 264
Policansky, David, 772*n*
polio, 345
  eliminating in Latin America, 338–40
political economy, inequality and, 179
political institutions, 80*b,* 81*b*
political stability, economic growth and, 75–76, 88, 90, 117
political systems:
  global information flow and, 16
  globalization and, 9
politics. *see* government
pollution, 762–63
  health and, 323
  income levels and, 759–60, 800
  informal regulation of, 781–82
  marketable permits for, 779–81, 780*f*
  multinational corporations and, 368
  optimal level of, 774*f*
  sustainable development and, 771, 772*f,* 774–76
  taxes and, 776–79
Poor Economics (Banerjee and Duflo), 317
population, population growth, 9, 308
  accumulation and, 237–38
  agriculture and, 241–42
  authoritarian approaches and, 249–51, 253
  brief history of, 218–20
  carbon dioxide emissions and, 794–95, 794*f*
  causes of, 231–36
  demographic dividend (or bonus) and, 238
  demographic future and, 227–28, 231
  demographic transition and, 220–24, 238
  economic development and, 236–38, 240–41
  economic growth and, 112*b*
  explosion in, 220
  family planning and, 246–49
  fertility and, 240–41
  GDP and, 124*f,* 237
  immigration and, 224, 226, 231
  implosion in, 226
  levels and trends in world population, 2009, 226*t*
  market failures and, 243–44
  present demographic situation and, 224, 226–27
  productivity and, 241–43
  rate of gross domestic product per, 37*t*
  reducing unwanted births, 247
  savings and, 238
  in Solow growth model, 106, 237, 237*n*
  sustainable development and, 758, 763, 787
  in twenty-first century, 253–54, 255
  world population growth through history, 219*f*
  world population historical trends and projections, 228*f*
  world's 10 most populous nations in 2050, 230*t*
  see also Malthus, Thomas R.
population-carrying capacity, 220
population growth rate, changes in Solow model, 109–13, 111*f*
population momentum, 228, 229–30*b*
population neutralists, 243
population optimists, 242, 243
population pessimists, 237, 242, 243
population policy, 245–46
population revisionists, 243, 244
population-weighted international inequality, 211
portfolio equity, 363, 386, 389
Portugal, 514
  financial crisis in, 455
POSCO, 734
positive checks, 232
positive externality, 126
positive real interest rates, 447, 454
positive spillovers, 371, 372
postindustrial countries, 11

potential level of output per worker, 109
poverty, 17, 20, 55, 56
  absolute, standard international definition for, 165
  agriculture and, 205, 588, 589, 610–17, 618
  in Bangladesh, 165–66
  climate change and, 796
  descriptions of, 197
  in developing world, 15
  dissenting opinions on, 191
  economic development and, 202*b*
  economic growth and, 58, 197–98, 203
  economic growth as good for the poor, 198–200, 199*f*
  economic growth as not being enough for poor, 200–201
  education and, 205, 206
  environment and, 785–92, 790*f,* 801
  in Ethiopia, 5–6
  extreme, 186
  financial crisis of 2008 and, 193
  food prices and, 662
  foreign aid and, 499, 500, 501–2, 516–17, 543
  gender and, 194–95
  geography of, 84
  global estimates of, 165
  global inequality and end of, 208–14
  health and, 205, 206, 344
  HIV/AIDS and, 327
  improved opportunities and, 205–6
  income transfers and, 206
  inequality and, 166, 178–79, 180
  intergenerational cycles of, 205, 206
  life expectancy and, 314, 314*f*
  living in, 195–97
  in Malaysia, 3
  MDGs and, 48, 49*f*
  measuring, 180–82, 184–90
  in Mexico, 166
  micro credit and, 451–52*b,* 452, 454
  multidimensional quality of, 181, 182, 182*n*
  political unrest and, 76
  population and, 193–95, 244
  reducing, strategies for, 197–208, 215
  rural, 611, 611*f*
  rural-urban divide and, 194, 214
  savings and, 374
  severity of, 185–86, 189–90
  since 2005, 192–93
  social safety nets and, 207
  subsidies and, 396
  temporary migration and, 753–55, 756
  trade and, 681, 736, 739–41, 743, 744, 756
  ulta-poor, 196–97, 197*n*
  urban, 611, 611*f*
  women and, 214
poverty gap (PG), 185, 190
  calculating, 186
poverty gap squared, 186
poverty lines, 181–91, 182*n,* 188*f,* 192*b*
  in Bangladesh, Mexico, and United States, 183–84*b*
  defined, 181
  reason for $1.25 a day, 186–90
  regional distribution of people living below $2-a-day poverty line, 2005, 193*f*
Poverty Reduction Strategy Papers, 478
praziquantel, 290
Prebisch, Raul, 136, 689, 690, 705, 708, 711*n*
Prebisch-Singer hypothesis, 605–6, 689, 692
present value (PV), 270–71, 351–54
Preston, Samuel, 316, 316*n*
Preston curve, 316, 316*f*
preventive checks, 232
price elasticity of demand, 136, 403
prices, pricing, 17
  inflation and, 203–4, 215, 427
  international, deriving, 30
  market functionality and, 150
  monetary policy and stability in, 432–42
  in open economies, 546–49, 552–56, 558–59
  output and, 648*f,* 649
  price ceilings, 18
  state-set pricing, 145–46
  see also oil prices
price takers, 546
primary export-led growth, empirical evidence on, 687–707
primary exports:
  economic growth and, 687–88
  governance and, 704–5, 708
primary products, trade in, 683–86, 684*f,* 686, 687, 707
principal-agent problem, 536–37, 539, 543
Principles of Political Economy and Taxation, The (Ricardo), 591, 674
Pritchett, Lant, 36*n,* 75*n,* 192*b,* 247*n,* 248*n,* 266*n,* 280*n,* 281*n,* 315*n,* 534, 534*n,* 738*n,* 754, 754*n*
private enterprise, economic growth and, 82–83, 88
private foreign saving, elements of, 386
private goods, public goods *vs.,* 134*b*
private marginal cost curve, 772*f*
private return, 269, 269*f*
private savings, 578

