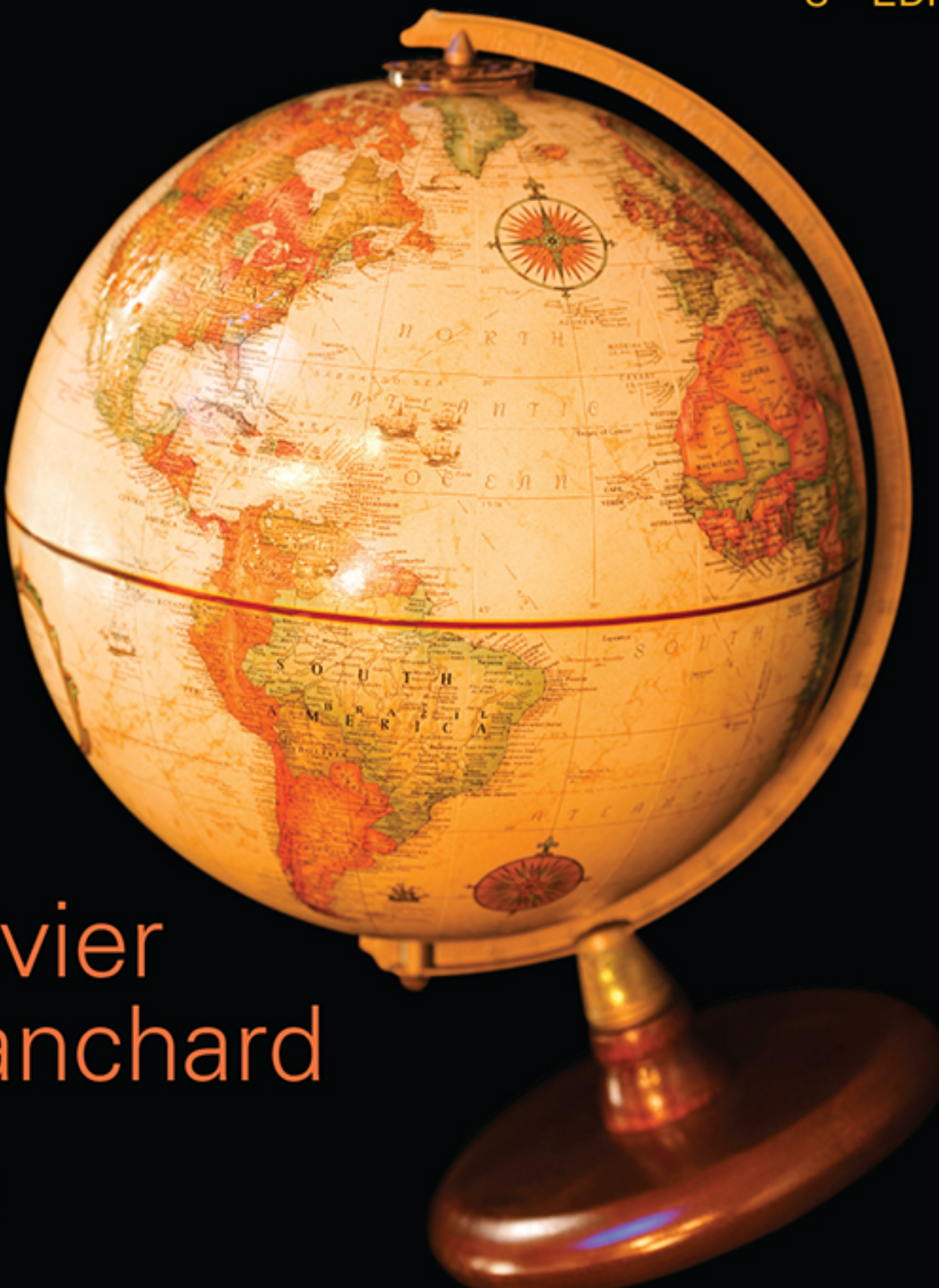
MACROECONOMICS

8TH EDITION



Olivier Blanchard

Eighth Edition

MACROECONOMICS

Olivier Blanchard

Pearson

Library of Congress Cataloging-in-Publication Data

Name: Blanchard, Olivier (Olivier J.), author.
Title: Macroeconomics / Olivier Blanchard.
Description: Eighth Edition. | Hoboken : Pearson, [2021] | Revised edition of Macroeconomics, [2017]
Identifiers: LCCN 2019015670| ISBN 9780134897899 | ISBN 0134897897
Subjects: LCSH: Macroeconomics.
Classification: LCC HB172.5 .B573 2019 | DDC 339—dc23 LC record available at https://lccn.loc.gov/2019015670

ScoutAutomatedPrintCode

Pearson

ISBN 10: 0-13-489789-7
ISBN 13: 978-0-13-489789-9

To Noelle

Flexible Organization

Macroeconomics, eighth edition is organized around two central parts: A core and a set of two major extensions. The text's **flexible organization** emphasizes an integrated view of macroeconomics, while enabling professors to focus on the theories, models, and applications that they deem central to their particular course.

The flowchart below quickly illustrates how the chapters are organized and fit within the book's overall structure. For a more detailed explanation of the **Flexible Organization**, and for an extensive list of **Alternative Course Outlines**, see pages **xiii–xiv** in the preface.

INTRODUCTION

A Tour of the World **Chapter 1**
A Tour of the Book **Chapter 2**

THE CORE

The Short Run

The Goods Market **Chapter 3**
Financial Markets I **Chapter 4**
Goods and Financial Markets: The *IS-LM* Model **Chapter 5**
Financial Markets II: The Extended *IS-LM* Model **Chapter 6**

The Medium Run

The Labor Market **Chapter 7**
The Phillips Curve, the Natural Rate of Unemployment, and Inflation **Chapter 8**
From the Short to the Medium Run: The IS-LM PC Model **Chapter 9**

The Long Run

The Facts of Growth **Chapter 10**
Saving, Capital Accumulation, and Output **Chapter 11**
Technological Progress and Growth **Chapter 12**
The Challenges of Growth **Chapter 13**

EXTENSIONS

EXPECTATIONS

Financial Markets and Expectations **Chapter 14**
Expectations, Consumption, and Investment **Chapter 15**
Expectations, Output, and Policy **Chapter 16**

THE OPEN ECONOMY

Openness in Goods and Financial Markets **Chapter 17**
The Goods Market in an Open Economy **Chapter 18**
Output, the Interest Rate, and the Exchange Rate **Chapter 19**
Exchange Rate Regimes **Chapter 20**

BACK TO POLICY

Should Policymakers Be Restrained? **Chapter 21**
Fiscal Policy: A Summing Up **Chapter 22**
Monetary Policy: A Summing Up **Chapter 23**

EPILOGUE

The Story of Macroeconomics **Chapter 24**

Brief Contents

THE CORE

Introduction 1

Chapter 1 A Tour of the World 3
Chapter 2 A Tour of the Book 19

The Short Run 43

Chapter 3 The Goods Market 45
Chapter 4 Financial Markets I 67
Chapter 5 Goods and Financial Markets: The IS-LM Model 87
Chapter 6 Financial Markets II: The Extended IS-LM Model 109

The Medium Run 133

Chapter 7 The Labor Market 135
Chapter 8 The Phillips Curve, the Natural Rate of Unemployment, and Inflation 155
Chapter 9 From the Short to the Medium Run: The IS-LM-PC Model 177

The Long Run 199

Chapter 10 The Facts of Growth 201
Chapter 11 Saving, Capital Accumulation, and Output 219
Chapter 12 Technological Progress and Growth 243
Chapter 13 The Challenges of Growth 265

EXTENSIONS

Expectations 285

Chapter 14 Financial Markets and Expectations 287
Chapter 15 Expectations, Consumption, and Investment 313
Chapter 16 Expectations, Output, and Policy 335

The Open Economy 353

Chapter 17 Openness in Goods and Financial Markets 355
Chapter 18 The Goods Market in an Open Economy 375
Chapter 19 Output, the Interest Rate, and the Exchange Rate 397
Chapter 20 Exchange Rate Regimes 419

Back to Policy 441

Chapter 21 Should Policymakers Be Restrained? 443
Chapter 22 Fiscal Policy: A Summing Up 461
Chapter 23 Monetary Policy: A Summing Up 485
Chapter 24 Epilogue: The Story of Macroeconomics 507

Contents

Preface xiii
About the Author xviii

THE CORE

Introduction 1

Chapter 1 A Tour of the World 3

1-1 The Crisis 4
1-2 The United States 6
Do Policymakers Have the Tools to Handle the Next Recession? 7 • How Worrisome Is Low Productivity Growth? 8
1-3 The Euro Area 9
Can European Unemployment Be Reduced? 11 • What Has the Euro Done for Its Members? 12
1-4 China 13
1-5 Looking Ahead 15
Appendix 1: Where to Find the Numbers 18
Appendix 2: What Do Macroeconomists Do? 18

Chapter 2 A Tour of the Book 19

2-1 Aggregate Output 20
GDP: Production and Income 20 • Nominal and Real GDP 22 • GDP: Level versus Growth Rate 24
2-2 The Unemployment Rate 25
Why Do Economists Care about Unemployment? 26
2-3 The Inflation Rate 27
The GDP Deflator 27 • The Consumer Price Index 29 • Why Do Economists Care about Inflation? 30
2-4 Output, Unemployment, and the Inflation Rate: Okun's Law and the Phillips Curve 31
Okun's Law 31 • The Phillips Curve 32
2-5 The Short Run, the Medium Run, and the Long Run 33
2-6 A Tour of the Book 34
The Core 35 • Extensions 35 • Back to Policy 35 • Epilogue 36
Appendix: The Construction of Real GDP and Chain-Type Indexes 40

The Short Run 43

Chapter 3 The Goods Market 45

3-1 The Composition of GDP 46
3-2 The Demand for Goods 47
Consumption (*C*) 48 • Investment (*I*) 50 • Government Spending (*G*) 50
3-3 The Determination of Equilibrium Output 51
Using Algebra 52 • Using a Graph 53 • Using Words 55 • How Long Does It Take for Output to Adjust? 56
3-4 Investment Equals Saving: An Alternative Way of Thinking about Goods-Market Equilibrium 58
3-5 Is the Government Omnipotent? A Warning 60

Chapter 4 Financial Markets I 67

4-1 The Demand for Money 68
Deriving the Demand for Money 69
4-2 Determining the Interest Rate: I 71
Money Demand, Money Supply, and the Equilibrium Interest Rate 71 • Monetary Policy and Open Market Operations 73 • Choosing Money or Choosing the Interest Rate? 75
4-3 Determining the Interest Rate: II 76
What Banks Do 76 • The Demand for and Supply of Central Bank Money 78 • The Demand for Central Bank Money 78 • The Federal Funds Market and the Federal Funds Rate 79
4-4 The Liquidity Trap 81

Chapter 5 **Goods and Financial Markets: The IS-LM Model 87**

5-1 The Goods Market and the IS Relation 88

Investment, Sales, and the Interest Rate 88 • Determining Output 89 • Deriving the IS Curve 91 • Shifts of the IS Curve 91

5-2 Financial Markets and the LM Relation 92

Real Money, Real Income, and the Interest Rate 92 • Deriving the LM Curve 93

5-3 Putting the IS and LM Relations Together 94

Fiscal Policy 94 • Monetary Policy 96

5-4 Using a Policy Mix 96

5-5 How Does the IS-LM Model Fit the Facts? 102

Chapter 6 **Financial Markets II: The Extended IS-LM Model 109**

6-1 Nominal versus Real Interest Rates 110

Nominal and Real Interest Rates in the United States since 1978 112 • Nominal and Real Interest Rates: The Zero Lower Bound and Deflation 113

6-2 Risk and Risk Premia 114

6-3 The Role of Financial Intermediaries 115

The Choice of Leverage 116 • Leverage and Lending 117

6-4 Extending the IS-LM Model 119

Financial Shocks and Policies 121

6-5 From a Housing Problem to a Financial Crisis 122

Housing Prices and Subprime Mortgages 122 • The Role of Financial Intermediaries 123 • Macroeconomic Implications 125 • Policy Responses 125

The Medium Run 133

Chapter 7 **The Labor Market 135**

7-1 A Tour of the Labor Market 136

The Large Flows of Workers 136

7-2 Movements in Unemployment 139

7-3 Wage Determination 141

Bargaining 142 • Efficiency Wages 142 • Wages, Prices, and Unemployment 144 • The Expected Price Level 144 • The Unemployment Rate 145 • The Other Factors 145

7-4 Price Determination 145

7-5 The Natural Rate of Unemployment 146

The Wage-Setting Relation 146 • The Price-Setting Relation 147 • Equilibrium Real Wages and Unemployment 148

7-6 Where We Go from Here 149

Appendix: Wage- and Price-Setting Relations versus Labor Supply and Labor Demand 154

Chapter 8 **The Phillips Curve, the Natural Rate of Unemployment, and Inflation 155**

8-1 Inflation, Expected Inflation, and Unemployment 156

8-2 The Phillips Curve and Its Mutations 158

The Original Phillips Curve 158 • The De-anchoring of Expectations 158 • The Re-anchoring of Expectations 161

8-3 The Phillips Curve and the Natural Rate of Unemployment 162

8-4 A Summary and Many Warnings 164

Variations in the Natural Rate over Time 165 • Variations in the Natural Unemployment Rate across Countries 165 • High Inflation and the Phillips Curve 166 • Deflation and the Phillips Curve 169

Appendix: Derivation of the Relation Between Inflation, Expected Inflation, and Unemployment 175

Chapter 9 **From the Short to the Medium Run: The IS-LM-PC Model 177**

9-1 The IS-LM-PC Model 178

9-2 From the Short to the Medium Run 181

9-3 Complications and How Things Can Go Wrong 183

The Zero Lower Bound and Deflation Spirals 184

9-4 Fiscal Consolidation Revisited 185

9-5 The Effects of an Increase in the Price of Oil 188

Effects on the Natural Rate of Unemployment 189

9-6 Conclusions 191

The Short Run versus the Medium Run 193 • Shocks and Propagation Mechanisms 193

The Long Run 199

Chapter 10 The Facts of Growth 201

10-1 Measuring the Standard of Living 202

10-2 Growth in Rich Countries since 1950 205

The Large Increase in the Standard of Living since 1950 207 • The Convergence of Output per Person 208

10-3 A Broader Look across Time and Space 209

Looking across Two Millennia 209 • Looking across Countries 209

10-4 Thinking about Growth: A Primer 212

The Aggregate Production Function 212 • Returns to Scale and Returns to Factors 213 • Output per Worker and Capital per Worker 213 • The Sources of Growth 214

Chapter 11 Saving, Capital Accumulation, and Output 219

11-1 Interactions between Output and Capital 220

The Effects of Capital on Output 220 • The Effects of Output on Capital Accumulation 221 • Output and Investment 221 • Investment and Capital Accumulation 222

11-2 The Implications of Alternative Saving Rates 223

Dynamics of Capital and Output 223 • The Saving Rate and Output 225 • The Saving Rate and Consumption 228

11-3 Getting a Sense of Magnitudes 230

The Effects of the Saving Rate on Steady-State Output 232 • The Dynamic Effects of an Increase in the Saving Rate 233 • The US Saving Rate and the Golden Rule 235

11-4 Physical versus Human Capital 236

Extending the Production Function 236 • Human Capital, Physical Capital, and Output 237 • Endogenous Growth 238

Appendix: The Cobb-Douglas Production Function and the Steady State 241

Chapter 12 Technological Progress and Growth 243

12-1 Technological Progress and the Rate of Growth 244

Technological Progress and the Production Function 244 • Interactions between Output and Capital 246 • Dynamics of Capital and Output 248 • Back to the Effects of the Saving Rate 249

12-2 The Determinants of Technological Progress 250

The Fertility of the Research Process 251 • The Appropriability of Research Results 252 • Innovation versus Imitation 254

12-3 Institutions, Technological Progress, and Growth 255

Appendix: How to Measure Technological Progress, and the Application to China 262

Chapter 13 The Challenges of Growth 265

13-1 The Future of Technological Progress 266

13-2 Robots and Unemployment 267

13-3 Growth, Churn, and Inequality 269

The Increase in Wage Inequality 271 • The Causes of Increased Wage Inequality 272 • Inequality and the Top 1% 274 • Growth and Inequality 275

13-4 Climate Change and Global Warming 278

EXTENSIONS

Expectations 285

Chapter 14 Financial Markets and Expectations 287

14-1 Expected Present Discounted Values 288

Computing Expected Present Discounted Values 288 • A General Formula 289 • Using Present Values: Examples 290 • Constant Interest Rates 290 • Constant Interest Rates and Payments 290 • Constant Interest Rates and Payments Forever 291 • Zero Interest Rates 291 • Nominal versus Real Interest Rates and Present Values 291

14-2 Bond Prices and Bond Yields 292

Bond Prices as Present Values 294 • Arbitrage and Bond Prices 295 • From Bond Prices to Bond Yields 296 • Reintroducing Risk 297 • Interpreting the Yield Curve 298

14-3 The Stock Market and Movements in Stock Prices 299

Stock Prices as Present Values 300 • The Stock Market and Economic Activity 302 • A Monetary Expansion and the Stock Market 303 • An Increase in Consumer Spending and the Stock Market 304

14-4 Risk, Bubbles, Fads, and Asset Prices 305

Stock Prices and Risk 306 • Asset Prices, Fundamentals, and Bubbles 306

Appendix: Deriving the Expected Present Discounted Value Using Real or Nominal Interest Rates 312

Chapter 15 Expectations, Consumption, and Investment 313

15-1 Consumption 314

The Very Foresighted Consumer 314 • An Example 315 • Toward a More Realistic Description 316 • Putting Things Together: Current Income, Expectations, and Consumption 319

15-2 Investment 320

Investment and Expectations of Profit 320 • Depreciation 321 • The Present Value of Expected Profits 321 • The Investment Decision 322 • A Convenient Special Case 322 • Current versus Expected Profit 324 • Profit and Sales 326

15-3 The Volatility of Consumption and Investment 328

Appendix: Derivation of the Expected Present Value of Profits under Static Expectations 333

Chapter 16 Expectations, Output, and Policy 335

16-1 Expectations and Decisions: Taking Stock 336

Expectations, Consumption, and Investment Decisions 336 • Expectations and the IS Relation 336