private voluntary assistance, 502
privatization:
of state-owned enterprises, 151, 151*n*
Washington Consensus and, 147
produced capital, 762, 764*t*
producers' surplus, 714
production, productivity, 87, 127
economic growth and, 60, 64, 66, 67*f,* 87–88, 90
education and, 281, 298
health and, 319–20
of investment, 351–57
life expectancy and, 77
population growth and, 241–43
technological progress and growth in, 113, 114
see also gross domestic product; gross national product
production function, 66
derivation of, 91
diminishing returns and, 117–21
in Harrod-Domar model, 94–96
in Solow diagram, 108
in Solow growth model, 105*f*
production isoquant, 636*b*
production possibility frontier, 638, 649
production subsidies, 720722
productivity growth:
definition of, 64
economic growth and, 74
as "manna from heaven" in Solow model, 117
measuring, 88
profits, maximization of, 150, 639, 640, 644
progessive taxes, 411
Programme for International Student Assessment (PISA), 264, 266
PROGRESA (Programa de Educación Salud y Alimentacíon), 207, 288*b*
progressive income taxes, 412
project appraisal, 351
property rights, 82, 83, 801
marketable permits and, 779–81
population and, 243
secure, costs/benefits of, 655
sustainable development and, 773, 774
Washington Consensus and, 147
women and, 194
see also land
property taxes, incidence problem and, 414–16
pro-poor growth, 201–4, 205, 215
protective tariffs, 713–18
direct effect of, 713
import quotas, 714, 717–18
markets and, 713–14
nominal, 713, 713*f*
world price and, 713, 717
Psacharopoulos, George, 273, 273*n,* 275, 276*n,* 277, 280
public goods, 133*b,* 393
market failure associated with, 133*b*
private goods *vs.,* 134*b*
public investment expenditures, 394
public sector investments, opportunity costs and, 355
Puerto Rico, 172
Punic Wars, 465*b*
purchasing power parity (PPP), 12, 28, 32, 53–54, 477
hypothetical conversion for two countries using two goods, 29*t,* 30
limitations with, 31–32
poverty line and, 187, 188

Qatar, 12
quantitative restrictions, on imports, 714
quantity regulations, 774
quota rents, 717, 717*n*
quotas, 712, 736, 750
market access and, 748–49

race, poverty and, 194
Radano, Alberto H. J., 406*b*
Radelet, Steven, 37*n,* 73, 73*n,* 118*n,* 155, 198*n,* 378, 384, 487, 491*n,* 517, 521*n,* 540*n,* 733*n*
Ramanathan, V., 795*n*
Ramsey (inverse elasticity) rule, 419
randomized controlled trials (RCTs), 286, 287, 293, 293*n,* 295*b,* 297, 298, 451*b*
randomized evaluations, 286
Rangarajan, C., 607*n,* 608*n*
Ranis, Gustav, 592, 592*n,* 603*b*
rapid inflation (hyperinflation), 573
rapidly growing countries, characteristics of, 74–79, 82–87
rates of return
in education, 269–71, 272, 273–79, 273*t,* 298
estimating from wage equations, 275–76*b*
rational panics, 488
Rausser, G., 608*n*
Ravallion, Martin, 45*n,* 175*n,* 176*n,* 187*n,* 188, 189, 201*n,* 203*n,* 205*n*
Rawlings, Jerry, 139*b,* 566*b*
Rawls, John, 411
real appreciation, 695
real depreciation, 695

real effective exchange rate (REER), 552*n*
real exchange rates (RER), 549*n,* 703
  definition for, 695
  Dutch disease paradox and, 696
  nominal exchange rates *vs.,* 550–52*b*
  in open economies, 549, 553
real GDP, 25
  computing, 25*n*
real GDP per capita, 25
real interest rates, 352, 445, 446, 446*t,* 447, 454
Reardon, Thomas, 786, 789, 790*n,* 791, 792
recessions, combating, international debt and, 441–42
recurrent costs, 420
recurrent expenditures, 394–95, 397
rediscount rates, 438, 440
redistributive land reforms, 657–58
reduction in debt, 471
Reed, Walter, 322*b*
refinancing, of debt, 471
regional development banks, 511
regulation:
  barriers to doing business and, 361, 362
  financial policy and, 423
  role of, 421–22
regulatory mechanisms, economic growth and, 83, 88
Reinikka, Ritva, 296*n,* 706*n*
Renelt, David, 75*n*
renewable resources, 758
  sustainable development and, 799
renminbi (RMB), 731*b*
rental rate of capital, 642
rent ceilings, 657
rent contracts, reform of, 656
rent control, 18
rent reduction, 656
rent seeking, 723, 724
rent-seeking behavior, 358, 361, 388, 725
Repetto, Robert, 402*n,* 784*n*
RER, 704
rescheduling of debt, 471
research and development (R&D), investments in, 126
reserve requirements, 439, 454
residual claimants, 655
resource balance, 470
resource cost, 714
resource curse, 683, 688, 707
  breaking, 705–7, 708
resource endowments, inequality and, 178
resource movement effect, 697*b*
resource rents, 686, 704, 705, 708
resources, access to, measuring, 42
resource trap, 703–5
retail price, 650
retirement savings, 254
return on investment, 117
returns to scale, 119
revaluation, 547
Ricardian equivalence, 383
Ricardo, David, 130, 159, 383, 384, 591, 592, 597, 674, 676, 677, 681, 682, 707, 709
rice market, events in, 661*f*
rice production, 178, 610, 628, 633, 643–44, 647–49, 702*b*
  biofuels and, 660
  price trends and, 659, 659*f,* 660
Riley, Nancy E., 250*n,* 252*n*
risk
  agricultural production and, 586
  transformation and distribution of, 426–27
river blindness (onchocerciasis), 321, 342, 500, 524–25*b*
Robinson, James, 59*n,* 82*n*
Robinson, Joan, 98*n,* 113*n*
Rodríguez, Francisco, 84, 84*n,* 738*n*
Rodríguez-Clare, Andrés, 736, 736*n,* 737
Rodrik, Dani, 59*n,* 80*b,* 80*n,* 84, 84*n,* 89, 90*n,* 180*n,* 202*b,* 546*n,* 729–30, 735, 735*n,* 736, 738*n*
Roemer, Michael, 198*n,* 406*b,* 408*n,* 428*n,* 694*n,* 733*n*
Roesad, Kurnya, 783*n*
Rogoff, Kenneth, 738*n*
Rohaly, Jeff, 404*n*
Romania, 328
Romer, David, 83, 83*n,* 736, 736*n,* 737
Romer, Paul, 125*n*
Roodman, David, 508*n,* 533*n*
Roosevelt, Franklin D., 339
Roosevelt, Theodore, 322*b*
Rosegrant, Mark, 622*n*
Rosenstein-Rodan, Paul, 132, 132*n,* 134
Rosensweig, M. R., 320*n*
Rosenzweig, Mark, 279*b,* 603*b*
Ross, Ronald, 322*b*
Rostow, W. E., 520*n*
Rostow, Walter, 131*n,* 520, 520*n*
Rotary International, 524
Roubini, Nouriel, 491*n,* 493, 493*n,* 496*n*
royalties, 399
rule of 70, 63*b*
  doubling times and, 62*b*
rule of law, economic growth and, 79, 82

runaway inflation (hyperinflation), 428
among developing countries, 1948-2008, 429–30*t*
rural areas
environmental degradation and, 786–87
market access and, 651
see also migrants, migration
rural population, agroecological potential and, 627
Russell, Steven, 327*n*
Russia, 7, 146, 360, 432, 661, 709
financial crisis in, 386, 457, 480, 488, 493
foreign direct investment in, 365
foreign savings in, 384
health in, 314
life expectancy in, 308
markets in, 150
MDR-TB in, 334
population growth, age structure, and dependency ratios in, 239–40*b*
population in, 231
state price controls in, 145
Washington Consensus and, 148
see also Soviet Union
Ruttan, Vernon, 635*b,* 637*n*
Rwanda, 328, 757, 791

Sabin, Albert, 339
Sachs, Jeffrey D., 84*n,* 85*n,* 118*n,* 124, 124*n,* 213, 332, 378, 464*n,* 467, 487, 491*n,* 500, 521*n,* 687, 687*n,* 688, 738, 738*n*
Sadoulet, Elisabeth, 611, 611*n,* 613, 613*n,* 615
Saggi, Kamal, 368*n*
Sala-i-Martin, Xavier, 75*n*
salaries. *see* wages
sales taxes, 402, 403, 413, 414, 417, 420
Salih, Kamal, 3*n*
Salk, Jonas, 339
Salter, W. E. G., 546*n*
Sam, Kim Young, 143*b*
Samoa, 673
Samuelson, Paul, 682*n*
Sandler, Todd, 542*n*
Sangraula, P., 187
Sanyal, Amal, 406*b*
Sarkozy, Nicolas, 26*n*
Sarraf, Maria, 775*n*
SARS, 309, 337
satellite communications, 87
Saudi Arabia, 546, 564, 693
foreign aid through, 505*n*
Saudi Fund for Development, 503
Save the Children, 502
saving function, in Solow diagram, 108
savings, saving rates, 65*t,* 92, 350
balance of payments and, 578–79
changes in Solow model, 109–11, 110*f,* 113
corporate saving, 379–80
domestic capital formation and, 375*t*
domestic savings, 374–75, 376, 389
economic growth and, 65–66, 90, 126
financial intermediation and, 426
food prices and, 610
foreign aid and, 521, 529–31, 543
foreign savings, 374, 384–88, 389
government savings, 380–84, 389
gross domestic savings, 65
in Harrod-Domar model, 98–99, 109
household saving and consumption, 376–79
investments and, 350, 374–88, 389
life-cycle hypothesis of, 241
life-cycle model of, 377–79
life expectancy and, 77
population growth and, 238
price inflation and, 204
retirement accounts, 254
in Solow model, 106–7, 116
sustainable development and, 765–67
in Thailand, 101*b*
savings gap, 131
Schady, Norbert, 207*n*
schistosomiasis, 319, 321
Schmidt, Robert, 236*n*
Schmidt-Hebbel, Klaus, 378
schooling. *see* education
Schultz, Paul, 276*n*
Schultz, Theodore W., 257, 257*n,* 258, 621, 621*n*
Scorse, Jason, 746*b,* 747*n*
sea levels, rise in, 758–59, 798
seasonality, agricultural production and, 585–86, 617, 621
sectoral labor productivity, agriculture and nonagriculture, 90*t*
self-discovery, 735–36
self-fulfilling creditor panics, 486–87*b,* 488
semiconductor industry, in Malaysia, 3–4
semidwarf disease resistant wheat, 628
Sen, Amartya, 39, 40, 252*b,* 252*n,* 253*n,* 602*b,* 603*b*
Senegal, 61*b,* 62*b,* 63*b,* 325, 402, 583, 686
education in, 284
September 11, 2001 attacks, 36, 505*n*
Serven, Luis, 378
service-based economies, 11
Sese Seko, Mobutu, 475, 476*b*
Setser, Brad, 491*n,* 493, 493*n,* 496*n*