16-2 Monetary Policy, Expectations, and Output 339

Monetary Policy Revisited 339

16-3 Deficit Reduction, Expectations, and Output 342

The Role of Expectations about the Future 343 • Back to the Current Period 343

The Open Economy 353

Chapter 17 Openness in Goods and Financial Markets 355

17-1 Openness in Goods Markets 356

Exports and Imports 356 • The Choice between Domestic Goods and Foreign Goods 358 • Nominal Exchange Rates 358 • From Nominal to Real Exchange Rates 359 • From Bilateral to Multilateral Exchange Rates 363

17-2 Openness in Financial Markets 364

The Balance of Payments 365 • The Choice between Domestic and Foreign Assets 367 • Interest Rates and Exchange Rates 369

17-3 Conclusions and a Look Ahead 371

Chapter 18 The Goods Market in an Open Economy 375

18-1 The IS Relation in the Open Economy 376

The Demand for Domestic Goods 376 • The Determinants of *C*, *I*, and *G* 376 • The Determinants of Imports 377 • The Determinants of Exports 377 • Putting the Components Together 377

18-2 Equilibrium Output and the Trade Balance 379

18-3 Increases in Demand—Domestic or Foreign 380

Increases in Domestic Demand 380 • Increases in Foreign Demand 382 • Fiscal Policy Revisited 383

18-4 Depreciation, the Trade Balance, and Output 385

Depreciation and the Trade Balance: the Marshall-Lerner Condition 386 • The Effects of a Real Depreciation 386 • Combining Exchange Rate and Fiscal Policies 387

18-5 Saving, Investment, and the Current Account Balance 390

Appendix: Derivation of the Marshall-Lerner Condition 395

Chapter 19 Output, the Interest Rate, and the Exchange Rate 397

19-1 Equilibrium in the Goods Market 398

19-2 Equilibrium in Financial Markets 399

Domestic Bonds versus Foreign Bonds 399

19-3 Putting Goods and Financial Markets Together 402

19-4 The Effects of Policy in an Open Economy 404

The Effects of Monetary Policy in an Open Economy 404 • The Effects of Fiscal Policy in an Open Economy 405

19-5 Fixed Exchange Rates 409

Pegs, Crawling Pegs, Bands, the EMS, and the Euro 409 • Monetary Policy When the Exchange Rate Is Fixed 410 • Fiscal Policy When the Exchange Rate Is Fixed 411

Appendix: Fixed Exchange Rates, Interest Rates, and Capital Mobility 416

Chapter 20 Exchange Rate Regimes 419

20-1 The Medium Run 420

The IS Relation under Fixed Exchange Rates 421 • Equilibrium in the Short and the Medium Run 421 • The Case for and against a Devaluation 422

20-2 Exchange Rate Crises under Fixed Exchange Rates 424

20-3 Exchange Rate Movements under Flexible Exchange Rates 427

Exchange Rates and the Current Account 428 • Exchange Rates and Current and Future Interest Rates 429 • Exchange Rate Volatility 429

20-4 Choosing between Exchange Rate Regimes 430

Common Currency Areas 431 • Hard Pegs, Currency Boards, and Dollarization 433

Appendix 1: Deriving the IS Relation under Fixed Exchange Rates 439

Appendix 2: The Real Exchange Rate and Domestic and Foreign Real Interest Rates 439

Back to Policy 441

Chapter 21 Should Policymakers Be Restrained? 443

21-1 Uncertainty and Policy 444

How Much Do Macroeconomists Actually Know? 444 • Should Uncertainty Lead Policymakers to Do Less? 446 • Uncertainty and Restraints on Policymakers 446

21-2 Expectations and Policy 447

Hostage Takings and Negotiations 448 • Inflation and Unemployment Revisited 448 • Establishing Credibility 449 • Time Consistency and Restraints on Policymakers 451

21-3 Politics and Policy 451

Games between Policymakers and Voters 452 • Games between Policymakers 453 • Politics and Fiscal Restraints 454

Chapter 22 Fiscal Policy: A Summing Up 461

22-1 What We Have Learned 462

22-2 The Government Budget Constraint: Deficits, Debt, Spending, and Taxes 463

The Arithmetic of Deficits and Debt 463 • Current versus Future Taxes 465 • Full Repayment in Year 2 465 • Full Repayment in Year *t* 466 • Debt Stabilization in Year *2* 467 • The Evolution of the Debt-to-GDP Ratio 468

22-3 Ricardian Equivalence, Cyclical Adjusted Deficits, and War Finance 470

Ricardian Equivalence 470 • Deficits, Output Stabilization, and the Cyclically Adjusted Deficit 471 • Wars and Deficits 472

22-4 The Dangers of High Debt 474

High Debt, Default Risk, and Vicious Cycles 474 • Debt Default 476 • Money Finance 476

22-5 The Challenges Facing US Fiscal Policy Today 479

Chapter 23 Monetary Policy: A Summing Up 485

23-1 What We Have Learned 486

23-2 From Money Targeting to Inflation Targeting 487

Money Targeting 487 • Inflation Targeting 489 • The Interest Rate Rule 490

23-3 The Optimal Inflation Rate 491

The Costs of Inflation 491 • The Benefits of Inflation 494 • The Optimal Inflation Rate: The State of the Debate 495

23-4 Unconventional Monetary Policy 496

Monetary Policy since the End of the Liquidity Trap 497

23-5 Monetary Policy and Financial Stability 498

Liquidity Provision and Lender of Last Resort 499 • Macroprudential Tools 499

Chapter 24 **Epilogue: The Story of Macroeconomics 507**

24-1 Keynes and the Great Depression 508

24-2 The Neoclassical Synthesis 508

Progress on All Fronts 509 • Keynesians versus Monetarists 510

24-3 The Rational Expectations Critique 511

The Three Implications of Rational Expectations 512 • The Integration of Rational Expectations 513

24-4 Developments in Macroeconomics up to the 2009 Crisis 514

New Classical Economics and Real Business Cycle Theory 515 • New Keynesian Economics 515 • New Growth Theory 516 • Toward an Integration 517

24-5 First Lessons for Macroeconomics after the Crisis 518

Appendix 1 **An Introduction to the National Income and Product Accounts A-1**

Appendix 2 **A Math Refresher A-7**

Appendix 3 **An Introduction to Econometrics A-12**

Glossary G-1

Index I-1

Credits C-1

Symbols Used in This Book S-1

Focus Boxes

Real GDP, Technological Progress, and the Price of Computers 25
Unemployment and Happiness 28
The Lehman Bankruptcy, Fears of Another Great Depression, and Shifts in the Consumption Function 57
The Paradox of Saving 61
Semantic Traps: Money, Income, and Wealth 69
Who Holds US Currency? 71
Will Bitcoins Replace Dollars? 80
The Liquidity Trap in Action 82
The US Recession of 2001 98
Deficit Reduction: Good or Bad for Investment? 101
Bank Runs 118
The Current Population Survey 138
From Henry Ford to Jeff Bezos 143
Theory ahead of Facts: Milton Friedman and Edmund Phelps 163
Changes in the US Natural Rate of Unemployment since 1990 166
What Explains European Unemployment? 167
Okun's Law across Time and Countries 180
Deflation in the Great Depression 186
Oil Price Increases: Why Were the 2000s So Different from the 1970s? 192
The Construction of PPP Numbers 204
Does Money Buy Happiness? 206
The Reality of Growth: A US Workingman's Budget in 1851 210
Capital Accumulation and Growth in France in the Aftermath of World War II 226
Social Security, Saving, and Capital Accumulation in the United States 231
Nudging US Households to Save More 236
The Diffusion of New Technology: Hybrid Corn 252
Management Practices: Another Dimension of Technological Progress 254
The Importance of Institutions: North Korea and South Korea 256
What Lies Behind Chinese Growth? 258
Job Destruction, Churn and Earnings Losses 270
The Long View: Technology, Education, and Inequality 273
Inequality and the Gini Coefficient 276
The Vocabulary of Bond Markets 294
Making (Some) Sense of (Apparent) Nonsense: Why the Stock Market Moved Yesterday, and Other Stories 305
Famous Bubbles: From Tulipmania in 17th-Century Holland to Russia in 1994 306
The Increase in US Housing Prices During the First Half of the 2000s: Fundamentals or Bubble? 308
Up Close and Personal: Learning from Panel Datasets 315
Do People Save Enough for Retirement? 318
Investment and the Stock Market 323
Profitability versus Cash Flow 326
Rational Expectations 341
Can a Budget Deficit Reduction Lead to an Output Expansion? Ireland in the 1980s 345
Uncertainty and Fluctuations 348
Can Exports Exceed GDP? 358
GDP versus GNP: The Example of Kuwait 368
Buying Brazilian Bonds 370
The G20 and the 2009 Fiscal Stimulus 384
The Disappearance of the Current Account Deficit in Greece: Good News or Bad News? 388
Capital Flows, Sudden Stops, and the Limits to the Interest Parity Condition 400
Monetary Contraction and Fiscal Expansion: The United States in the Early 1980s 407
US Trade Deficits and Trump Administration Trade Tariffs 408
German Reunification, Interest Rates, and the EMS 412
The Return of Britain to the Gold Standard: Keynes versus Churchill 423
The 1992 EMS Crisis 426
The Euro: A Short History 433
Lessons from Argentina's Currency Board 434
Was Alan Blinder Wrong in Speaking the Truth? 451
Euro Area Fiscal Rules: A Short History 456
Inflation Accounting and the Measurement of Deficits 464
How Countries Decreased Their Debt Ratios after World War II 469
Deficits, Consumption, and Investment in the United States during World War II 473
Money Financing and Hyperinflations 478
Money Illusion 493
LTV Ratios and Housing Price Increases from 2000 to 2007 501

Preface

I had two main goals in writing this book:

- To make close contact with current macroeconomic events. What makes macroeconomics exciting is the light it sheds on what is happening around the world, from the major economic crisis which engulfed the world in the late 2000s, to monetary policy in the United States, to the problems of the Euro area, to growth in China. These events—and many more—are described in the book, not in footnotes, but in the text or in detailed boxes. Each box shows how you can use what you have learned to get an understanding of these events. My belief is that these boxes not only convey the life of macroeconomics, but also reinforce the lessons from the models, making them more concrete and easier to grasp.

- To provide an integrated view of macroeconomics. The book is built on one underlying model, a model that draws the implications of equilibrium conditions in three sets of markets: the goods market, the financial markets, and the labor market. Depending on the issue at hand, the parts of the model relevant to the issue are developed in more detail while the other parts are simplified or lurk in the background. But the underlying model is always the same. This way, you will see macroeconomics as a coherent whole, not a collection of models. And you will be able to make sense not only of past macroeconomic events but also of those that unfold in the future.

Solving Learning and Teaching Challenges

Flexible Organization

The book is organized around two central parts: a core, and a set of two major extensions. An introduction precedes the core. The two extensions are followed by a review of the role of policy. The book ends with an epilogue. The flowchart at the beginning of this book makes it easy to see how the chapters are organized and fit within the book's overall structure.

- Chapters 1 and 2 introduce the **basic facts and issues of macroeconomics**. Chapter 1 takes you on an economic tour of the world, from the United States, to the Euro area, and to China. Some instructors will prefer to cover Chapter 1 later, perhaps after Chapter 2, which introduces basic concepts, articulates the notions of short run, medium run, and long run, and gives the reader a quick tour of the book.

  While Chapter 2 gives the basics of national income accounting, I have put a detailed treatment of national income accounts in Appendix 1 at the end of the book. This decreases the burden on the beginning reader and allows for a more thorough treatment in the appendix.

- Chapters 3 through 13 constitute the **core**.

  Chapters 3 through 6 focus on the **short run**. Chapters 3 to 5 characterize equilibrium in the goods market and in the financial markets, and derive the basic model used to study short-run movements in output, the IS-LM model. Chapter 6 extends the basic IS-LM model to reflect the role of the financial system. It then uses it to describe what happened during the initial phase of the financial crisis.

  Chapters 7 through 9 focus on the **medium run**. Chapter 7 focuses on equilibrium in the labor market and introduces the notion of the natural rate of unemployment. Chapter 8 derives and discusses the relation between unemployment and inflation, known as the Phillips curve. Chapter 9 develops the IS-LM-PC (PC for Phillips curve) model which takes into account equilibrium in the goods market, in the financial markets, and in the labor market. It shows how this model can be used to understand movements in activity and movements in inflation, both in the short and in the medium run.

  Chapters 10 through 13 focus on the **long run**. Chapter 10 describes the facts, showing the evolution of output across countries and over long periods of time. Chapters 11 and 12 develop a model of growth and describe how capital accumulation and technological progress determine growth. Chapter 13, which is new, focuses on the challenges to growth, from inequality to climate change.

- Chapters 14 through 20 cover the two major **extensions**.

  Chapters 14 through 16 focus on the role of expectations in the short run and in the medium run. Expectations play a major role in most economic decisions and, by implication, play a major role in the determination of output.

Chapters 17 through 20 examine the implications of openness of modern economies. Chapter 20 looks at the implications of different exchange rate regimes, from flexible exchange rates, to fixed exchange rates, currency boards, and dollarization.

- Chapters 21 through 23 return to **macroeconomic policy**. Although most of the first 20 chapters constantly discuss macroeconomic policy in one form or another, the purpose of Chapters 21 through 23 is to tie the threads together. Chapter 21 looks at the role and the limits of macroeconomic policy in general. Chapters 22 and 23 review fiscal and monetary policy. Some instructors may want to use parts of these chapters earlier. For example, it is easy to move forward the discussion of the government budget constraint in Chapter 22 or the discussion of inflation targeting in Chapter 23.
- Chapter 24 serves as an **epilogue**; it puts macroeconomics in historical perspective by showing the evolution of macroeconomics over the last 80 years, discussing current directions of research, and the lessons of the crisis for macroeconomics.

Alternative Course Outlines

Within the book's broad organization, there is plenty of opportunity for alternative course organizations. I have made the chapters shorter than is standard in textbooks, and, in my experience, most chapters can be covered in an hour and a half. A few (Chapters 5 and 9 for example) might require two lectures to sink in.

- Short courses (15 lectures or less)

  A short course can be organized around the two introductory chapters and the core (Chapter 13 can be excluded at no cost in continuity). Informal presentations of one or two of the extensions, based, for example, on Chapter 16 for expectations (which can be taught as a standalone) and on Chapter 17 for the open economy, can then follow, for a total of 14 lectures.

  A short course might leave out the study of growth (the long run). In this case, the course can be organized around the introductory chapters and Chapters 3 through 9 in the core; this gives a total of 9 lectures, leaving enough time to cover, for example, Chapter 16 on expectations and Chapters 17 through 19 on the open economy, for a total of 13 lectures.

- Longer courses (20 to 25 lectures)

  A full semester course gives more than enough time to cover the core, plus one or both of the two extensions, and the review of policy. The extensions assume knowledge of the core, but are otherwise mostly self-contained. Given the choice, the order in which they are best taught is probably the order in which they are presented in the book. Having studied the role of expectations first helps students to understand the interest parity condition and the nature of exchange rate crises.

Innovative Features

I have made sure never to present a theoretical result without relating it to the real world. In addition to discussions of facts in the text itself, I have written many Focus boxes, which discuss particular macroeconomic events or facts from the United States or from around the world. Many of those are new to this edition.

I have tried to re-create some of the student–teacher interactions that take place in the classroom by the use of margin notes, which run parallel to the text. The margin notes create a dialogue with the reader and, in so doing, smooth the more difficult passages and give a deeper understanding of the concepts and the results derived along the way.

For students who want to explore macroeconomics further, I have introduced the following two features:

- Short appendixes to some chapters, which expand on points made within the chapter.
- A 'Further Readings' section at the end of most chapters, indicating where to find more information, including key Internet addresses.

Each chapter starts with a one- or two-sentence summary at the end of the introduction, and ends with three ways of making sure that the material in the chapter has been digested:

- A summary of the chapter's main points.
- A list of key terms.
- A series of end-of-chapter exercises. "Quick Check" exercises are easy. "Dig Deeper" exercises are a bit harder, and "Explore Further" activities typically require either access to the Internet or use of a spreadsheet program.
- A list of symbols at the end of the book makes it easy to recall the meaning of the symbols used in the text.

What's New in This Edition

A new Chapter 13 on the challenges to growth. Topics include whether the introduction of robots will lead to mass unemployment, the relation between growth and inequality, and the challenges of climate change.

A revised Chapter 8 on the Phillips curve, reflecting a major change in the US economy. The Phillips curve is now a relation between inflation and unemployment rather than between the change in inflation and unemployment.

A revised Chapter 9 showing how the changes in the Phillips curve relation have led to changes in monetary policy.

A new appendix in Chapter 1, 'What Do Macroeconomists Do?', which will give you a sense of what careers you may pursue if you were to specialize in macroeconomics.

Updated Focus Boxes include:

- NEW! Will Bitcoins Replace Dollars? (Chapter 4)
- From Henry Ford to Jeff Bezos (Chapter 7)
- NEW! Nudging US Households to Save More (Chapter 11)
- What Lies Behind Chinese Growth? (Chapter 12)
- Uncertainty and Fluctuations (Chapter 16)
- NEW! US Trade Deficits and Trump Administration Trump Tariffs (Chapter 19)

Acknowledgments and Thanks

This book owes much to many. I thank Adam Ashcraft, Peter Berger, Peter Benczur, Efe Cakarel, Francesco Furno, Harry Gakidis, Ava Hong, David Hwang, Kevin Nazemi, David Reichsfeld, Jianlong Tan, Stacy Tevlin, Gaurav Tewari, Corissa Thompson, John Simon, and Jeromin Zettelmeyer for their research assistance over the years. I thank the generations of students in 14.02 at MIT who have freely shared their reactions to the book over the years.

I have benefited from comments from many colleagues and friends. Among them are John Abell, Daron Acemoglu, Tobias Adrian, Chuangxin An, Roland Benabou, Samuel Bentolila, and Juan Jimeno (who have adapted the book for a Spanish edition); Francois Blanchard, Roger Brinner, Ricardo Caballero, Wendy Carlin, Martina Copelman, Henry Chappell, Ludwig Chincarini, and Daniel Cohen (who has adapted the book for a French edition); Larry Christiano, Bud Collier, Andres Conesa, Peter Diamond, Martin Eichenbaum, Gary Fethke, David Findlay, Francesco Giavazzi, and Alessia Amighini (who adapted the book first for an Italian edition, and then for a European edition); Andrew Healy, Steinar Holden, and Gerhard Illing (who has adapted the book for a German edition); Yannis Ioannides, Jose Isidoro Garcia de Paso Gomez, Per Krusell (who, with others, has adapted the book for a Swedish edition), Angelo Melino (who adapted the book for a Canadian edition), P. N. Junankar, Sam Keeley, Bernd Kuemmel, Paul Krugman, Antoine Magnier, Peter Montiel, Bill Nordhaus, Tom Michl, Dick Oppermann, Athanasios Orphanides, and Daniel Pirez Enri (who has adapted the book for a Latin American edition); Michael Plouffe, Zoran Popovic, Jim Poterba, and Jeff Sheen (who has adapted the book for an Australasian edition); Ronald Schettkat, and Watanabe Shinichi (who has adapted the book for a Japanese edition); Francesco Sisci, Brian Simboli, Changyong Rhee, Julio Rotemberg, Robert Solow, Andre Watteyne, and Michael Woodford. Thanks also go to David Johnson who wrote the end-of-chapter exercises for this edition.

I want to thank the reviewers that provided feedback for this new edition:

- Sahar Bahmani, University of Wisconsin Parkside
- Robert Blecker, American University
- LaTanya Brown-Robertson, Bowie State University
- Martina Copelman, University of Maryland, College Park
- Satyajit Ghosh, University of Scranton
- Ting Levy, Florida Atlantic University
- Sokchea Lim, John Carroll University
- Yu Peng Lin, University of Detroit-Mercy

I have also benefited from comments from many readers, reviewers, and class testers. Among them:

- John Abell, Randolph, Macon Woman's College
- Carol Adams, Cabrillo College
- Gilad Aharonovitz, School of Economic Sciences
- Terence Alexander, Iowa State University
- Roger Aliaga-Diaz, Drexel University
- Robert Archibald, College of William & Mary
- John Baffoe-Bonnie, La Salle University
- Fatolla Bagheri, University of North Dakota
- Stephen Baker, Capital University
- Erol Balkan, Hamilton College
- Jennifer Ball, Washburn University
- Richard Ballman, Augustana College
- King Banaian, St. Cloud State University
- Charles Bean, London School of Economics and Political Science
- Scott Benson, Idaho State University
- Gerald Bialka, University of North Florida
- Robert Blecker, American University

- Scott Bloom, North Dakota State University
- Pim Borren, University of Canterbury, New Zealand
- LaTanya Brown-Robertson, Bowie State University
- James Butkiewicz, University of Delaware
- Colleen Callahan, American University
- Bruce Carpenter, Mansfield University
- Kyongwook Choi, Ohio University College
- Michael Cook, William Jewell College
- Nicole Crain, Lafayette College
- Rosemary Cunningham, Agnes Scott College
- Evren Damar, Pacific Lutheran University
- Dale DeBoer, University of Colorado at Colorado Springs
- Adrian de Leon-Arias, Universidad de Guadalajara
- Brad DeLong, UC Berkeley
- Firat Demir, University of Oklahoma
- Wouter Denhaan, UC San Diego
- John Dodge, King College
- F. Trenery Dolbear, Brandeis University
- Patrick Dolenc, Keene State College
- Brian Donhauser, University of Washington
- Michael Donihue, Colby College
- Vincent Dropsy, California State University
- Justin Dubas, St. Norbert College
- Amitava Dutt, University of Notre Dame
- John Edgren, Eastern Michigan University
- Eric Elder, Northwestern College
- Sharon J. Erenburg, Eastern Michigan University
- Antonina Espiritu, Hawaii Pacific University
- J. Peter Federer, Clark University
- Rendigs Fels, Vanderbilt University
- John Flanders, Central Methodist University
- Marc Fox, Brooklyn College
- Yee-Tien (Ted) Fu, Stanford University
- Yee-Tien Fu, National Cheng-Chi University, Taiwan
- Scott Fullwiler, Wartburg College
- Julie Gallaway, University of Missouri–Rolla
- Bodhi Ganguli, Rutgers, The State University of New Jersey
- Fabio Ghironi, Boston College
- Alberto Gomez-Rivas, University of Houston–Downtown
- Fidel Gonzalez, Sam Houston State University
- Harvey Gram, Queens College, City University of New York
- Randy Grant, Linfield College
- Alan Gummerson, Florida International University
- Reza Hamzaee, Missouri Western State College
- Michael Hannan, Edinboro University
- Kenneth Harrison, Richard Stockton College
- Mark Hayford, Loyola University
- Thomas Havrilesky, Duke University
- George Heitmann, Muhlenberg College
- Ana Maria Herrera, Michigan State University
- Peter Hess, Davidson College
- Eric Hilt, Wellesley College
- John Holland, Monmouth College
- Mark Hopkins, Gettysburg College
- Takeo Hoshi, University of California, San Diego
- Ralph Husby, University of Illinois, Urbana–Champaign
- Yannis Ioannides, Tufts University
- Aaron Jackson, Bentley College
- Bonnie Johnson, California Lutheran University
- Louis Johnston, College of St. Benedict
- Barry Jones, SUNY Binghamton
- Fred Joutz, George Washington University
- Cem Karayalcin, Florida International University
- Okan Kavuncu, University of California
- Miles Kimball, University of Michigan
- Paul King, Denison University
- Michael Klein, Tufts University
- Mark Klinedinst, University of Southern Mississippi
- Shawn Knabb, Western Washington University
- Todd Knoop, Cornell College
- Paul Koch, Olivet Nazarene University
- Ng Beoy Kui, Nanyang Technical University, Singapore
- Leonard Lardaro, University of Rhode Island
- James Leady, University of Notre Dame
- Charles Leathers, University of Alabama
- Hsien-Feng Lee, National Taiwan University
- Jim Lee, Texas A&M University–Corpus Christi
- John Levendis, Loyola University New Orleans
- Frank Lichtenberg, Columbia University
- Mark Lieberman, Princeton University
- Shu Lin, Florida Atlantic University
- Maria Luengo-Prado, Northeastern University
- Mathias Lutz, University of Sussex
- Bernard Malamud, University of Nevada, Las Vegas
- Ken McCormick, University of Northern Iowa
- William McLean, Oklahoma State University
- B. Starr McMullen, Oregon State University

- Mikhail Melnik, Niagara University
- O. Mikhail, University of Central Florida
- Fabio Milani, University of California, Irvine
- Rose Milbourne, University of New South Wales
- Roger Morefield, University of Saint Thomas
- Shahriar Mostashari, Campbell University
- Eshragh Motahar, Union College
- Nick Noble, Miami University
- Ilan Noy, University of Hawaii
- John Olson, College of St. Benedict
- Brian O'Roark, Robert Morris University
- Jack Osman, San Francisco State University
- Emiliano Pagnotta, Northwestern University
- Biru Paksha Paul, SUNY Cortland
- Andrew Parkes, Mesa State College
- Allen Parkman, University of Mexico
- Jim Peach, New Mexico State University
- Gavin Peebles, National University of Singapore
- Michael Quinn, Bentley College
- Charles Revier, Colorado State University
- Jack Richards, Portland State University
- Raymond Ring, University of South Dakota
- Monica Robayo, University of North Florida
- Malcolm Robinson, Thomas Moore College
- Brian Rosario, University of California, Davis
- Kehar Sangha, Old Dominion University
- Ahmad Saranjam, Bridgewater State College
- Carol Scotese, Virginia Commonwealth University
- John Seater, North Carolina State University
- Peter Sephton, University of New Brunswick
- Ruth Shen, San Francisco State University
- Kwanho Shin, University of Kansas
- Tara Sinclair, The George Washington University
- Aaron Smallwood, University of Texas, Arlington
- David Sollars, Auburn University
- Liliana Stern, Auburn University
- Edward Stuart, Northeastern Illinois University
- Abdulhanid Sukaar, Cameron University
- Peter Summers, Texas Tech University
- Mark Thomas, University of Maryland Baltimore County
- Brian Trinque, The University of Texas at Austin
- Marie Truesdell, Marian College
- David Tufte, Southern Utah University
- Abdul Turay, Radford University
- Frederick Tyler, Fordham University
- Pinar Uysal, Boston College
- Evert Van Der Heide, Calvin College
- Kristin Van Gaasbeck, California State University, Sacramento
- Lee Van Scyoc, University of Wisconsin, Oshkosh
- Paul Wachtel, New York University Stern Business School
- Susheng Wang, Hong Kong University
- Donald Westerfield, Webster University
- Christopher Westley, Jacksonville State University
- David Wharton, Washington College
- Jonathan Willner, Oklahoma City University
- Mark Wohar, University of Nebraska, Omaha
- Steven Wood, University of California, Berkeley
- Michael Woodford, Princeton University
- Ip Wing Yu, University of Hong Kong
- Chi-Wa Yuen, Hong Kong University of Science and Technology
- Liping Zheng, Drake University
- Christian Zimmermann, University of Connecticut

They have helped me beyond the call of duty, and each has made a difference to the book.

I have many people to thank for this edition. Cameron Fletcher was a totally reliable proofreader and checker. At Pearson, I want to thank Chris DeJohn, content strategy manager; Samantha Lewis, product manager; Shweta Jain, content producer; Nayke Heine, product marketer; and Ashley Bryan, product and solutions specialist. At Integra, Denise Forlow, Gowthaman Sadhanandham and his team, for their work on the galleys.

Finally, I want to single out Steve Rigolosi, the editor for the first edition; and Michael Elia, the editor for the second and third editions. Steve forced me to clarify. Michael forced me to simplify. Together, they made all the difference to the process and to the book. I thank them deeply. I have also benefited from often-stimulating suggestions from my daughters, Serena, Giulia, and Marie: I did not, however, follow all of them. At home, I continue to thank Noelle for preserving my sanity.

Olivier Blanchard
Washington,
November 2019

About the Author

A citizen of France, **Olivier Blanchard** has spent most of his professional life in Cambridge, U.S.A. After obtaining his Ph.D. in economics at the Massachusetts Institute of Technology in 1977, he taught at Harvard University, returning to MIT in 1982. He was chair of the economics department from 1998 to 2003. In 2008, he took a leave of absence to be the Economic Counsellor and Director of the Research Department of the International Monetary Fund. Since October 2015, he has been the Fred Bergsten Senior Fellow at the Peterson Institute for International Economics, in Washington. He also remains Robert M. Solow Professor of Economics emeritus at MIT.

He has worked on a wide set of macroeconomic issues, from the role of monetary policy, to the nature of speculative bubbles, to the nature of the labor market and the determinants of unemployment, to transition in former communist countries, and to forces behind the recent global crisis. In the process, he has worked with numerous countries and international organizations. He is the author of many books and articles, including a graduate level textbook with Stanley Fischer.

He is a past editor of the Quarterly Journal of Economics, of the NBER Macroeconomics Annual, and founding editor of the AEJ Macroeconomics. He is a fellow and past council member of the Econometric Society, a past president of the American Economic Association, and a member of the American Academy of Sciences.

Introduction

The first two chapters of this book introduce you to the issues and the approach of macroeconomics.

Chapter 1

Chapter 1 takes you on a macroeconomic tour of the world. It starts with a look at the economic crisis that has shaped the world economy since the late 2000s. The tour then stops at each of the world's major economic powers: the United States, the euro area, and China.

Chapter 2

Chapter 2 takes you on a tour of the book. It defines the three central variables of macroeconomics: output, unemployment, and inflation. It then introduces the three time periods around which the book is organized: the short run, the medium run, and the long run.

This page is intentionally left blank

A Tour of the World

1

What is macroeconomics? The best way to answer is not to give you a formal definition, but rather to take you on an economic tour of the world, to describe both the main economic evolutions and the issues that keep macroeconomists and macroeconomic policymakers awake at night.

At the time of this writing (the start of 2019), policymakers are sleeping better than they did a decade ago. In 2008, the world economy entered a major macroeconomic crisis, the deepest since the Great Depression. World output growth, which typically runs at 4% to 5% a year, was negative in 2009. Since then, growth has turned positive, and the world economy has largely recovered. But the crisis, now known as the **Great Financial Crisis**, has left several scars, and some worries remain.

My goal in this chapter is to give you a sense of these events and of some of the macroeconomic issues confronting different countries today. I shall start with an overview of the crisis, and then focus on the three main economic powers of the world: the United States, the euro area, and China.

Section 1-1 looks at the crisis.

Section 1-2 looks at the United States.

Section 1-3 looks at the euro area.

Section 1-4 looks at China.

Section 1-5 concludes and looks ahead.

Read this chapter as you would read an article in a newspaper. Do not worry about the exact meaning of the words or about understanding the arguments in detail: The words will be defined, and the arguments will be developed in later chapters. Think of this chapter as background, intended to introduce you to the issues of macroeconomics. If you enjoy reading this chapter, you will probably enjoy reading this book. Indeed, once you have read it, come back to this chapter; see where you stand on the issues, and judge how much progress you have made in your study of macroeconomics.

◀ If you do not, please accept my apologies…

If you remember one basic message from this chapter, it should be: Economies, like people, get sick—high unemployment, recessions, financial crises, low growth. Macroeconomics is about why it happens, and what can be done about it. ▶▶▶

1-1 THE CRISIS

Figure 1-1 shows output growth rates for the world economy, for advanced economies, and for emerging and developing economies, separately, since 2000. As you can see, from 2000 to 2007 the world economy had a sustained expansion. Annual average world output growth was 4.5%, with advanced economies (the group of 30 or so richest countries in the world) growing at 2.7% per year, and emerging and developing economies growing at an even faster 6.6% per year.

In 2007, however, signs that the expansion might be coming to an end started to appear. US housing prices, which had doubled since 2000, started declining. Economists started to worry. Optimists believed that, although lower housing prices might lead to lower housing construction and to lower spending by consumers, the Federal Reserve Bank (the US central bank, called the Fed for short) could lower interest rates to stimulate demand and avoid a recession. Pessimists believed that the decrease in interest rates might not be enough to sustain demand and that the United States might go through a short recession.

Even the pessimists turned out not to be pessimistic enough. As housing prices continued to decline, it became clear that the problems were deeper. Many of the mortgages that had been sold during the previous expansion were of poor quality. Many of the borrowers had taken too large a loan and were increasingly unable to make the monthly payments. And, with declining housing prices, the value of their mortgage often exceeded the price of the house, giving them an incentive to default. This was not the worst of it: The banks that had issued the mortgages had often bundled and packaged them together into new securities and then sold these securities to other banks and investors. These securities had then often been repackaged into yet new securities, and so on. The result is that many banks, instead of holding the mortgages themselves, held these securities, which were so complex that their value was nearly impossible to assess.

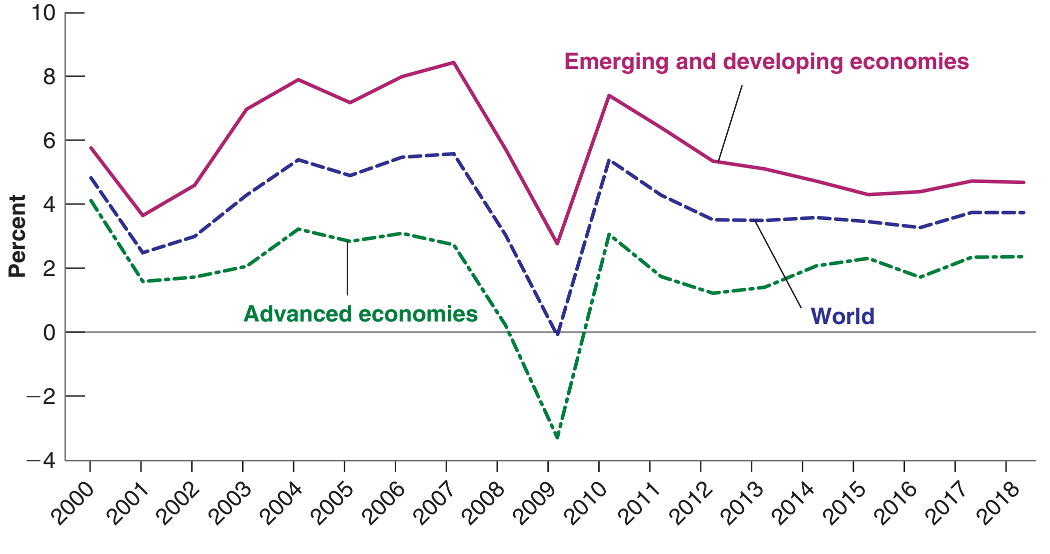
"Banks" here actually means "banks and other financial institutions." But this is too long to write and I do not want to go into these complications in Chapter 1. ▶

This complexity and opaqueness turned a housing price decline into a major financial crisis, a development that few economists had anticipated. Not knowing the quality of the assets that other banks had on their balance sheets, banks became reluctant to lend to each other for fear that the bank to which they lent might not be able to repay. Unable to borrow, and with assets of uncertain value, many banks found themselves in trouble. On September 15, 2008, a major bank, Lehman Brothers, went bankrupt.

Figure 1-1

Output Growth Rates for the World Economy, for Advanced Economies, and for Emerging and Developing Economies, 2000–2018

Source: IMF, World Economic Outlook Database, July 2018. NGDP_RPCH.A.