settled agriculture, 621, 622, 663
Seva Mandir (India), 294–96*b*
Sevilla, Jaypee, 241*n*
sex-selective abortions, 251*b*
sexually transmitted diseases, 324*b,* 336
Seychelles, 92
shadow prices (economic opportunity costs), 355–56
  effects of, on cost-benefit analysis, 357*t*
  estimating, 356
Shah, Anwar, 784*n*
shallow finance, 443
shallow financial strategy, 444–47
Shambaugh, Jay C., 495*n*
Shanghai Automotive Industry Corporation, 380
share cropping, 655
Sheahan, John, 140*n*
shifting cultivation, 621, 663
shock therapy approach to reform, 151, 158
shorting the peso, 483*n*
short-term capital flows, 495, 497
short-term foreign debt and reserves, 485*t*
short-term foreign debt/foreign exchange reserves, 461, 462
short-term loans, 484
Shoup, Carl, 403*n*
Sicat, Gerry, 403*n*
Sierra Leone, 89, 304, 432, 519, 703, 705
  health in, 299–302, 300*f,* 301*t,* 306
  natural resources in, 370
Silent Spring (Carson), 333
Silvani, Carlos A., 406*b*
Simon, Julian, 242, 242*n*
Sinding, Steven, 236*n*
Singapore, 82, 84, 85, 122*f,* 124*n,* 141, 176*n,* 401, 440, 687, 711
  doing business in, 360
  economic growth in, 58, 59*b*
  financial policy in, 448
  fiscal policy in, 441
  government's role in development of, 142
  health in, 321
  liberalization and, 728, 728*n*
  savings rate in, 107
  semiconductor industry in, 4
  trade and, 668, 671, 726, 741
  see also East Asian tigers
Singapore Airlines, 728
Singer, Hans, 136, 689, 690, 705, 708, 711*n*
Sinopec Corporation, 380
size distributions, 214
  calculating, 169
  of consumption or income in Bangladesh, Mexico, and United States, 170*t*
Sjöholm, Fredrick, 369*n*
Slade, Roger, 607*n,* 608*n*
slash and burn agriculture, 621
sleeping sickness, 321
small business, 83
small economies, 546, 547
smallpox, 34, 322*b,* 524
  eradicating, 337–38, 345
Smith, Adam, 60, 89, 129, 130, 132, 159, 672, 672*n,* 709
Smith, Lisa C., 197*n*
Smoot-Hawley tariffs, 136, 710–11
smuggling, 401
Snodgrass, Donald, 408*n*
Soares, Rodrigo R., 308*n*
social cost-benefit analysis, 417
social discount rate, 356
social evaluation, 388
socialist economies, socialism, 738–39
social marginal cost curve, 772*f*
social project appraisal, 357
social returns, of education, 271, 272*f,* 273*t*
social safety nets, 207–8, 215
Social Security, 254, 399, 400, 404
Society, Politics, and Economic Development—A Quantitative Approach (Morris), 74, 74*n*
soft governments, 158
soft loans, 386
Solimano, Andrés, 560*n*
Solow, Robert, 64, 64*n,* 66, 66*n,* 70, 90, 91*n,* 103, 103*n,* 113*n,* 126
Solow (neoclassical) growth model, 64–66, 68, 73, 103–17, 118*b,* 121, 123, 127, 128, 457, 520, 520*n,* 521
  basic equations of, 104–7
  change in saving rate and population growth rate in, 109–11, 113
  changes in population growth rate in, 111*f*
  diagram of, 108–9, 108*f*
  increase in saving rate in, 110*f*
  moving beyond, 125–27, 128
  population growth, economic growth and, 112*b*
  powerful implications tied to, 128
  production function in, 105*f*
  strengths and weaknesses of, 116–17
  technology and, 113–16, 115*f*
Solow residual, 69, 70
Somalia, 241*n,* 337, 757
Somoza, Anastasio, 476*b*
son preference, 251–52*b*
source of capital, 365
sources of growth analysis, 68, 71, 88, 126

South Africa, 84, 325, 326, 546, 781
education in, 266
inequality in, 177, 209
land reform in, 658
PES program in, 779
Truth and Reconciliation Commission in, 476*b*
South America, 144, 159
import substituting industrialization in, 135, 136
land distribution in, 653
South Asia, 144
average TFP growth in, 71, 73*t*
cereal yields and poverty in, 612*f*
foreign aid in, 500
poverty in, 189
rate of GDP per population growth (percent/year), 37*n*
Southeast Asia, 402
under-five mortality rate in, 302
Southern Africa Customs Union, 60*b*
South Korea, 141, 142, 158, 160, 189, 468, 507*b,* 709
debt restructuring in, 493–94
declining effectiveness of government intervention in, 143–44*b*
economic growth in, 58
financial crisis in, 481, 491, 493
financial policy in, 440
land reform in, 177, 206
personal income taxes in, 412
savings rate in, 378
semiconductor industry in, 4
see also East Asian tigers
sovereign debt, 469
sovereign lending default, 464–66*b*
sovereign wealth funds, 705–6
Soviet Bloc, trade and, 151
Soviet command model, transition to market economies, 148–53
Soviet economic model, 153
Soviet growth model, 159
Soviet Union, 152, 430, 432, 514, 518, 546, 726
collapse of, 7, 149
collectivization in, 39
growth model pioneered by, 135, 137
structural change in, 151
trade and, 671
see also Russia
Spain:
doing business in, 360
financial crisis in, 455
Spandana, 450–51*b*
Spanish-American War, 465*b,* 476*b*
special economic zones (SEZs), 731*b*
specialization, foreign direct investment and, 366
special safeguard mechanism, 752–53
specific-factor models, 741*n*
Spence, M., 754*n*
spillovers, 367–68, 369*n,* 371
Squire, Lyn, 175*n,* 176*n*
Sri Lanka, 172, 396, 402, 409, 426, 439, 748
Srinivasan, T. N., 378
stabilization:
in Chile, 559–60*t*
economic growth and, 75–76
in financial system, 427
macroeconomic, 149, 203, 204
in open economies, 560–63
programs, 545
stabilization and structural adjustment programs, 472
stabilization package, from inflation and a deficit, 569, 570–71, 570*f,* 573–74
Stages of Economic Growth, The (Rostow), 131*n,* 520
stamp taxes in, 417
Standard Oil, 685
stand-by credits, 511
standstill, 493
Stark, O., 320*n*
state-owned enterprises (SOEs), subsidies to, 396, 399
static conditions, 674, 682
steady state, of Solow model, 109, 110, 115, 117
sterilization, money supply and, 558
sterilization campaigns, 249
Stern, Joseph, 408*n*
Stern, Nicholas, 133*n,* 750*n,* 796, 796*n*
Stern, Robert, 744*n*
Sterner, Thomas, 781*n*
Stern Review, 797, 798
Stevenson, Betsey, 51*n*
Stiglitz, Joseph, 26*n,* 85*n,* 404*n,* 491*n,* 500
stimulus packages, 442
stock markets, 150, 448, 454
stocks, in education, 259
Stolper, Wolfgang, 682*n*
Strauss, John, 276*n*
Strengthening of Rural Health Delivery Project, 341
structural adjustment reforms, 145–46
structural rigidities, 558
structural transformation:
in agriculture, 587–90, 588*f,* 617
phases in, 588–89

stunted growth, 609*b*
stunting, 301
  poverty and, 192*b*
Subramanian, Arvind, 80*b,* 80*n,* 89, 90*n*
Sub-Saharan Africa:, rate of GDP per population growth (percent/year), 37, 37*n*
subsidies, 155, 158, 395–96, 722, 736, 750
  environment and, 776–79
  food or fuel, 145
  impact of, on firms competing with imports, 721*f*
  production, 720–22
  to state-owned enterprises, 399
  Washington Consensus and, 147
subsistence level, 593
substitutes, substitution, between labor and capital, 100
Sudan, 87, 500, 703*n,* 705
Suharto, 494
Summers, Lawrence H., 75*n,* 276*n,* 315*n*
Summit of the Americas, 665
supply curve of labor facing the industrial sector, 593–94
supply response, 647
surplus labor:
  in China, 598–99*b*
  debates over, 601–3*b*
surplus labor model, 174–75
sustainable development, 700, 757–801
  concept of, 761–63, 765, 801
  global climate change and, 792–800, 801
  harvests and, 787–88
  international climate negotiations and, 799–800
  market failures and, 769–72
  policy failures and, 782–800
  policy solutions for, 773–82
  savings and, 765–67
  wealth and, 762
sustainable economic growth, 74
sustained economic development, import substitution and, 711–12
Svensson, Jakob, 296*n,* 539*n,* 706*n*
Swan, Trevor W., 546*n*
Swaroop, Vinaya, 531*n*
Swaziland, 58, 325
sweatshops, 743–47
Sweden, 412, 505, 507*b*
Swedish International Development Cooperation Agency (SIDA), 503
Switzerland, 86, 673
Tait, Alan A., 410*n*
Taiwan, 141, 142, 158, 160, 176*n,* 189, 500, 514, 666, 687, 711
  doing business in, 360
  economic growth in, 205
  financial policy in, 440
  foreign debt in, 496
  Gini coefficient for, 175, 176*f*
  land reform in, 657
  savings rate in, 378
  semiconductor industry in, 4
  trade and, 668, 671, 726, 729, 730, 741
  see also East Asian tigers
Tajikistan, 187, 241*n,* 378, 428
Tanzania, 187, 327, 478, 509, 520, 542
  agriculture in, 613, 650
  education in, 277, 287
  health in, 319
  state owned enterprises in, 396
  tax rates in, 412
tariff escalation, 686, 748
tariff peaks, 748
tariff rates, 156, 715*b*
tariffs, 145, 372, 413, 414, 710–11, 715*b,* 717, 718, 721, 722, 725, 733, 736, 738, 747, 750
  average unweighted tariff rates by region, 727*f*
  definition of, 712
  protective, 713–18
  reducing average tariffs, trend in, 727
Tarp, Finn, 522*n*
Tata Steel (India), 380
Tax, Sol, 195*n*
tax administration, changes in, 405
taxation, taxes, 18, 420, 686
  composition of tax systems by type of tax, 2009, 400*t*
  corporate income taxes, 404
  effective taxes, 412
  efficient tax system, 418–19, 419*f*
  environment and, 776–79
  excise taxes, 402–3
  FDI and, 372–73
  government expenditures and, 399–401
  government saving and, 381
  income distribution and, 411–16
  on international trade, 401–2
  neutral taxes, 418, 419–20
  new sources of tax revenues, 404–5
  personal income taxes, 404
  progessive taxes, 411
  sales taxes, 402, 403, 417