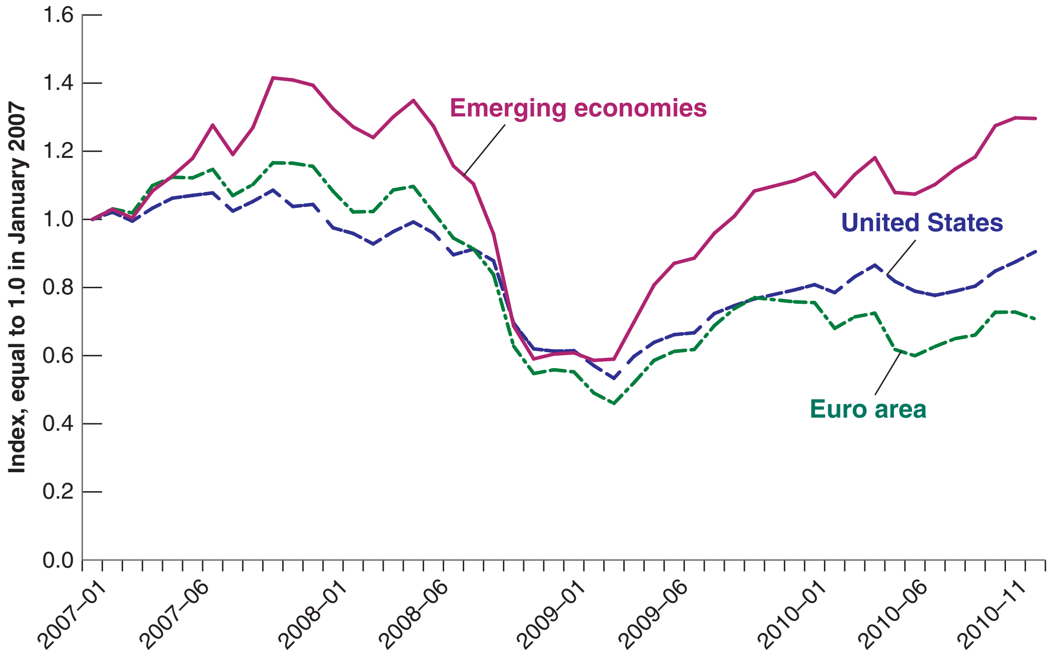


Figure 1-2

Stock Prices in the United States, the Euro Area, and Emerging Economies, 2007–2010

Source: Haver Analytics USA (S111ACD), Eurogroup (S023ACD), all emerging markets (S200ACD), all monthly averages.

The effects were dramatic. Because the links between Lehman and other banks were so opaque, many other banks appeared at risk of going bankrupt as well. For a few weeks, it looked as if the whole financial system might collapse.

I started my job as chief economist of the International Monetary Fund two weeks before the Lehman bankruptcy. I faced a steep learning curve.

This financial crisis quickly turned into a major economic crisis. Stock prices collapsed. Figure 1-2 plots stock price indexes for the United States, the euro area, and emerging economies from the beginning of 2007 to the end of 2010. The indexes are set equal to 1 in January 2007. Note that, by the end of 2008, stock prices had lost half or more of their value from their previous peak. Note also that, even though the crisis originated in the United States, European and emerging market stock prices decreased by as much as their US counterparts; I shall return to this later.

Hit by the decrease in housing prices and the collapse in stock prices, and worried that this might be the beginning of another Great Depression, people sharply cut their consumption. Worried about sales and uncertain about the future, firms sharply cut back their investment. With housing prices dropping and many vacant homes on the market, very few new homes were built. Despite strong actions by the Fed, which cut interest rates all the way down to zero, and by the US government, which cut taxes and increased spending, demand decreased, and so did output. In the third quarter of 2008, US output growth turned negative and remained so in 2009.

One might have hoped that the crisis would remain largely contained in the United States. As Figures 1-1 and 1-2 both show, this was not the case. The US crisis quickly became a world crisis. Other countries were affected through two channels.

The first channel was trade. As US consumers and firms cut spending, part of the decrease fell on imports of foreign goods. Looking at it from the viewpoint of countries exporting to the United States, their exports went down, and so, in turn, did their output.

The second channel was finance. US banks, badly needing funds in the United States, repatriated funds from other countries, creating problems for banks in those countries as well. As those banks got in trouble, lending came to a halt, leading to a decrease in spending and in output. Also, in several European countries, governments had accumulated high levels of debt and were now running large deficits. Investors began to worry about whether debt could be repaid and asked for much higher interest rates. Confronted with those high interest rates, governments drastically reduced their deficits, through a combination of lower spending and higher taxes. This led in turn to a further decrease in demand and in output. In Europe, the decline in output was so bad that this aspect of the crisis acquired its own name, the *Euro Crisis.* In short, the US recession turned into a world recession. By 2009, average growth in advanced economies was −3.4%, by far the lowest annual growth rate since the Great Depression. Growth in emerging and developing economies remained positive but was 3.5 percentage points lower than the 2000–2007 average.

Thanks to strong monetary and fiscal policies and to the gradual repair of the financial system, economies turned around and started recovering. As you can see from Figure 1-1, growth in advanced countries turned positive in 2010 and has remained positive since. In some advanced countries, most notably the United States, unemployment is now very low. The euro area, however, is still struggling; growth is positive, but unemployment remains high. Growth in emerging and developing economies has also recovered, but, as you can see from Figure 1-1, it is lower than it was before the crisis.

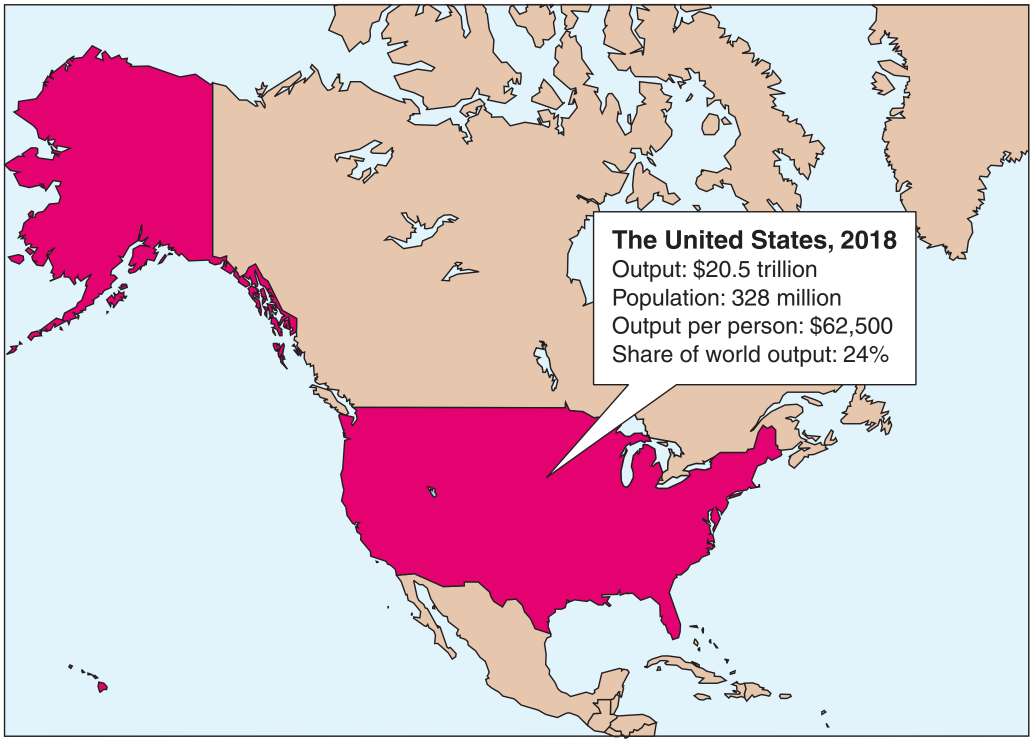
Now that I have set the stage, let me take you on a tour of the three main economic powers in the world: the United States, the euro area, and China.

1-2 THE UNITED STATES

When economists look at a country, the first two questions they ask are: How big is the country from an economic point of view? And what is its standard of living? To answer the first, they look at output—the level of production of the country as a whole. To answer the second, they look at output per person. The answers, for the United States, are given in Figure 1-3: The United States is big, with an output of $20.5 trillion in 2018, accounting for 24% of world output. And the standard of living in the United States

Figure 1-3

The United States, 2018



is high: Output per person is $62,500. It is not the country with the highest output per person in the world, but it is close to the top.

Can you guess some of the countries with a higher standard of living than the United States? *Hint*: Think of oil producers and financial centers. For answers, look for "Gross Domestic Product per capita, in current prices" in the WEO database (see the chapter appendix for the web address).

When economists want to dig deeper and look at the health of the country, they look at three basic variables:

- *Output growth*—the rate of change of output
- The *unemployment rate*—the proportion of workers in the economy who are not employed and are looking for a job
- The *inflation rate*—the rate at which the average price of goods in the economy is increasing over time

Numbers for these three variables for the US economy are given in Table 1-1. To put current numbers in historical perspective, the first column gives the average value of each of the three variables for the period 1990 up to 2007, the year before the crisis. The second column shows numbers for the acute part of the crisis, the years 2008 and 2009. The third column shows the numbers from 2010 to 2017, and the last column gives the numbers for 2018.

By looking at the numbers for 2018, you can see why economists are upbeat about the US economy at this point. Growth in 2018 is 2.9%, close to the 1990–2007 average. The unemployment rate, which increased during the crisis and its aftermath (it reached 10% during 2010), has steadily decreased and is now 3.7%, substantially lower than the 1990–2007 average. Inflation is also low, equal to its 1990–2007 average. In short, the US economy seems to be in good shape, having largely left the effects of the crisis behind.

So what are the main macroeconomic problems facing US policymakers? I shall pick two. The first concerns the short run, namely whether policymakers have the necessary tools to handle a recession. The second is how to increase productivity growth in the long run. Let's look at both issues in turn.

Do Policymakers Have the Tools to Handle the Next Recession?

The recovery from the financial crisis started in the United States in June 2009. Since then, output growth has been positive, and at the time of writing, the expansion has gone on for 115 months. If it goes on until July 2019, it will be the longest expansion on record since 1945.

If history is any guide, however, the sad reality is that expansions do not go on forever, and the United States will, sooner or later, go through another recession. It may come from several places. It may be triggered by a trade war, leading, for example, to a sharp decrease in exports. It may come from increased uncertainty, leading people to consume less and firms to invest less. It may come from another financial crisis, despite

Table 1-1 Growth, Unemployment, and Inflation in the United States, 1990–2018

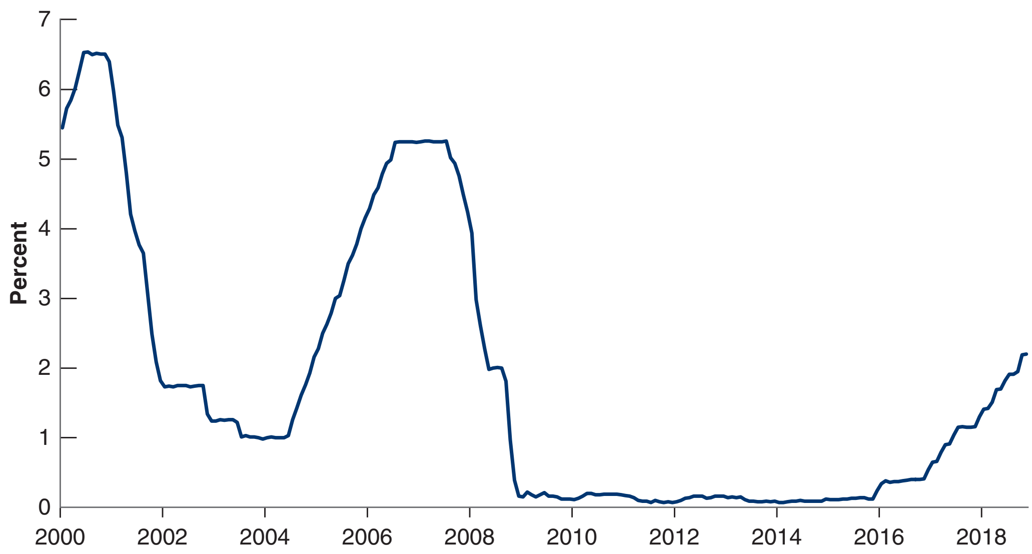
Percent	1990–2007 (average)	2008–2009 (average)	2010–2017 (average)	2018
Output growth rate	3.0	–1.3	2.2	2.9
Unemployment rate	5.4	7.5	6.8	3.7
Inflation rate	2.3	1.3	1.6	2.3

Output growth rate: annual rate of growth of output (GDP). Unemployment rate: average over the year. Inflation rate: annual rate of change of the price level (GDP deflator).

Source: IMF, World Economic Outlook, October 2018.

Figure 1-4

The US Federal Funds Rate, since 2000



the measures that have been taken since 2009 to decrease risk. Or it may come, as has happened many times in the past, from events we simply have not thought about.

Donald Rumsfeld, a past secretary of defense, had a very insightful quote. "There are known unknowns. But there are also unknown unknowns. And it is the latter category that tend to be the difficult ones."

When the recession comes, the question will be what policymakers can do to limit the decline in output. The Fed will have to play a central role. This is for two reasons. First, because part of the mandate of the Fed is indeed to fight recessions. Second, because it has the best policy instrument to do so, namely control of the interest rate. By decreasing the interest rate, the Federal Reserve can stimulate demand, increase output, and decrease unemployment. By increasing the interest rate, it can slow down demand and increase unemployment.

The problem that the Fed faces at this point, however, is shown in Figure 1-4, which shows the path of the policy interest rate (called the *Federal Funds Rate*) since 2000. Note how much the Fed decreased the interest rate when the crisis hit, from 5.3% in July 2008 to close to 0% in December 2008. Note then that the rate remained close to 0% until the end of 2015, and how it has increased a little since then and now stands at 2.4%.

Why did the Fed stop at zero? It would have liked to decrease the interest rate further, but it could not because the interest rate cannot be negative. If it were, then nobody would hold bonds; everybody would want to hold cash instead—because cash pays a zero interest rate. This constraint is known as the *zero lower bound*, and this is the bound the Fed ran into in December 2008.

Because keeping cash in large sums is inconvenient and dangerous, people might be willing to hold some bonds even if those pay a small negative interest rate. But there is a clear limit to how negative the interest rate can go before people switch to cash.

Now that the interest rate has increased, why is the zero lower bound still an issue? Because the interest rate remains very low by historical standards. And this implies that there is little room for the Fed to decrease it. If another recession were to happen, the Fed could decrease the policy rate by only about 2%, not enough to have a large effect on demand.

Are there other tools that the Fed could use? Can fiscal policy help? The answer to both questions, as we shall see later in the book, is yes. But whether these other tools will be enough is far from certain. This is why many economists are worried that it might be difficult to limit the depth of the next recession.

By the time you read this book, a recession may have started. If so, you will know what the correct answer was.

How Worrisome Is Low Productivity Growth?

In the short run, what happens to the economy depends, as we just discussed, on movements in demand and the decisions of the central bank. In the longer run, however, growth is determined by other factors, the main one being productivity growth: Without

Table 1-2 Labor Productivity Growth, by Decade, 1990–2018			
Percent change; year on year (average)	**1990s**	**2000s**	**2010–2018**
Private nonfarm business sector	**2.2**	**2.8**	**0.9**
Manufacturing	**4.1**	**3.6**	**0.4**

Source: FRED database. PRS85006092, MPU490063

productivity growth, there just cannot be a sustained increase in income per person. And, here, the news is worrisome. Table 1-2 shows average US productivity growth by decade since 1990 for the private nonfarm business sector and for the manufacturing sector. As you can see, productivity growth in the 2010s has been, so far, much lower than it was in the previous two decades.

How worrisome is this? Productivity growth varies a lot from year to year, and some economists believe that it may just be a few bad years and not much to worry about. Others believe that measurement issues make it difficult to measure output and that productivity growth may be underestimated. For example, how do you measure the productivity of a new smartphone relative to an older model? For the same price as an older model, it does many things that the older model could not do. Put another way, it is much more productive, and we may not be very good at measuring the improvement in productivity. Yet others believe that the United States has truly entered a period of lower productivity growth, that the major gains from the current innovations in information technology (IT) may already have been obtained, and that progress is likely to be less rapid, at least for some time.

One particular reason to worry is that this slowdown in productivity growth is happening in the context of growing inequality. When productivity growth is high, most are likely to benefit, even if inequality increases. The poor may benefit less than the rich, but they still see their standard of living increase. This is not the case today in the United States. Since 2000, the real earnings of workers with a high school education or less have actually decreased. If policymakers want to invert this trend, they need to either raise productivity growth or limit the rise of inequality, or both. These are two major challenges facing US policymakers today.

Increasing inequality is a problem affecting not just the United States but many advanced economies. It has serious political implications.

1-3 THE EURO AREA

In 1957, six European countries decided to form a common European market—an economic zone where people, goods, and services could move freely. Over time, 22 more countries joined, bringing the total to 28. This group is now known as the **European Union (EU)** and its scope extends beyond just economic issues. In 2016, the United Kingdom held a referendum in which the government was given the mandate to exit the Union. At this juncture, negotiations are still going on, but, if and when the United Kingdom leaves, this will leave 27 members.

Until a few years ago, the official name was the *European Community*, or EC. You may still encounter that name. EC now stands for *European Commission,* the executive arm of the European Union.

In 1999, the EU decided to go a step further and started the process of replacing national currencies with one common currency, called the *euro*. Only 11 countries participated at the start; since then, 8 more have joined. Nineteen countries now belong to this **common currency area**, known as the **euro area**.

The area also goes by the names of “Eurozone” or “Euroland.” The first sounds too technocratic, and the second reminds one of Disneyland. I shall avoid them.

As you can see from the numbers in Figure 1-5, the euro area is a strong economic power. At the current exchange rate between the euro and the dollar, its output is equal to two-thirds of US output. (The EU as a whole has an output equal to 90% of that of the United States.)

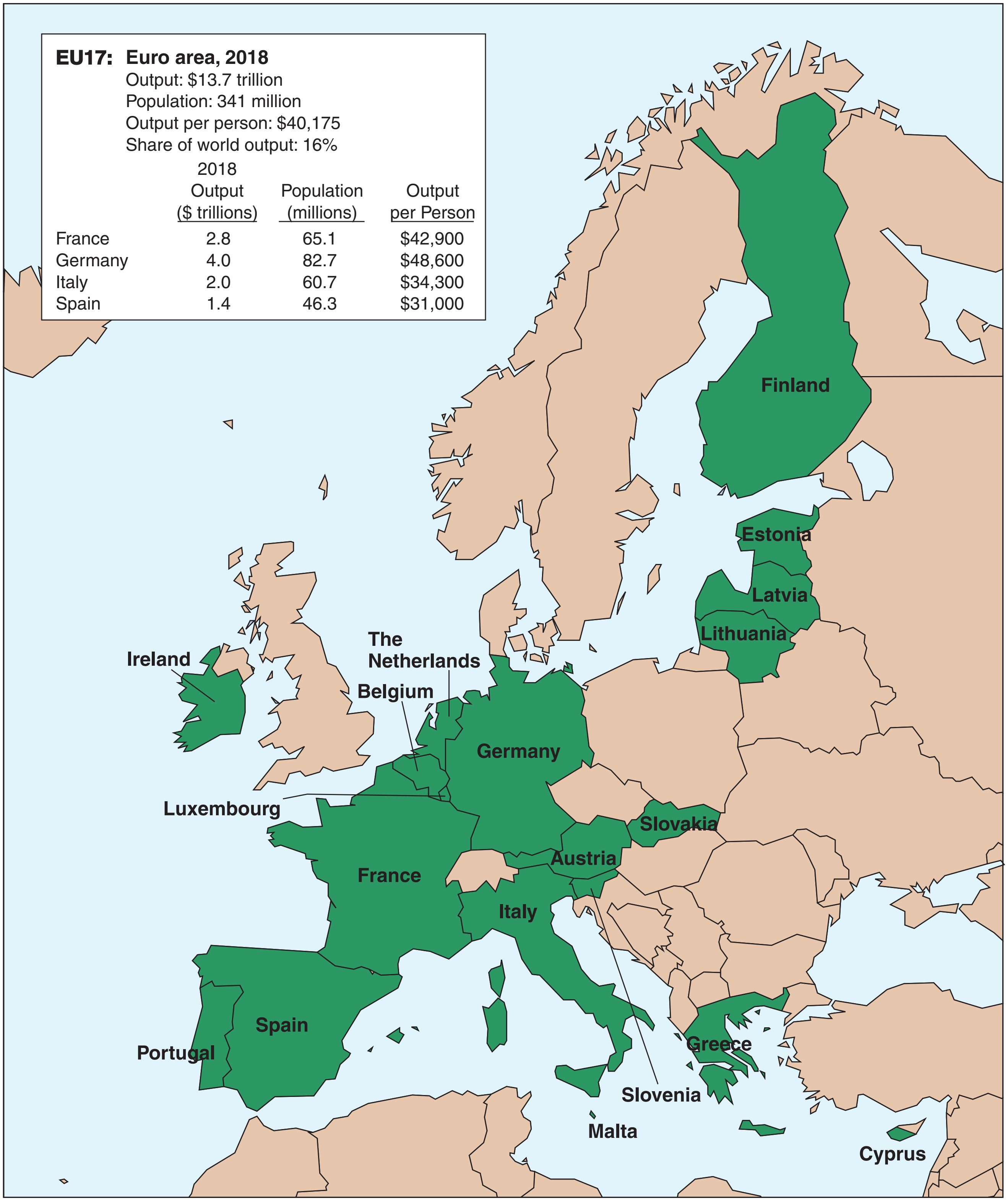


Figure 1-5

The Euro Area, 2018

Table 1-3	Growth, Unemployment, and Inflation in the Euro Area, 1990–2018

Percent	1990–2007 (average)	2008–2009 (average)	2010–2017 (average)	2018
Output growth rate	2.1	–2.0	1.3	2.0
Unemployment rate	9.4	8.6	10.6	8.3
Inflation rate	2.1	1.5	1.0	1.5

Output growth rate: annual rate of growth of output (GDP). Unemployment rate: average over the year. Inflation rate: annual rate of change of the price level (GDP deflator).

Source: IMF, World Economic Outlook, October 2018.

Table 1-3 gives the numbers for output growth, the unemployment rate, and the inflation rate for 1990–2007, 2008–2009, 2010–2017, and 2018. Just as in the United States, the acute phase of the crisis, 2008–2009, was characterized by negative growth. Whereas the United States recovered, growth in the euro area remained anemic. Indeed, while this is not shown in the table, growth was negative in both 2012 and 2013. Growth has now increased, reaching 2% in 2018, but the unemployment rate remains high, at 8.3%. Inflation remains too low, below the 2% target of the European Central Bank (ECB).

The euro area faces two main issues today. The first is how to reduce unemployment. Second is whether and how it can function efficiently as a common currency area. Let's look at the two issues in turn.

Can European Unemployment Be Reduced?

The high average unemployment rate for the euro area, 8.3% in 2018, hides large variations across the euro countries. At one end, Greece and Spain have unemployment rates of 20% and 15%, respectively. At the other, Germany's unemployment rate is close to 3%. In the middle are countries like France and Italy, with unemployment rates of 9% and 11%, respectively. Thus, how to reduce unemployment must be tailored to the specifics of each country.

To show the complexity of the issues, it is useful to look at a country with high unemployment, say Spain. Figure 1-6 shows the striking evolution of the Spanish

Figure 1-6

Unemployment in Spain since 1990

(*Source:* International Monetary Fund, World Economic Outlook, October 2018).

unemployment rate since 1990. After a long boom starting in the mid-1990s, the unemployment rate decreased from a high of nearly 25% in 1994 to 8% by 2007. But, with the crisis, unemployment exploded again, exceeding 25% in 2013. It has declined since then, but still stands at 15%.

The figure suggests two conclusions:

- Part of the high unemployment rate today is probably still a result of the crisis and the sudden collapse in demand we discussed in the first section. A housing boom that turned into a housing bust, plus a sudden increase in interest rates, triggered the increase in unemployment from 2008 on. One can hope that, eventually, demand will continue to increase, and unemployment will decrease further.
- Even at the peak of the boom, the unemployment rate in Spain never went below 8%, nearly three times the unemployment rate in Germany today. This suggests that more is at work than the crisis and the fall in demand. The fact that, for most of the last 20 years, unemployment has exceeded 10%, points to problems in the labor market. The challenge is then to identify exactly what these problems are.

Some economists believe the main problem is that European states protect workers too much. To prevent workers from losing their jobs, they make it expensive for firms to lay off workers. One of the unintended results of this policy is to deter firms from hiring workers in the first place, thus increasing unemployment. Also, to protect workers who become unemployed, European governments provide generous unemployment insurance. But, by doing so, they decrease the incentives for the unemployed to take jobs rapidly; this also may increase unemployment. The solution, these economists argue, is to be less protective, to eliminate these *labor market rigidities*, and to adopt US-style labor market institutions. This is what the United Kingdom has largely done, and its unemployment rate is low.

Others, and this includes me, are more skeptical. They point to the fact that unemployment is not high everywhere in Europe. Yet most European countries provide protection and generous social insurance to workers. This suggests that the problem may lie not so much with the degree of protection but with the way it is implemented. The challenge, these economists argue, is to understand what the low-unemployment European countries are doing right, and whether what they do right can be exported to the other European countries.

Resolving these questions is one of the major tasks facing European macroeconomists and policymakers.

What Has the Euro Done for Its Members?

Supporters of the euro point to its enormous symbolic importance. In light of the many past wars among European countries, what better proof of the permanent end to conflict than the adoption of a common currency? They also point to the economic advantages of having a common currency: no more changes in exchange rates for European firms to worry about; no more need to change currencies when crossing borders. Together with the removal of other obstacles to trade among European countries, the euro contributes, they argue, to the creation of a large economic power in the world. There is little question that the move to the euro was indeed one of the main economic events of the start of the 21st century.

Others worry, however, that the symbolism of the euro has come with substantial economic costs. Even before the crisis, they pointed out that a common currency means a common monetary policy, which means the same interest rate across the euro countries. What if, they argue, one country plunges into recession while another is in the middle of an economic boom? The first country needs lower interest rates to increase spending and output; the second country needs higher interest rates to slow down its economy. If interest rates must be the same in both countries, what will happen? Isn't there the risk that one country will remain in recession for a long time or that the other will not be able to slow down its booming economy? A common currency also means the loss of the exchange rate as an instrument of adjustment within the euro area. What if, they argue, a country has a large trade deficit and needs to become more competitive? If it cannot adjust its exchange rate, it must adjust by decreasing prices relative to its competitors. This is likely to be a painful and long process.

Until the euro crisis, the debate had remained somewhat abstract. It no longer is. As a result of the crisis, several euro members, from Ireland and Portugal to Greece, have gone through deep recessions. If they had their own currency, they could have depreciated their currency vis-à-vis other euro members to increase the demand for their exports. Because they shared a currency with their neighbors, this was not possible. Thus, some economists conclude, some countries should drop out of the euro and recover control of their monetary policy and their exchange rate. Others argue that such an exit would be both unwise because it would give up the other advantages of being in the euro and extremely disruptive, leading to even deeper problems for the country that exited. This issue is likely to remain a hot one for some time to come.

At the time of this writing, some Italian politicians argue that Italy, which suffers from low growth, would be better off outside the euro and advocate euro exit.

1-4 CHINA

China is in the news every day. It is perceived as one of the major economic powers in the world. Is the attention justified? A first look at the numbers in Figure 1-7 suggests it may not. True, the population of China is enormous, more than four times that of the United States. But its output, expressed in dollars by multiplying the number in yuan (the Chinese currency) by the dollar–yuan exchange rate, is still only \$13.5 trillion, about 60% of that of the United States. Output per person is about \$9,700, only roughly 15% of output per person in the United States.

So why is so much attention paid to China? There are two main reasons:

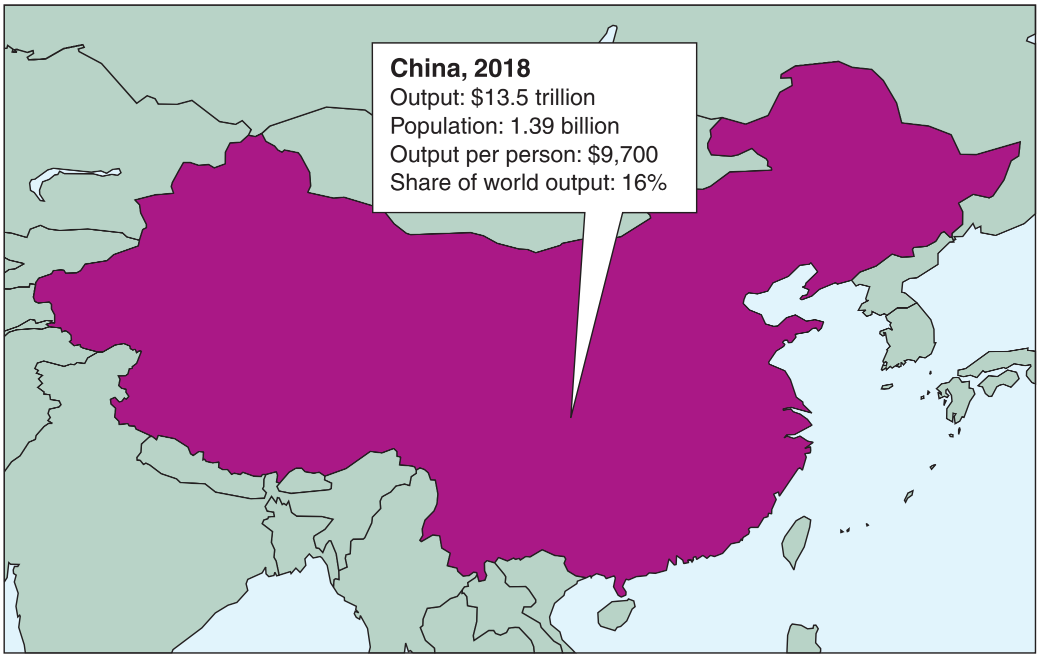


Figure 1-7

China, 2018

Source: IMF, World Economic Outlook, October 2018.

To understand the first, we need to go back to the number for output per person. When comparing output per person in a rich country like the United States and a relatively poor country like China, one must be careful. The reason is that many goods are cheaper in poor countries. For example, the average price of a restaurant meal in New York City is about $40; the average price of a restaurant meal in Beijing is about 50 yuan, or, at the current exchange rate, about $7.50. Put another way, the same income (expressed in dollars) buys you much more in Beijing than in New York City. If we want to compare standards of living, we must correct for these differences; measures that do so are called PPP (for *purchasing power parity*) measures. Using such a measure, China's output is estimated to be $25.3 trillion, thus higher than that of the United States. And output per person in China is estimated to be about $18,100, a bit less than one-third of the output per person in the United States. This gives a more accurate picture of the standard of living in China. It is obviously still much lower than that of the United States or other rich countries. But it is higher than suggested by the numbers in Figure 1-7.

The issue is less important when comparing two rich countries. Thus, this was not a major issue when comparing standards of living in the United States and the euro area. ▶

Second, and more importantly, China has been growing very rapidly for more than three decades. This is shown in Table 1-4, which, like the previous tables for the United States and the euro area, gives output growth, unemployment, and inflation for 1990–2007, 2008–2009, 2010–2017, and 2018.

The first line of the table tells the basic story. From 1990 (indeed, from 1980, if we were to extend the table back by another 10 years) to the late 2000s, China grew at close to 10% a year. This represents a doubling of output every 7 years. Compare this to the numbers for the United States and for Europe, and you understand why the weight of the emerging economies in the world economy, China being the main one, is increasing so rapidly.

A useful rule, called the rule of 70: The number of years it takes for a variable to double is equal to 70 divided by the growth rate of the variable. ▶

There are two other interesting aspects to Table 1-4.

The first is how difficult it is to see the effects of the crisis in the data. Growth barely decreased during 2008 and 2009, and unemployment barely increased. The reason is not that China is closed to the rest of the world. Chinese exports slowed during the crisis. But the adverse effect on demand was nearly fully offset by a major fiscal expansion by the Chinese government, with, in particular, a major increase in public investment. The result was sustained growth of demand and, in turn, of output.

The second is the decline in growth rates from 10% before the crisis to less than 8% after the crisis, and to 6.6% for 2018. This raises questions both about how China maintained such a high growth rate for so long, and whether it is now entering a period of lower growth.

A preliminary question is whether the numbers are for real. Could it be that Chinese growth was and is still overstated? After all, China is still officially a communist country, and government officials may have incentives to overstate the economic performance of their sector or their province. Economists who have looked at this carefully conclude that this is probably not the case. The statistics are not as reliable as they are in richer countries, but there is no major bias. Output growth has indeed been very high in China.

Table 1-4 Growth, Unemployment, and Inflation in China, 1990–2018

Percent	1990–2007 (average)	2008–2009 (average)	2010–2017 (average)	2018
Output growth rate	10.2	9.4	7.9	6.6
Unemployment rate	3.3	4.3	4.1	4.0
Inflation rate	5.9	3.7	2.9	2.2

Output growth rate: annual rate of growth of output (GDP). Unemployment rate: average over the year. Inflation rate: annual rate of change of the price level (GDP deflator).

Source: IMF, World Economic Outlook, October 2018.

So where has growth come from? It has come from two sources: The first was high accumulation of capital. The investment rate (the ratio of investment to output) in China is 46%, a very high number. For comparison, the investment rate in the United States is only 21%. More capital means higher productivity and higher output. The second is rapid technological progress. One of the strategies followed by the Chinese government has been to encourage foreign firms to relocate and produce in China. As foreign firms are typically much more productive than Chinese firms, this has increased productivity and output. Another aspect of the strategy has been to encourage joint ventures between foreign and Chinese firms. By making Chinese firms work with and learn from foreign firms, the productivity of the Chinese firms has increased dramatically.

This transfer of technology is the subject of strong criticism by the United States government, which argues that part of it has been done illegally, and is a source of trade tensions between the two countries.

When described in this way, achieving high productivity and high output growth appears easy and a recipe that every poor country could and should follow. In fact, things are less obvious. China is one of several countries that made the transition from central planning to a market economy. Most of the other countries, from those in Central Europe to Russia and the other former Soviet republics, experienced a large decrease in output at the time of transition. Most still have growth rates far below that of China. In many countries, widespread corruption and poor property rights make firms unwilling to invest. So why has China fared so much better? Some economists believe that this is the result of a slower transition: The first Chinese reforms took place in agriculture as early as 1980, and even today, many firms remain owned by the state. Others argue that the fact that the communist party has remained in control has actually helped the economic transition; tight political control has allowed for better protection of property rights, at least for new firms, giving them incentives to invest. Getting the answers to these questions, and thus learning what other poor countries can take from the Chinese experience, can clearly make a huge difference, not only for China but for the rest of the world.

Tight political control has allowed corruption to develop, and corruption can also threaten investment. China is now in the midst of a strong anti-corruption campaign.

At the same time, the recent growth slowdown raises a new set of questions: Where does the slowdown come from? Should the Chinese government try to maintain high growth or accept the lower growth rate? Most economists and, indeed, the Chinese authorities themselves believe that lower growth is now desirable, that the Chinese people will be better served if the investment rate decreases, allowing more output to go to consumption. Achieving the transition from investment to consumption is the major challenge facing the Chinese authorities today.

1-5 LOOKING AHEAD

This concludes our whirlwind world tour. There are many other regions of the world and many other macroeconomic issues we could have looked at:

- India, another poor and large country, with a population of 1,330 million, which, like China, is now growing very fast and becoming a world economic power.
- Japan, whose growth performance for the 40 years following World War II was so impressive that it was referred to as an economic miracle, but has done very poorly in the last two decades. Since a stock market crash in the early 1990s, Japan has been in a prolonged slump, with average output growth only around 1% per year.
- Latin America, which went from high inflation to low inflation in the 1990s, and then sustained strong growth. Recently, however, its growth has slowed, as a result, in part, of a decline in the price of commodities.
- Central and Eastern Europe, which shifted from central planning to a market system in the early 1990s. In most countries, the shift was characterized by a sharp decline in output at the start of transition. Since then, however, most countries have achieved high growth rates, and are catching up with Western Europe.

- Sub-Saharan Africa, which has suffered decades of economic stagnation, but where, contrary to common perceptions, growth has been high since 2000, averaging 5% per year and reflecting growth in most of the countries of the continent.

There is a limit to how much you can absorb in this first chapter. Think about the issues to which you have been exposed:

- The big issues triggered by the crisis: What caused the crisis? Why was it transmitted from the United States to the rest of the world? In retrospect, what could and should have been done to prevent it? Were the monetary and fiscal responses appropriate? Why has the recovery been so slow in Europe? How was China able to maintain high growth during the crisis?
- How can monetary and fiscal policies be used to fight recessions? What are the pros and cons of joining a common currency area such as the euro area? What measures could be taken in Europe to reduce persistently high unemployment?
- Why do growth rates differ so much across countries, even over long periods of time? Can advanced economies achieve sustained growth without increasing inequality? Can poor countries emulate China and grow at the same rate? Should China slow down?

The purpose of this book is to give you a way of thinking about these questions. As we develop the tools you need, I shall show you how to use them by returning to these questions and showing you the answers that the tools suggest.