sources of inefficiency, 417–18
stamp taxes, 417
temporary migration and, 755
value-added tax, 403, 410
see also protective tariffs
tax codes, simplified, 420
tax form, Indonesian, 407–8*b*
tax holidays, 372–73, 388–89
tax reform
fundamental, 405–6, 408–11
Washington Consensus and, 147
Taylor, Alan M., 577*n*
tea, 402
technical ceiling:
in Mosher framework, 623, 624, 624*f,* 625
raising, 627–35
technical change, Solow model with, 115*f*
technically efficient crop mix, 639
technology, technological change, 16
in agriculture, 586, 627, 630, 633, 635–37*b,* 651–52*b,* 663
environment and, 759
FDI and, 370
job creation and, 10
labor-saving, 635*b*
land-saving, 635*b*
productivity and, 64
in Solow model, 113–16, 114*n,* 115*f*
trade and, 680, 689
technology diffusion, 627
temporary migration, 756
textiles, footwear, and apparel (TFA), 746–47*b*
textiles and apparel, 601–2, 686, 742, 756
market access and, 748–49
Thailand, 82, 176*n,* 189, 326, 327, 589, 656, 660, 700
baht-dollar exchange rate and, 433
beer tariff in, 718
doing business in, 360
economic growth in, 15, 56, 101*b*–102*b*
education in, 259*f,* 260
financial crisis in, 386, 457, 480, 481, 483, 484, 488, 489
HIV/AIDS prevention in, 336
intersectoral labor shifts in, 590
mean levels of life satisfaction in, 53
trade and, 726, 729, 739, 748
Theodore Schultz, 602*b,* 603*b*
Theory of Economic Growth, The (Lewis), 53
theory of induced innovation, 635*b*
Third International Mathematics and Science Study, 266
third world, 10–11
Thomas, Vinod, 75*n*
Tiananmen Square, 146
timber, 402
sustainable yields of, 787–88
time deposits, 424, 425
time profile, of net cash flow, 352, 353*f*
Timmer, C. Peter, 584, 584*n,* 587, 587*n,* 588*n,* 608*n,* 623*n,* 625*n*
Timor-Leste, 360
tin, 402
Tinbergen, Jan, 562
Tobago, 665
Toder, Eric, 404*n*
Togo, 45, 401, 681
total debt stock, 459
total economic production, 63
total factor productivity (TFP), 68–71, 73–74, 644–45
sources of growth in East Asia and other regions, 1960-2000, 72*t*
total fertility rates (TFRs), 224, 225–26*b,* 226, 227, 228, 231, 250
total liquid liabilities, 424
total suspended particulates (TSP), trading program in Chile, 781
total wealth, 762, 763
township and village industries, in China, 152
Trabajar program (Argentina), 208
tradable goods and services, 547, 548
disequilibrium in, 576
Dutch disease and, 697–99*b*
exchange-rate management and, 722
market for, 554*f,* 555
natural disasters and, 574
in open economies, 553, 561, 564
prices for, 694, 695
trade, 27–31, 665–708
aggregate gains from, 682
agriculture and, 749–50, 749*t*
balance of trade, 549
benefits of, 677–81, 678*f*
comparative advantage and, 674–83
declining terms of, 567–68, 569*f,* 690–93, 690*f*
distributional consequences of, 681–83
economic growth and, 60, 83–84, 118*b,* 666–67, 668i, 682–83, 685, 707, 736–39, 756
export orientation and, 711, 725–30, 733–34
favorable views toward, 709
foreign investment and, 685–86
geography and, 671–74, 737
importance of, to economies, 1970-2007, 669*f*