KEY TERMS

Great Financial Crisis, 3
European Union (EU), 9
euro area, 9
common currency area, 9

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. Output growth was negative in advanced as well as emerging and developing countries in 2009.
b. World output growth recovered to its pre-recession level after 2009.
c. Stock prices around the world fell between 2007 and 2010 and then recovered to their pre-recession level.
d. The rate of unemployment in the United Kingdom is much lower than in much of the rest of Europe.
e. China's seemingly high growth rate is a myth; it is a product solely of misleading official statistics.
f. The high rate of unemployment in Europe started when a group of major European countries adopted a common currency.
g. The Federal Reserve lowers interest rates when it wants to avoid a recession and raises interest rates when it wants to slow the rate of growth in the economy.
h. Output per person is different in the euro area, the United States, and China.
i. Interest rates in the United States were at or near zero from 2009 to 2018.

2. Macroeconomic policy in Europe

Beware of simplistic answers to complicated macroeconomic questions. Consider each of the following statements and comment on whether there is another side to the story.

a. There is a simple solution to the problem of high European unemployment: Reduce labor market rigidities.
b. What can be wrong about joining forces and adopting a common currency? Adoption of the euro is obviously good for Europe.

DIG DEEPER

3. Chinese economic growth is the outstanding feature of the world economic scene over the past two decades.

a. In 2018, US output was $20.5 trillion, and Chinese output in 2017 was $13.5 trillion. Suppose that from 2017 the output of China grows at an annual rate of 7.9%, whereas the output of the United States grows from 2018 at an annual rate of 2.2%. These are the values in each country for the most recent periods in Tables 1-1 and 1-4, respectively. Using these assumptions and a spreadsheet, calculate and plot US and Chinese output from 2017 or 2018 over the next 100 years. How many years will it take for China to have a total level of output equal to that of the United States?

b. When China catches up with the United States in total output, will residents of China have the same standard of living as US residents? Explain.

c. Another term for *standard of living* is *output per person*. How has China raised its output per person in the last two decades? Are these methods applicable to the United States?

d. Do you think China's experience in raising its standard of living (output per person) provides a model for developing countries to follow? Explain.

4. The rate of growth of output per person was identified as a major issue facing the United States as of the writing of this chapter. Go to the 2018 Economic Report of the President (www.whitehouse.gov/wp-content/uploads/2018/02/ERP_2018_Final-FINAL.pdf) and find a table titled "Productivity and Related Data" (Table B-16). It can be downloaded as an Excel file.

a. Find the column with numbers that describe the level of output per hour worked of all persons in the nonfarm business sector. This value is presented as an index number equal to 100 in 2009. Calculate the percentage increase in output per hour worked from 2009 to 2010. What does that value mean?

b. Now use the spreadsheet to calculate the average percent increase in output per hour worked for the decades 1970–1979, 1980–1989, 1990–1999, 2000–2009, and 2010–2017. How does productivity growth in the most recent decade compare to the other decades?

c. If a more recent *Economic Report of the President* is available, update your estimate of the average growth rate of output per hour worked to include years past 2017. Is there any evidence of an increase in productivity growth?

EXPLORE FURTHER

5. US recessions

This section looks at US recessions over the past 60 years. To work out this problem, first obtain quarterly data on US output growth for the period 1960 to the most recent data from www.bea.gov. Table 1.1.1 presents the percent change in real gross domestic product (GDP). The data can be downloaded to a spreadsheet. Plot the quarterly GDP growth rates from 1960:1 to the latest observations. Which, if any, quarters have negative growth? Using the definition of a recession as two or more consecutive quarters of negative growth, answer the following questions.

a. How many recessions has the US economy undergone since 1960, quarter 2?

b. How many quarters has each recession lasted?

c. In terms of length and magnitude, which two recessions have been the most severe?

6. From Problem 5, write down the quarters in which the recessions started. Find the monthly series in the Federal Reserve Bank of St. Louis (FRED) database for the seasonally adjusted unemployment rate in the United States entitled ***civilian unemployment rate****. Retrieve the monthly data series on the unemployment rate for the period 1969 to the end of the data. Make sure all data series are seasonally adjusted.*

a. Look at each recession since 1969. What was the unemployment rate in the first month of the first quarter of negative growth? What was the unemployment rate in the last month of the last quarter of negative growth? By how much did the unemployment rate change?

b. Which recession had the largest increase in the rate of unemployment? Begin with the month before the quarter in which output first falls and measure to the highest level of the unemployment rate before the next recession.

7. European unemployment

The FRED database contains updates of the unemployment rate in Spain (Figure 1-6) as well as unemployment rates for the European Union as a whole and for individual countries for the seasonally adjusted unemployment rate. Retrieve the monthly data series for the unemployment rates below starting in the year 2000 to the latest data:

Harmonized Unemployment Rate: Total: All Persons for the European Union

Harmonized Unemployment Rate: Total: All Persons for Spain

Harmonized Unemployment Rate: Total: All Persons for the United Kingdom

a. Is the most recent unemployment rate in the United Kingdom much lower than that in the European Union or in Spain?

b. How does the change in Spanish unemployment from its peak near April 2013 compare to the change in the unemployment rate for the European Union as a whole from its peak in May 2013?

APPENDIX 1: Where to Find the Numbers

Suppose you want to find the numbers for inflation in Germany over the past five years. Fifty years ago, the answer would have been to learn German, find a library with German publications, find the page where inflation numbers were given, write them down, and plot them by hand on a clean sheet of paper. Today, improvements in the collection of data, the development of computers and electronic databases, and access to the internet make the task much easier. This appendix will help you find the numbers you are looking for, be it inflation in Malaysia last year, or consumption in the United States in 1959, or unemployment in Ireland in the 1980s.

To Find Data

Four good sources of free and easily downloadable data are:

- FRED: Federal Reserve Economic Database. A continuously updated database maintained by the Federal Reserve Bank of Saint Louis, giving many macroeconomic and financial data, mostly for the United States, but also for other countries. https://fred.stlouisfed.org/
- WEO: World Economic Outlook Database. A database maintained by the **International Monetary Fund (IMF)**, an international organization including most countries (at this point, 189) in the world. Updated twice a year, giving basic macroeconomic data for all member countries. The October 2018 data can be found at www.imf.org/external/pubs/ft/weo/2018/02/weodata/index.aspx
- OECD.stat: A database maintained by the **Organization for Economic Cooperation and Development (OECD)**, an international organization that includes most of the rich countries in the world. Together, these countries account for about 70% of the world's output. One strength of the OECD data is that, for many variables, the OECD tries to make the variables comparable across member countries (or tells you when they are not comparable). https://stats.oecd.org/
- AMECO: An *annual macroeconomic* database, maintained by the Commission of the European Union, that gives detailed macroeconomic data for all European Union members. http://ec.europa.eu/economy_finance/ameco/user/serie/SelectSerie.cfm

For (sometimes much) longer historical time series, the following are good sources:

- For the United States, *Historical Statistics of the United States, Colonial Times to 1970*, Parts 1 and 2, published by the US Department of Commerce, Bureau of the Census (www.census.gov/prod/www/statistical_abstract.html).
- For long-term historical statistics for several countries, Angus Maddison's *Monitoring the World Economy,* 1820–1992, Development Centre Studies, OECD, Paris, 1995. This study gives data going back to 1820 for 56 countries. Two even longer and broader sources are *The World Economy: A Millennial Perspective*, Development Studies, OECD, 2001, and *The World Economy: Historical Statistics*, Development Studies, OECD 2004, both also by Angus Maddison.

To Keep Current

By the time you read this chapter, many new events will have occurred. If you want to keep informed of current economic events, you will find the following three sources very useful:

- The WEO, which describes major economic events in the world and in specific member countries.
- The OECD Economic Outlook, published by the OECD twice a year, which describes major economic events in OECD countries. www.oecd.org/eco/outlook/economic-outlook/
- *The Economist*, published each week. *The Economist* is a highly informative, often opinionated magazine on economic and political events around the world. The last four pages give the most recent numbers on output, unemployment, inflation, exchange rates, interest rates, and stock prices for a large number of countries. Unfortunately, most of the articles and data are behind a paywall.

Key Terms

Organization for Economic Cooperation and Development (OECD), 18
International Monetary Fund (IMF), 18

APPENDIX 2: What Do Macroeconomists Do?

There might be many reasons why you are taking a course in macroeconomics. Some of you simply want to have a better understanding of what is happening around you. Some of you are taking the class to fulfill a major, get an undergraduate degree in economics, and go to the job market. Others need the course to get a more advanced degree, be it a master's or a doctorate in economics.

For those who want to specialize in macroeconomics, you may want to know what jobs you can expect to get, what you will be doing in those jobs, and how much you can expect to earn. In short, with an undergraduate degree, you can expect to work in the private sector, be it in large firms or in financial institutions, helping them assess the economic situation. Jobs in central banks such as the Fed, or in international organizations such as the IMF or the World Bank, are likely to require you to have a PhD. Jobs in academia have a similar requirement. You can get much more information by going to the American Economic Association website page devoted to careers in economics. www.aeaweb.org/resources/students/careers

A Tour of the Book

2

The words *output, unemployment*, and *inflation* appear daily in newspapers and on the evening news. So when I used these words in Chapter 1, you knew roughly what we were talking about. It is now time to define these words more precisely, and this is what I do in the first three sections of this chapter.

Section 2-1 looks at output.

Section 2-2 looks at the unemployment rate.

Section 2-3 looks at the inflation rate.

Section 2-4 introduces two important relations between these three variables: Okun's law and the Phillips curve.

Section 2-5 then introduces the three central concepts around which the book is organized:

- **The *short run*: What happens to the economy from year to year**
- **The *medium run*: What happens to the economy over a decade or so**
- **The *long run*: What happens to the economy over a half century or longer**

Building on these three concepts, Section 2-6 gives you a road map to the rest of the book.

If you remember one basic message from this chapter, it should be: The three central macroeconomic variables are output, unemployment, and inflation. ▶▶▶

2-1 AGGREGATE OUTPUT

Economists studying economic activity in the 19th century or even during the Great Depression had no measure of aggregate activity (*aggregate* is the word macroeconomists use for *total*) on which to rely. They had to put together bits and pieces of information, such as the shipments of iron ore, or sales at some department stores, to try to infer what was happening to the economy as a whole.

Two economists, Simon Kuznets, from Harvard University, and Richard Stone, from Cambridge University, received the Nobel Prize for their contributions to the development of the national income and product accounts—a gigantic intellectual and empirical achievement.

It was not until the end of World War II that **national income and product accounts** (or national income accounts, for short) were put together. Measures of aggregate output have been published on a regular basis in the United States since October 1947. (Measures of aggregate output for earlier times have been constructed retrospectively.)

Like any accounting system, the national income accounts first define concepts and then construct measures corresponding to these concepts. You need only to look at statistics from countries that have not yet developed such accounts to realize that precision and consistency in such accounts are crucial. Without precision and consistency, numbers that should add up do not; trying to understand what is going on feels like trying to balance someone else's checkbook. I shall not burden you with the details of national income accounting here. But because you will occasionally need to know the definition of a variable and how variables relate to each other, Appendix 1 at the end of the book gives you the basic accounting framework used in the United States (and, with minor variations, in most other countries) today. You will find it useful whenever you want to look at economic data on your own.

You may come across another term, **gross national product**, or **GNP**. There is a subtle difference between "domestic" and "national," and thus between GDP and GNP. We examine the distinction in Chapter 18 and in Appendix 1 at the end of the book. For now, ignore it.

GDP: Production and Income

The measure of **aggregate output** in the national income accounts is called the **gross domestic product, or GDP**. To understand how GDP is constructed, it is best to work with a simple example. Consider an economy composed of just two firms:

In reality, not only workers and machines are required for steel production, but so are iron ore, electricity, and so on. I ignore these to keep the example simple.

- Firm 1 produces steel, employing workers and using machines to produce the steel. It sells the steel for $100 to Firm 2, which produces cars. Firm 1 pays its workers $80, leaving $20 in profit to the firm.
- Firm 2 buys the steel and uses it, together with workers and machines, to produce cars. Revenues from car sales are $200. Of the $200, $100 goes to pay for steel and $70 goes to workers in the firm, leaving $30 in profit to the firm.

We can summarize this information in a table:

Steel Company (Firm 1)			**Car Company (Firm 2)**		
Revenues from sales		$100	Revenues from sales		$200
Expenses		$80	Expenses		$170
Wages	$80		Wages	$70	
			Steel purchases	$100	
Profit		$20	Profit		$30

An intermediate good is a good used in the production of another good. Some goods can be both final goods and intermediate goods. Potatoes sold directly to consumers are final goods. Potatoes used to produce potato chips are intermediate goods. What are some other examples?

How would you define aggregate output in this economy? As the sum of the values of all goods produced in the economy—the sum of $100 from the production of steel and $200 from the production of cars, so $300? Or as just the value of cars, which is equal to $200?

Some thought suggests that the right answer must be $200. Why? Because steel is an **intermediate good**: It is used in the production of cars. Once we count the production of cars, we do not want to also count the production of the goods that went into the production of these cars.

This leads to the first definition of GDP:

1. *GDP Is the Value of the Final Goods and Services Produced in the Economy during a Given Period.*

The important word here is *final*. We want to count only the production of **final goods**, not intermediate goods. Using our example, we can make this point in another way. Suppose the two firms merged, so that the sale of steel took place in the new firm and was no longer recorded. The accounts of the new firm would be given by the following table:

Steel and Car Company	
Revenues from sales	$200
Expenses (wages)	$150
Profit	$50

All we would see would be one firm selling cars for $200, paying workers $80 + $70 = $150 and making $20 + $30 = $50 in profits. The $200 measure would remain unchanged—as it should. We do not want our measure of aggregate output to depend on whether firms decide to merge or not.

This first definition gives us one way to construct GDP: by recording and adding up the production of all final goods—and this is indeed roughly the way actual GDP numbers are put together. But it also suggests a second way of thinking about and constructing GDP.

2. *GDP Is the Sum of Value Added in the Economy during a Given Period.*

The **value added** by a firm is defined as the value of its production minus the value of the intermediate goods used in production.

In our two-firms example, the steel company does not use intermediate goods. Its value added is simply equal to the value of the steel it produces, $100. The car company, however, uses steel as an intermediate good. Thus, the value added by the car company is equal to the value of the cars it produces minus the value of the steel it uses in production, $200 − $100 = $100. Total value added in the economy, or GDP, equals $100 + $100 = $200. (Note that aggregate value added would remain the same if the steel and car firms merged and became a single firm. In this case, we would not observe intermediate goods at all—because steel would be produced and then used to produce cars within the single firm—and the value added in the single firm would simply be equal to the value of cars, $200.)

This definition gives us a second way of thinking about GDP. Put together, the two definitions imply that the value of final goods and services—the first definition of GDP—can also be thought of as the sum of the value added by all the firms in the economy—the second definition of GDP.

So far, we have looked at GDP from the *production side*. The other way of looking at GDP is from the *income side*. Go back to our example and think about the revenues left to a firm after it has paid for its intermediate goods: Some of the revenues go to pay workers—this component is called *labor income*. The rest goes to the firm—that component is called *capital income* or *profit income* (the reason it is called capital income is that you can think of it as remuneration for the owners of the capital used in production).

Of the $100 of value added by the steel manufacturer, $80 goes to workers (labor income) and the remaining $20 goes to the firm (capital income). Of the $100 of value added by the car manufacturer, $70 goes to labor income and $30 to capital income.

The labor share in the example is thus 75%. In advanced countries, the share of labor is indeed typically between 60% and 75%.

For the economy as a whole, labor income is equal to $150($80 + $70) and capital income is equal to $50($20 + $30). Value added is equal to the sum of labor income and capital income: $200($150 + $50).

This leads to the third definition of GDP.

3. *GDP Is the Sum of Incomes in the Economy during a Given Period.*

Two lessons to remember:

i. GDP is the measure of aggregate output, which we can look at from the production side (aggregate production) or the income side (aggregate income); and

ii. Aggregate production and aggregate income are always equal.

To summarize: You can think about aggregate output—*GDP*—in three different but equivalent ways.

- From the *production side*: GDP equals the value of the final goods and services produced in the economy during a given period.
- Also from the *production side*: GDP is the sum of value added in the economy during a given period.
- From the *income side*: GDP is the sum of incomes in the economy during a given period.

Nominal and Real GDP

US GDP was $20,500 billion in 2018, compared to $543 billion in 1960. Was US output really almost 38 times higher in 2018 than in 1960? Obviously not: Much of the increase reflected an increase in prices rather than an increase in quantities produced. This leads to the distinction between nominal GDP and real GDP.

Warning! People often use *nominal* to denote small amounts. Economists use *nominal* for variables expressed in current prices, and they surely do not refer to small amounts: The numbers typically run in the billions or trillions of dollars.

Nominal GDP is the sum of the quantities of final goods produced times their current price. This definition makes clear that nominal GDP increases over time for two reasons:

- First, the production of most goods increases over time.
- Second, the price of most goods also increases over time.

If our goal is to measure production and its change over time, we need to eliminate the effect of increasing prices on our measure of GDP. That's why **real GDP** is constructed as the sum of the quantities of final goods times *constant* (rather than *current*) prices.

You may wonder why I chose these three particular years. The explanation will be given when I look at the actual numbers for the United States.

If the economy produced only one final good, say, a single car model, constructing real GDP would be easy: We would use the price of the car in a given year and multiply the quantity of cars produced in each year. An example will help here. Consider an economy that only produces cars—and to avoid issues we shall tackle later, assume the same model is produced every year. Suppose the number and the price of cars in three successive years are given by:

Year	Quantity of Cars	Price of Cars	Nominal GDP	Real GDP (in 2012 dollars)
2011	10	$20,000	$200,000	$240,000
2012	12	$24,000	$288,000	$288,000
2013	13	$26,000	$338,000	$312,000

Nominal GDP, which is equal to the quantity of cars times their price, goes up from $200,000 in 2011 to $288,000 in 2012—a 44% increase—and from $288,000 in 2012 to $338,000 in 2013—a 17% increase.

- To construct real GDP, we need to multiply the number of cars in each year by a *common* price. Suppose we use the price of a car in 2012 as the common price. This approach gives us *real GDP in 2012 dollars*.