trade *(continued)*
industrial policy and, 734–36
international terms of, 677
investment and, 9, 701, 745
labor and, 683, 685, 686, 696, 742–47
to less populous nations, importance of, 672*f*
liberalization strategy, 728
markets and, 674–76, 679–81, 750–53
migration and, 753–55
net barter terms of, 692, 693*f*
policy debate about, 710–11
poverty and, 739–41, 743, 744, 756
poverty reduction and, 736
in primary products, 683–86, 684*f,* 686, 687, 707
sustainable development and, 700
technology and, 689
trends and patterns in, 667–71
winners and losers, 681–83, 740
see also import substitution; primary exports; protective tariffs
trade balance, 385, 577, 578
trade barriers, 136
trade policy(ies)
inward-oriented, 83
outward-oriented, 83, 84
Sachs-Warner measure of, 738–39
trade volumes *vs.,* 737, 739
trade protection, politics and, 718, 720–25
trade reforms, poverty and, 204
trade-related intellectual property rights (TRIPS), 751
trade volumes:
measuring, 737
trade policies *vs.,* 737, 739
traditional agriculture, 620–22, 663
Tragedy of the Commons, population growth and, 243
training of workers and managers, 368
transferable deposits, 424
transfer of technology, skills, and ideas, 367
Trans Nzoia, 650
Transparency International, 358–59, 361, 362
transportation, 666
population and, 220
transport costs, geography and, 86–87
traps, 704
Trebbi, Francesco, 89, 90*n*
Trevedi, Mandar, 788*n*
Trinidad, 665
tropical countries, growth rate per capita in, 85
tropical rain forests, loss of, 787–88
Truman, Harry, 503*b*
trypanosomiasis, 321
tuberculosis (TB), 9, 312, 321, 323, 324–25*b,* 330*b,* 333–34, 336–37, 345
Tunisia, 27, 371, 726
Turkey, 122*f,* 123, 396, 448, 538, 727
financial crisis in, 457, 480, 481, 491, 493
Washington Consensus and, 148
Turkmenistan, 430
Turner, R. Kerry, 780*n*
turning point, 594
two-country model, with a tariff, 719–20*b*
two-gap model, 131
two-sector labor surplus model, 596*f*
two-sector models of development, 590–602, 618
labor surplus model, 591–97, 596*f*
neoclassical, 599–602
typhoid, 312
Uganda, 87, 325, 432, 478, 500, 511, 613, 706
debt relief in, 478, 479*b*
education in, 287, 294, 296
Ukraine, 6–8, 9*n,* 432, 661
UNAIDS, 328
uncertainty, agricultural production and, 586
unconditional convergence, 121, 125
UNCTAD, 365*n*
underemployment, 591
under-five mortality rate, 299–300, 302
undervaluation, 557
undervalued exchange rates, 724
underweight children, 301, 344
unemployment, 427, 555–56, 558, 567, 696
labor surplus and, 591
markets and, 150
UNESCO, 284
UNICEF, 341, 525
United Arab Emirates, 700
United Climate Change Conferences, 799
United Fruit, 685, 685*n*
United Kingdom, 113, 510, 514
foreign aid through, 505
savings rate in, 375
top marginal tax rate cut in, 408
United Nations, 11, 30, 227, 231, 249, 253, 503, 508, 512
International Comparison Program, 187
System of National Accounts, 579*n*
World Population Conference, 246
United Nations Children's Fund, 512
United Nations Conference on Trade and Development (UNCTAD), 691
United Nations Development Programme (UNDP), 25, 30, 40, 42, 44, 198, 512, 524*b*

United Nations Framework Convention on Climate Change (UNFCCC), 799, 800
United Nations High Commissioner for Refugees (UNHCR), 512
United Nations Industrial Development Organization (UNIDO), 136
"United Nations Millennium Declaration," adoption of, 46
United Nations Population Fund (UNFPA), 512
United States, 12, 84, 113, 122*f,* 123, 172, 180, 226, 387, 415, 474, 475, 507*b,* 514, 661, 666, 685, 709, 747, 748, 752
  average farm size in, 653
  as benchmark for convergence, 121
  carbon dioxide emissions and, 794
  economic growth in, 444
  education in, 259–60, 259*f,* 264
  financial crisis of 2008-09 in, 8, 391, 422, 427, 494
  foreign aid and, 504, 514, 516, 543
  foreign borrowing in, 458
  foreign income taxes in, 410
  foreign savings in, 384
  global inequality and consumerism in, 212
  income distribution in, 167, 168*f*
  income inequality in, 177*n*
  life expectancy in, 312
  maintenance and replacement expenditures in, 395
  market exchange rate *vs.* purchasing power parity methods of converting GDP in, 28–30, 29*t*
  mean levels of life satisfaction in, 53
  NAFTA and, 710, 740
  population of, 231
  poverty lines in, 183–84*b,* 184, 184*n*
  sales taxes in, 417
  savings rate in, 375
  size distributions of consumption or income in, 170–71, 170*t*
  smallpox eradication and, 338
  tax avoidance in, 412
  top marginal tax rate cut in, 408
  trade and, 674, 675, 676, 677, 681, 682
  world markets and, 742
units of account, 423
upper-middle-income countries, 11, 12, 13*t*
  central government expenditure as share of GDP in, 393*t*
  composition of tax systems in, 400*t*
  see also middle-income countries
urbanization
  first indications of, 218
  health and, 323
  inequality and, 174
Uruguay, 447, 448
  financial crisis in, 480
Uruguay Round of trade negotiations, 751–52
U.S. Agency for International Development (USAID), 341, 343, 503, 525, 542, 646*b*
U.S. Census Bureau, 227*n*
U.S. Export-Import Bank, 386
U.S. Treasury bills, 387, 388
usury laws, 444
Uttar Pradesh, 194, 260
Uzbekistan, 428
vaccines, 302, 312, 319, 344
  making markets for, 330–32*b*
  polio, 339
  smallpox, 338
Vaillant, Charlotte, 650*n*
value added, 25, 715*b*
value-added taxes (VATs), 403, 410–11, 420
van der Ploeg, Frederick, 694*n,* 701*n*
Vedeld, Paul, 788, 788*n*
Venezuela, 58, 360, 361, 396, 473, 685, 706
  energy pricing and, 784
  financial crisis in, 480
  inflation in, 428
  natural resources in, 370
  SOE deficits in, 396
  Washington Consensus and, 155
vent for surplus, 685
vertical spillovers, 368
Vietnam, 144, 146, 157, 442, 509, 514, 660, 687, 718, 726, 742, 744, 782
  agriculture in, 613, 615
  changing market participation among farming households in, 614, 614*t*
  electronics industry in, 5
  foreign direct investment in, 366
  market reforms in, 160
  reform path in, 152
Vietnam War, 468
Viktor, 6–8, 10
Vincent, Jeffrey R., 766, 766*n,* 777*n,* 784*n,* 786*n,* 787, 787*n,* 795*n*
Voss, Maj-Lis, 248*n*
Vosti, Stephen A., 789, 790*n,* 791, 792