- Using this approach, real GDP in 2011 (in 2012 dollars) equals 10 cars × $24,000 per car = $240,000. Real GDP in 2012 (in 2012 dollars) equals 12 cars × $24,000 per car = $288,000, the same as nominal GDP in 2012. Real GDP in 2013 (in 2012 dollars) is equal to 13 × $24,000 = $312,000.
  So real GDP goes up from $240,000 in 2011 to $288,000 in 2012—a 20% increase—and from $288,000 in 2012 to $312,000 in 2013—an 8% increase.
- How different would our results have been if we had decided to construct real GDP using the price of a car in, say, 2013 rather than 2012? Obviously, the level of real GDP in each year would be different (because the prices are not the same in 2013 than in 2012); but its rate of change from year to year would be the same as shown.

To check, compute real GDP in 2013 dollars, and compute the rate of growth from 2011 to 2012, and from 2012 to 2013.

The problem when constructing real GDP in practice is that there is obviously more than one final good. Real GDP must be defined as a weighted average of the output of all final goods, and this brings us to what the weights should be.

The *relative prices* of the goods would appear to be the natural weights. If one good costs twice as much per unit as another, then that good should count for twice as much as the other in the construction of real output. But this raises the question: What if, as is typically the case, relative prices change over time? Should we choose the relative prices of a particular year as weights, or should we change the weights over time? More discussion of these issues, and of the way real GDP is constructed in the United States, is in the appendix to this chapter. Here, what you should know is that the measure of real GDP in the US national income accounts uses weights that reflect relative prices and change over time. The measure is called **real GDP in chained (2012) dollars**. It says 2012 because, as in our example, at this point in time 2012 is the year when, by construction, real GDP is equal to nominal GDP. It is our best measure of the output of the US economy, and its evolution shows how US output has increased over time.

The year used to construct prices, at this point 2012, is called the *base year*. The base year is changed from time to time, and when you read this book, it may have changed again.

Figure 2-1 plots the evolution of both nominal GDP and real GDP since 1960. By construction, the two are equal in 2012. Real GDP in 2018 was about 5.7 times its level of 1960—a considerable increase, but clearly much less than the 38-fold increase in

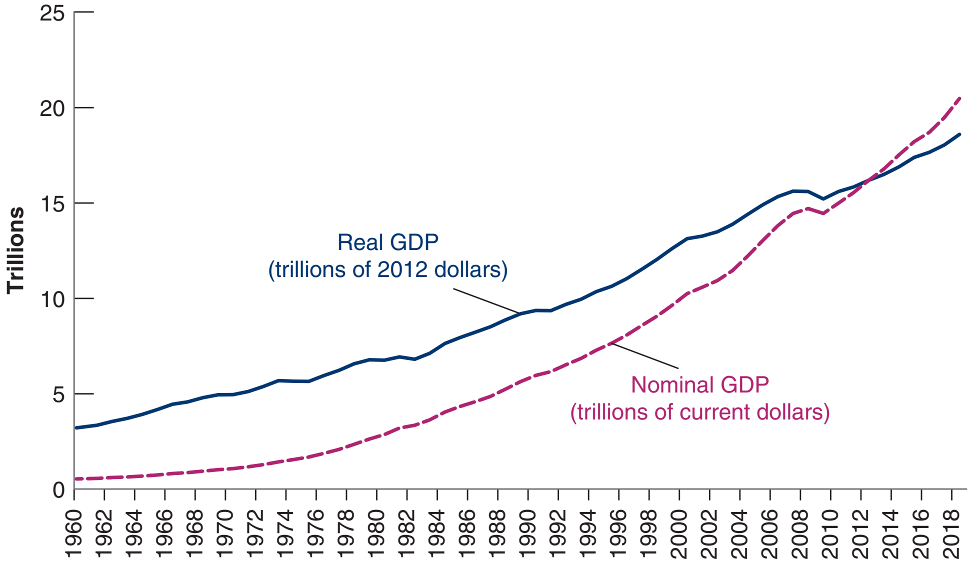


Figure 2-1

Nominal and Real US GDP, 1960–2018.

From 1960 to 2018, nominal GDP increased by a factor of 38. Real GDP increased by a factor of 5.7.

Source: FRED. Series GDPC, GDP.

Suppose real GDP was measured in 2000 dollars rather than 2012 dollars. Where would the nominal GDP and real GDP lines on the graph intersect?

nominal GDP over the same period. The difference between the two results comes from the increase in prices over the period.

The terms *nominal GDP* and *real GDP* each have many synonyms, and you are likely to encounter them in your readings:

- Nominal GDP is also called **dollar GDP** or **GDP in current dollars**.
- Real GDP is also called **GDP in terms of goods, GDP in constant dollars, GDP adjusted for inflation**, or **GDP in chained (2012) dollars**, or **GDP in 2012 dollars** (if the year in which real GDP is set equal to nominal GDP is 2012, as is the case in the United States at this time).

In the chapters that follow, unless I indicate otherwise,

- GDP will refer to *real GDP* and Y_t will denote *real GDP in year t*.
- Nominal GDP, and variables measured in current dollars, will be denoted by a dollar sign in front of them—for example, $\$Y_t$ for nominal GDP in year t.

GDP: Level versus Growth Rate

Warning: One must be careful about how one does the comparison: Recall the discussion in Chapter 1 about the standard of living in China. This is discussed further in Chapter 10.

We have focused so far on the *level* of real GDP. This is an important number that gives the economic size of a country. A country with twice the GDP of another country is economically twice as big as the other country. Equally important is the level of **real GDP per person**, the ratio of real GDP to the population of the country. It gives us the average standard of living of the country.

In assessing the performance of the economy from year to year, economists focus however on the rate of growth of real GDP, often called just **GDP growth**. Periods of positive GDP growth are called **expansions**. Periods of negative GDP growth are called **recessions**.

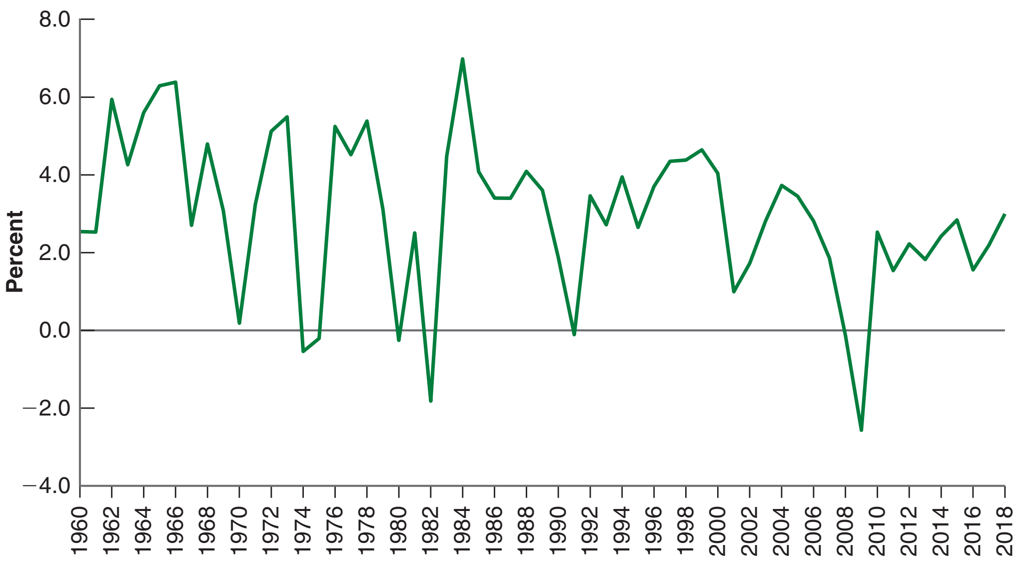
GDP growth in the United States since 1960 is given in Figure 2-2. GDP growth in year t is constructed as $(Y_t - Y_{t-1})/Y_{t-1}$ and expressed as a percentage. The figure shows how the US economy has gone through a series of expansions (periods of positive growth), interrupted by short recessions. Again, you can see the effects of the recent crisis: zero growth in 2008, and a large negative growth rate in 2009.

Figure 2-2

Growth Rate of US GDP, 1960–2018.

Since 1960, the US economy has gone through a series of expansions, interrupted by short recessions. The 2008–2009 recession was the most severe recession in the period from 1960 to 2018.

Source: Calculated using series GDPC in Figure 2-1.



FOCUS

Real GDP, Technological Progress, and the Price of Computers

A tough problem in calculating real GDP is how to deal with changes in quality of existing goods. One of the most difficult cases is computers. It would clearly be absurd to assume that a personal computer in 2019 is the same good as a personal computer produced, say, 20 years ago: The 2019 version can clearly do much more than the 1999 version. But how much more? How do we measure it? How do we take into account the improvements in internal speed, the size of the RAM (random access memory) or of the hard disk, faster access to the internet, and so on?

The approach used by economists to adjust for these improvements is to look at the market for computers and how it values computers with different characteristics in a given year. Example: Suppose the evidence from prices of different models on the market shows that people are willing to pay 10% more for a computer with a speed of 4 GHz (4,000 megahertz) rather than 3 GHz. (The first edition of this book, published in 1996, compared two computers, with speeds of 50 and 16 megahertz, respectively. This change is a good indication of technological progress.) Suppose new computers this year have a speed of 4 GHz compared to a speed of 3 GHz for new computers last year. (A further indication of the complexity of technological progress is that, in the more recent past, progress has been made not so much by increasing the speed of processors but rather by using multicore processors, which allow for faster parallel processing. We shall leave this aspect aside here, but people in charge of national income accounts cannot.) And suppose the dollar price of new computers this year is the same as the dollar price of new computers last year. Then economists in charge of computing the adjusted price of computers will conclude that new computers are in fact 10% cheaper than last year.

This approach, which treats goods as providing a collection of characteristics—for computers, speed, memory, and so on—each with an implicit price, is called **hedonic pricing** ("hedone" means "pleasure" in Greek. What matters in assessing the value of a good is how much utility ("pleasure") it provides). It is used by the Department of Commerce—which constructs real GDP—to estimate changes in the price of complex and fast-changing goods, such as automobiles and computers. Using this approach, the Department of Commerce estimates, for example, that for a given price, the quality of new laptops has increased on average by 20% a year since 1999 (if you want to look, the series is given by PCU33411133411172 in the FRED database). Put another way, a typical laptop in 2019 delivers $1.20^{21} = 46$ times the computing services a typical laptop delivered in 1999. (Interestingly, in light of the discussion of slowing US productivity growth in Chapter 1, the rate of quality improvement has decreased substantially in the recent past, and is now closer to 10%.)

Not only do laptops deliver more services, they have become cheaper as well: Their dollar price has declined by about 7% a year since 1999. Putting this together with the information in the previous paragraph, this implies that their quality-adjusted price has fallen at an average rate of $20\% + 7\% = 27\%$ per year. Put another way, a dollar spent on a laptop today buys $1.27^{21} = 151$ times more computing services than a dollar spent on a laptop in 1999.

2-2 THE UNEMPLOYMENT RATE

Because it is a measure of aggregate activity, GDP is obviously the most important macroeconomic variable. But two other variables, unemployment and inflation, tell us about other important aspects of how an economy is performing. This section focuses on the unemployment rate.

We start with two definitions: **Employment** is the number of people who have a job. **Unemployment** is the number of people who do not have a job but are looking for one. The **labor force** is the sum of employment and unemployment:

$$\underset{\text{labor force}}{L} = \underset{\text{employment}}{N} + \underset{\text{unemployment}}{U}$$

The **unemployment rate** is the ratio of the number of people who are unemployed to the number of people in the labor force:

$$u = \frac{U}{L}$$

$$\text{unemployment rate} = \text{unemployment/labor force}$$

Constructing the unemployment rate is less obvious than it might seem. Determining whether somebody is employed is relatively straightforward. Determining whether somebody is unemployed is more difficult. Recall from the definition that, to be classified as unemployed, a person must meet two conditions: he or she does not have a job, and he or she is looking for one; this second condition is harder to assess.

Until the 1940s in the United States, and until more recently in most other countries, the only available source of data on unemployment was the number of people registered at unemployment offices, and so only those workers who were registered in unemployment offices were counted as unemployed. This system led to a poor measure of unemployment. The number of those who were looking for jobs and were registered at the unemployment office varied both across countries and across time. Those who had no incentive to register—for example, those who had exhausted their unemployment benefits—were unlikely to take the time to come to the unemployment office, so they were not counted. Countries with less generous benefit systems were likely to have fewer unemployed people registered, and therefore smaller measured unemployment rates.

The 60,000 households are chosen as a representative sample of the whole US population. Thus, the sample provides good estimates of what is happening for the population as a whole.

Today, most rich countries rely on large surveys of households to compute the unemployment rate. In the United States, this survey is called the **Current Population Survey (CPS)**. It relies on interviews of 60,000 households every month. The survey classifies a person as employed if he or she has a job at the time of the interview; it classifies a person as unemployed if he or she does not have a job *and has been looking for a job in the last four weeks*. Most other countries use a similar definition of unemployment. In the United States, estimates based on the CPS show that, in December 2018, an average of 157 million people were employed, and 6.3 million people were unemployed, so the unemployment rate was $6.3/(157 + 6.3) = 3.9\%$.

Note that only those *looking for a job* are counted as unemployed; those who do not have a job and are not looking for one are counted as **not in the labor force**. When unemployment is high, some of the unemployed give up looking for a job and therefore are no longer counted as unemployed. These people are known as **discouraged workers**. Take an extreme example: If all workers without a job gave up looking for one, the unemployment rate would go to zero. This would make the unemployment rate a poor indicator of what is actually happening in the labor market. This example is too extreme; in practice, when the economy slows down, we typically observe both an increase in unemployment and an increase in the number of people who drop out of the labor force. Equivalently, a higher unemployment rate is typically associated with a lower **participation rate**, defined as the ratio of the labor force to the total population of working age.

Suppose that, in a given month, both employment and unemployment go up. What do you conclude?

Figure 2-3 shows the unemployment rate in the United States since 1960. It has fluctuated between 3% and 11%, going up during recessions and down during expansions. Again, you can see the effect of the recent crisis, with the unemployment rate reaching a peak at nearly 10% in 2010, the highest since the 1980s, followed by a steady decline since then.

Why Do Economists Care about Unemployment?

Economists care about unemployment for two reasons. First, they care about unemployment because of its direct effect on the welfare of the unemployed. Although unemployment benefits are more generous today than they were during the Great Depression, unemployment is still associated with financial and psychological suffering. The extent of suffering depends on the nature of unemployment.

One image of unemployment is that of a stagnant pool, of people remaining unemployed for long periods of time. In normal times, in the United States, this image is not

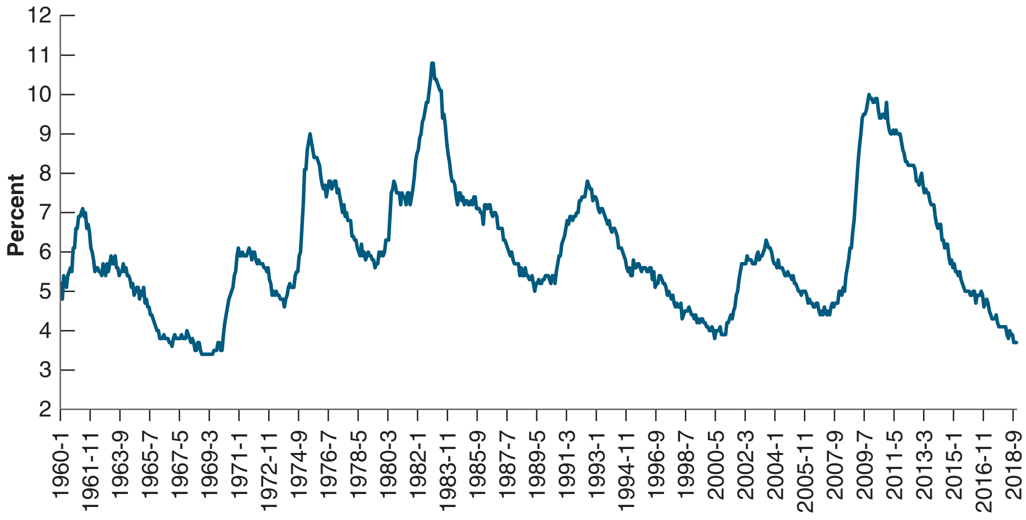


Figure 2-3

US Unemployment Rate, 1960–2018.

Since 1960, the US unemployment rate has fluctuated between about 3% and 11%.

Source: FRED Series: UNRATE.

right: Every month, many people become unemployed, and many of the unemployed find jobs. When unemployment is high, however, as it was during the crisis, another image becomes more relevant. Not only are more people unemployed, but also many of them are unemployed for a long time. For example, the mean duration of unemployment was 16 weeks on average during 2000–2007, but increased to 40 weeks in 2011. When unemployment increases, not only does unemployment become more widespread, it also becomes more painful for those who are unemployed.

Second, economists also care about the unemployment rate because it provides a signal that the economy is not using some of its resources. When unemployment is high, many workers who want to work do not find jobs; the economy is clearly not using its human resources efficiently. What about when unemployment is low? Can very low unemployment also be a problem? The answer is yes. Like an engine running at too high a speed, an economy in which unemployment is very low may be overusing its resources and run into labor shortages. How low is "too low"? This is a difficult question, and a question that, as of early 2019, is very relevant. The current rate of unemployment is below 4%, which is, as you can see from Figure 2-3, historically low. Whether it should be allowed to decrease further, or instead stabilized at the current level, is one of the main policy issues facing the Fed today.

It is probably because of statements like this that economics is known as the "dismal science."

2-3 THE INFLATION RATE

Inflation is a sustained rise in the general level of prices—the **price level**. The **inflation rate** is the rate at which the price level increases. (Symmetrically, **deflation** is a sustained decline in the price level. It corresponds to a negative inflation rate.)

Deflation is rare, but it happens. The United States experienced sustained deflation in the 1930s during the Great Depression (see the Focus Box in Chapter 9). Japan has had deflation, off and on, since the late 1990s. More recently, the euro area has had short spells of deflation.

The practical issue is how to define the price level so the inflation rate can be measured. Macroeconomists typically look at two measures of the price level, two *price indexes*: the GDP deflator and the Consumer Price Index.

The GDP Deflator

We saw how increases in nominal GDP can come either from an increase in real GDP, or from an increase in prices. Put another way, if we see nominal GDP increase faster than real GDP, the difference must come from an increase in prices.