Wacziarg, Romain, 84, 84*n,* 738*n*
Wade, Robert, 729*n*
wage equations, rates of return estimated from, 275–76*b*
wages, 28
  agriculture and, 599, 601–2
  education and, 267–69, 270, 278, 525

wages (*continued*)
health and, 320
labor and, 7
in semiconductor industry, 4
temporary migration and, 753–55
trade and, 743–44
see also incomes; labor, labor sector
Wang, Hua, 759*n*
Warden, Staci, 734*n*
Warner, Andrew, 84*n,* 124, 124*n,* 198*n,* 687, 687*n,* 688, 738, 738*n*
wars, 76, 77*f,* 156, 297, 430, 504, 705
Washington Consensus, 146, 150, 152, 153, 724
components of, 147
inflation and, 442
Latin America and, 144, 146, 148, 154–55, 160, 161
poverty and, 197, 203
success or failure?, 153–59, 161
water resources, 762–63
climate change and, 798
wealth, 764*t*
comprehensive, 762, 765, 801
distribution of, 166–67
Wealth of Nations, The (Smith), 60, 129
Webb, Richard, 431*b*
Weismann, Doris M., 197*n*
Weitzman, Martin, 798, 798*n*
Welch, Karen Horn, 84, 84*n,* 738*n*
welfare weights, 356–57
Western Europe, levels of GDP per capita in, 1820-2008, 35*f*
What Works Working Group, 335
wheat, 633
Green Revolution and, 628
trends in prices of, 659, 659*f,* 660–61
Wheeler, David, 759*n*
White, Howard, 522*n*
wider band system, of exchange rate, 434, 434*f*
Williams, Roberton, 404*n*
Williamson, Jeffrey, 241*n*
Williamson, John, 146, 146*n,* 147, 434*n,* 466*n,* 477*n*
Wink, Claudia, 252*n*
Witcover, Julie, 622*n*
Wolfensohn, James, 218
Wolfers, Justin, 51*n*
women:
childbearing, wages and, 235
education and, 263, 267, 276, 298
HIV/AIDS and, 326
micro credit and, 450
population and, 233, 245
poverty and, 194–95, 214
worker-to-population ratio, 112*b*
World Bank, 11*n,* 24, 25, 36, 79*n,* 102, 137, 139*b,* 140, 142*n,* 172, 177, 192*b,* 193, 197, 206, 214, 217, 351, 358, 374, 400, 441, 459, 474, 477, 496, 511, 517, 524*b,* 525, 526, 564, 570, 627, 645, 646*b,* 651, 749, 750
barriers to doing business and, 359
Berg Report published by, 146
classification of world economies by, 11, 12, 13*t*
DOTS strategy and, 334
establishment of, 508
foreign aid and, 144, 503, 508, 509, 510, 512, 537, 538, 540, 541, 542
foreign debt and, 472, 479*b*
global warming and, 796
international poverty line of, 165*n*
international prices and, 30
Living Standards Measurement Study, 195
population projections by, 227*n*
poverty line defined by, 186
structural adjustment reforms and, 145
sustainable development and, 762*n,* 763, 775, 788
World Bank Institute, 81*b*
World Development Indicators, 3*n*
World Development Report (WDR), 186, 197, 613, 633, 647*b*
world exports, growth in, by region, 1974–2008, 668*f*
World Food Program (WFP), 512
World Health Organization (WHO), 60*b,* 249, 301, 341, 512, 524, 524*b*
health-related measures for regions, 2009, 303*t*
World Health Report (2002), 343
world price, 713
World Trade Organization (WTO), 136, 373, 402, 510*n,* 665, 681, 691, 740, 751, 752, 753
World Vision, 502
World War I, 10, 135, 135*n,* 136, 384, 465*b,* 710, 750
World War II, 135, 141, 149, 177, 220, 374, 441, 455, 466*b,* 468, 490, 508, 711, 742, 750
development thinking after, 130–32, 134–37, 159
Marshall Plan and, 503–4*b*

worm infestation, in school children, 290–91
Wriston, Walter, 465*b*
Wunder, Sven, 779, 779*n*

Yamano, Takashi, 652*b*
yellow fever, Panama Canal and, 322*b*
Yeltsin, Boris, 149
Yemen, 112*b,* 703*n*
yield, 623
yield gap, 634, 639
Young, Mei Ling, 3*n*
Younger, Stephen, 527*n*
Ythier, Jean Mercier, 539*n*
yuan, 731*b*
Yulia, 6–8, 10
Yunus, Muhammad, 449

Zaire. *see* Congo, Democratic Republic of
Zambia, 87, 325, 330, 546, 558, 684, 687, 703, 726, 741, 765
  economic growth in, 56, 58
  education in, 282, 287
  natural resources in, 370
Zedillo, Ernesto, 288*b*
Zeufack, Albert, 5*n*
Zheng He, 10
Zhu, Min, 531*n*
Zhuravskaya, Ekaterina, 53*n*
Zimbabwe, 42, 45, 156, 206, 247, 325, 326, 430, 432, 658, 765
Zoellick, Robert, 218, 218*n*
Zoido-Lobatón, Pablo, 82*n*
zones of imbalance, 555, 556*f*



Income status, 2011

Low income | Lower-middle income | Upper-middle income | High income

Source: World Bank, http://data.worldbank.org/about/country-classifications/country-and-lending-groups

SOUTH ASIA



Income status, 2011

Low income | Lower-middle income | Upper-middle income | High income

Source: World Bank, http://data.worldbank.org/about/country-classifications/country-and-lending-groups

EAST ASIA