FOCUS

Unemployment and Happiness

How painful is unemployment? To answer this question, one needs information about particular individuals and how their happiness varies as they become unemployed. This information is available from the German Socio-Economic Panel survey. The survey has followed about 11,000 households each year since 1984, asking each member of the household a number of questions about their employment status, their income, and their happiness. The specific question in the survey about happiness is the following: "How satisfied are you at present with your life as a whole?", with the answer rated from 0 ("completely dissatisfied") to 10 ("completely satisfied").

The effect of unemployment on happiness defined in this way is shown in Figure 1. The figure plots the average life satisfaction for individuals who were unemployed during one year, and employed in the four years before and in the four years after. Year 0 is the year of unemployment. Years −4 to −1 are the years before unemployment, years 1 to 4 the years after.

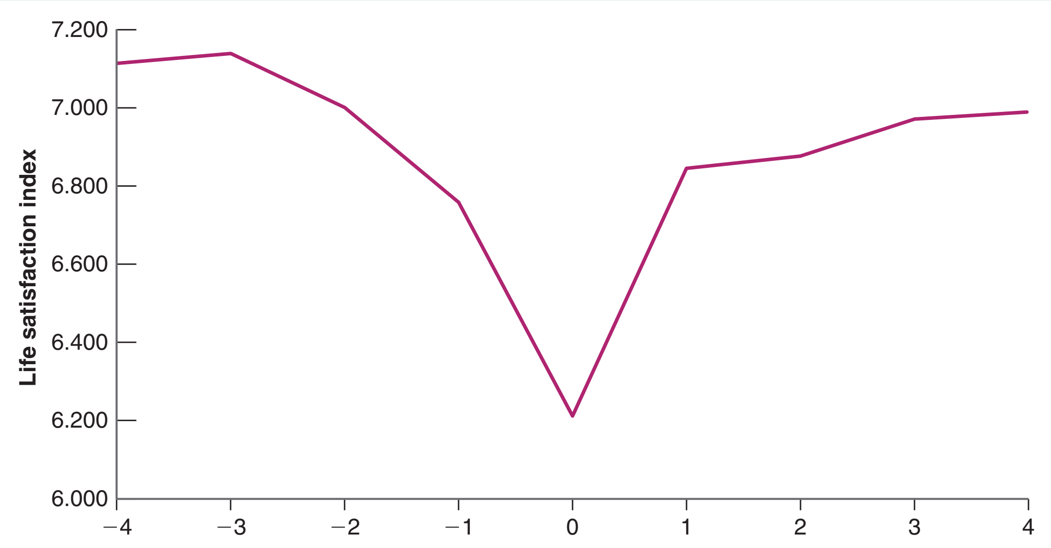
The figure suggests three conclusions. The first and main one is indeed that becoming unemployed leads to a large decrease in happiness. To give you a sense of scale, other studies suggest that this decrease in happiness is close to the decrease triggered by a divorce or a separation. The second is that happiness declines before the actual unemployment spell. This suggests that either workers know they are more likely to become unemployed, or that they like their job less and less. The third is that happiness does not fully recover even four years after the unemployment spell. This suggests that unemployment may do some long-lasting damage, either because of the experience of unemployment itself or because the new job is not as satisfying as the old one.

In thinking about how to deal with unemployment, it is essential to understand how unemployment decreases happiness. One important finding in this respect is that the decrease in happiness does not depend very much on the generosity of unemployment benefits. In other words, unemployment affects happiness not so much through financial channels as through psychological channels. To cite George Akerlof, a Nobel Prize–winning economist, "A person without a job loses not just his income but often the sense that he is fulfilling the duties expected of him as a human being."[1]

Figure 1

Effects of Unemployment on Happiness.

Source: Winkelmann 2014.



This remark motivates the definition of the GDP deflator. The **GDP deflator** in year t, P_t, is defined as the ratio of nominal GDP to real GDP in year t:

$$P_t = \frac{\text{Nominal GDP}_t}{\text{Real GDP}_t} = \frac{\$Y_t}{Y_t}$$

Note that, in the year in which, by construction, real GDP is equal to nominal GDP (2012 at this point in the United States), this definition implies that the price level is equal to 1. This is worth emphasizing: The GDP deflator is called an **index number**. Its level is

[1]*The material in this box, and in particular the figure, comes in part from "Unemployment and Happiness," by Rainer Winkelmann,* IZA World of Labor, 2014: *94, pp. 1–9.*

chosen arbitrarily—here it is equal to 1 in 2012—and has no economic interpretation. But its rate of change, $(P_t - P_{t-1})/P_{t-1}$ (which we shall denote by π_t in the rest of the book), has a clear economic interpretation: It gives the rate at which the general level of prices increases over time—the rate of inflation.

Index numbers are often set equal to 100 (in the base year) rather than to 1. If you look at the series for the GDP deflator in FRED (GDPDEF), it is equal to 100 for 2012 (the base year), 101.7 in 2013, and so on.

One advantage to defining the price level as the GDP deflator is that it implies a simple relation between *nominal GDP, real GDP*, and the *GDP deflator*. To see this, reorganize the previous equation to get:

$$\$Y_t = P_t Y_t$$

Nominal GDP is equal to the GDP deflator times real GDP. Or, putting it in terms of rates of change: The rate of growth of nominal GDP is equal to the rate of inflation plus the rate of growth of real GDP.

Compute the GDP deflator and the associated rate of inflation from 2011 to 2012 and from 2012 to 2013 in our car example in Section 2-1, when real GDP is constructed using the 2012 price of cars as the common price. (For a refresher on going from levels to rates of change, see Appendix 2 at the end of the book, Proposition 7.)

The Consumer Price Index

The GDP deflator gives the average price of output—the final goods *produced* in the economy. But consumers care about the average price of consumption—the goods they *consume*. The two prices need not be the same: The set of goods produced in the economy is not the same as the set of goods purchased by consumers, for two reasons:

- Some of the goods in GDP are sold not to consumers but to firms (machine tools, for example), to the government, or to foreigners.
- Some of the goods bought by consumers are not produced domestically but are imported from abroad.

To measure the average price of consumption, or, equivalently, the **cost of living**, macroeconomists look at another index, the **Consumer Price Index, or CPI**. The CPI has been in existence in the United States since 1917 and is published monthly (in contrast, numbers for GDP and the GDP deflator are constructed and published only quarterly).

Do not confuse the CPI with the PPI, or *producer price index*, which is an index of prices of domestically produced goods in manufacturing, mining, agriculture, fishing, forestry, and electric utility industries.

The CPI gives the cost in dollars of a specific list of goods and services over time. The list, which is based on a detailed study of consumer spending, attempts to represent the consumption basket of a typical urban consumer and is updated every two years.

Each month, Bureau of Labor Statistics (BLS) employees visit stores to find out what has happened to the price of the goods on the list; prices are collected for 211 items in 38 cities. These prices are then used to construct the CPI.

Like the GDP deflator (the price level associated with aggregate output, GDP), the CPI is an index. It is set equal to 100 in the period chosen as the base period and so its level has no particular significance. The current base period is 1982 to 1984, so the average for that period is equal to 100. In 2018, the CPI was 250; thus, it cost two and a half times as much in dollars to purchase the same consumption basket than in 1982–1984.

Do not ask why such a strange base period was chosen. Nobody seems to remember.

You may wonder how the rate of inflation differs depending on whether the GDP deflator or the CPI is used to measure it. The answer is given in Figure 2-4, which plots the two inflation rates since 1960 for the United States. The figure yields two conclusions:

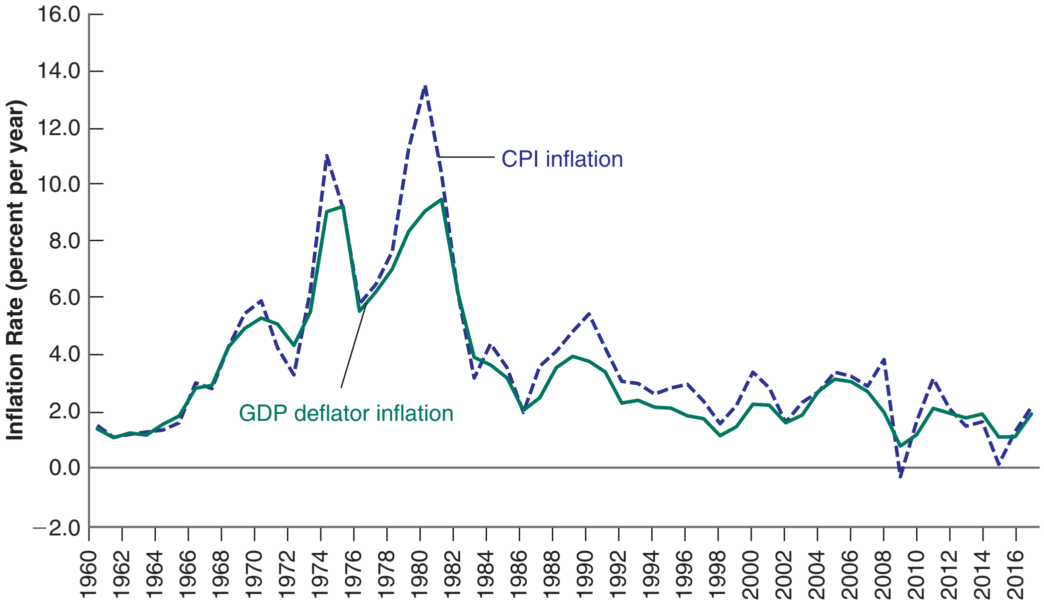
- The CPI and the GDP deflator move together most of the time. In most years, the two inflation rates differ by less than 1%.
- But there are clear exceptions. In 1979 and 1980, the increase in the CPI was significantly larger than the increase in the GDP deflator. The reason is not hard to find. Recall that the GDP deflator is the price of goods *produced* in the United States, whereas the CPI is the price of goods *consumed* in the United States. That

Figure 2-4

Inflation Rate, Using the CPI and the GDP Deflator, 1960–2018.

The inflation rates, computed using either the CPI or the GDP deflator, are largely similar.

Source: FRED: CPIAUCSL and GDPDEF.



means when the price of imported goods increases relative to the price of goods produced in the United States, the CPI increases faster than the GDP deflator. This is precisely what happened in 1979 and 1980. The price of oil doubled. And although the United States was a producer of oil, it produced less than it consumed: It was an oil importer. The result was a large increase in the CPI compared to the GDP deflator.

In what follows, we shall typically assume that the two indexes move together so we do not need to distinguish between them. We shall simply talk about *the price level* and denote it by P_t, without indicating whether we have the CPI or the GDP deflator in mind.

Why Do Economists Care about Inflation?

If a higher inflation rate meant just a faster but proportional increase in all prices and wages—a case called *pure inflation*—inflation would be only a minor inconvenience because relative prices would be unaffected.

Take, for example, the workers' *real wage*—the wage measured in terms of goods rather than dollars. Suppose that price inflation was 2%, and wage inflation 4%, so real wages increased by 2% a year, reflecting productivity growth. Now suppose that price inflation was instead 4% and wage inflation 6%. Real wages would still increase at $6\% - 4\% = 2\%$, the same as before. In other words, higher inflation would not affect real wages (or other relative prices). Inflation would not be entirely irrelevant; people would have to keep track of the increase in prices and wages when making decisions. But this would be a small burden, hardly justifying making control of the inflation rate one of the major goals of macroeconomic policy.

So why do economists care about inflation? Precisely because there is no such thing as pure inflation:

- During periods of inflation, not all prices and wages rise proportionately. Because they don't, inflation affects income distribution. For example, retirees in some countries receive payments that do not keep up with the price level, so they lose in relation to other groups when inflation is high. This is not the case in the United States, where Social Security benefits automatically rise with the CPI, protecting retirees from inflation. But during the very high inflation that took place in Russia in the 1990s, retirement pensions did not keep up with inflation, and many retirees were pushed to near starvation.

- Inflation leads to other distortions. Variations in relative prices also lead to more uncertainty, making it harder for firms to make decisions about the future, such as investment decisions. Some prices, which are fixed by law or by regulation, lag behind the others, leading to changes in relative prices. Taxation interacts with inflation to create more distortions. If tax brackets are not adjusted for inflation, for example, people move into higher and higher tax brackets as their nominal income increases, even if their real income remains the same.

This is known as *bracket creep*. In the United States, the tax brackets are adjusted automatically for inflation: If inflation is 5%, all tax brackets also go up by 5%—in other words, there is no bracket creep. By contrast, in Italy, where inflation averaged 17% a year in the second half of the 1970s, bracket creep led to a rise of almost 9 percentage points in the rate of income taxation.

If inflation is so bad, does this imply that deflation (negative inflation) is good?

Newspapers sometimes confuse deflation and recession. They may happen at the same time but they are not the same. Deflation is a decrease in the price level. A recession is a decrease in real output.

The answer is no. First, high deflation (a large negative rate of inflation) would create many of the same problems as high inflation, from distortions to increased uncertainty. Second, as we shall see in Chapter 4, even a low rate of deflation limits the ability of monetary policy to affect output. So what is the "best" rate of inflation? Most macroeconomists believe that the best rate of inflation is low and stable, somewhere between 1% and 4%.

We shall look at the pros and cons of different rates of inflation in Chapter 23.

2-4 OUTPUT, UNEMPLOYMENT, AND THE INFLATION RATE: OKUN'S LAW AND THE PHILLIPS CURVE

We have looked separately at the three main dimensions of aggregate economic activity: output growth, the unemployment rate, and the inflation rate. Clearly, they are not independent, and much of this book will be spent looking at the relations among them in detail. But it is useful to have a first pass now.

Okun's Law

Intuition suggests that if output growth is high, unemployment will decrease, and this is indeed true. This relation was first examined by US economist Arthur Okun and for this reason has become known as **Okun's law**. Figure 2-5 plots quarterly changes in the unemployment rate on the vertical axis against the quarterly rate of growth of output on the horizontal axis for the United States since the first quarter of 2000. It also draws the line that best fits the cloud of points. Looking at the figure and the line suggests two conclusions:

Arthur Okun was an adviser to President John F. Kennedy in the 1960s. Okun's law is, of course, not a law but an empirical regularity.

- The line is downward sloping and fits the cloud of points quite well. Put in economic terms: There is a strong relation between the two variables: Higher output growth leads to a decrease in unemployment. The slope of the line is—0.3. This implies that, on average, an increase in the growth rate of 1% decreases the unemployment rate by roughly—0.3%. This is why unemployment goes up in recessions and down in expansions. This relation has a simple but important implication: The key to decreasing unemployment is a high enough rate of growth.

Such a graph, plotting one variable against another, is called a *scatterplot*. The line is called *a regression line*. For more on regressions, see Appendix 3 at the end of the book.

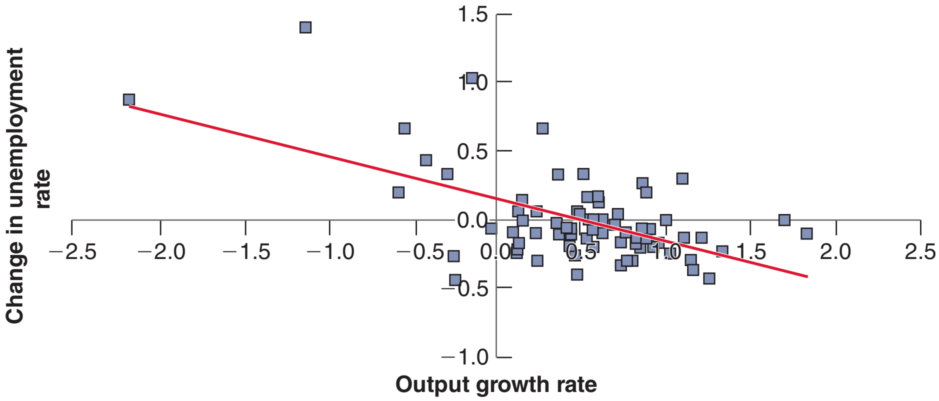
- This line crosses the horizontal axis at the point where quarterly output growth is roughly equal to 0.5%, equivalently when annual output growth is equal to 2%. In economic terms: It takes an annual growth rate of about 2% to keep unemployment

Figure 2-5

Changes in the Unemployment Rate versus Growth in the United States, 2000 Q1 to 2018 Q4.

Output growth that is higher than usual is associated with a reduction in the unemployment rate; output growth that is lower than usual is associated with an increase in the unemployment rate.

Source: FRED: Series GDPC, UNRATE.



constant. This is for two reasons. The first is that population, and thus the labor force, increases over time, so employment must grow over time just to keep the unemployment rate constant. The second is that output per worker is also increasing with time, which implies that output growth is higher than employment growth. Suppose, for example, that the labor force grows at 1% and that output per worker grows at 1%. Then output growth must be equal to 1% + 1% = 2% just to keep the unemployment rate constant.

The Phillips Curve

It should probably be known as the Phillips relation, but it is too late to change the name. ▶

Okun's law implies that, with strong enough growth, one can decrease the unemployment rate to very low levels. But intuition suggests that, when unemployment becomes very low, the economy is likely to overheat, and that this will lead to upward pressure on inflation. And, to a large extent, this is true. This relation was first explored in 1958 by a New Zealand economist, A. W. Phillips, and has become known as the **Phillips curve**. Phillips plotted the rate of inflation against the unemployment rate. Figure 2-6 does the same by plotting, on the vertical axis, the quarterly **core inflation rate**, which is the inflation rate constructed by leaving out volatile prices, such as food and energy, against the unemployment rate on the horizontal axis, together with the line that fits the cloud of points best, for the United States, quarterly since the first quarter of 2000. Looking at the figure again suggests two conclusions:

- The line is downward sloping, although the fit is definitely not as good as it was for Okun's law: Higher unemployment is associated, on average, with lower inflation; lower unemployment is associated with higher inflation. But this is only true on average. As we shall see later in Chapter 8, not only is the Phillips curve relation not as tight as Okun's law, but it has evolved over time, complicating in important ways the job of central banks, which have to care about both inflation and unemployment.
- Using the regression line, we can compute the rate of unemployment associated with a given rate of inflation. If, for example, we want the inflation rate to be 2%, which is the current target of the Fed and many other central banks, the line implies that the unemployment rate has to be roughly equal to 5%. In economic terms, since 2000, when unemployment has been below 5%, inflation has typically been above 2%. When unemployment has been above 5%, inflation has typically been above 2%. But again, the relation is not tight enough that the required unemployment rate can be pinned down precisely. Indeed, at the time of writing, unemployment is lower than 4% and core inflation is 2.2%, barely above 2%.

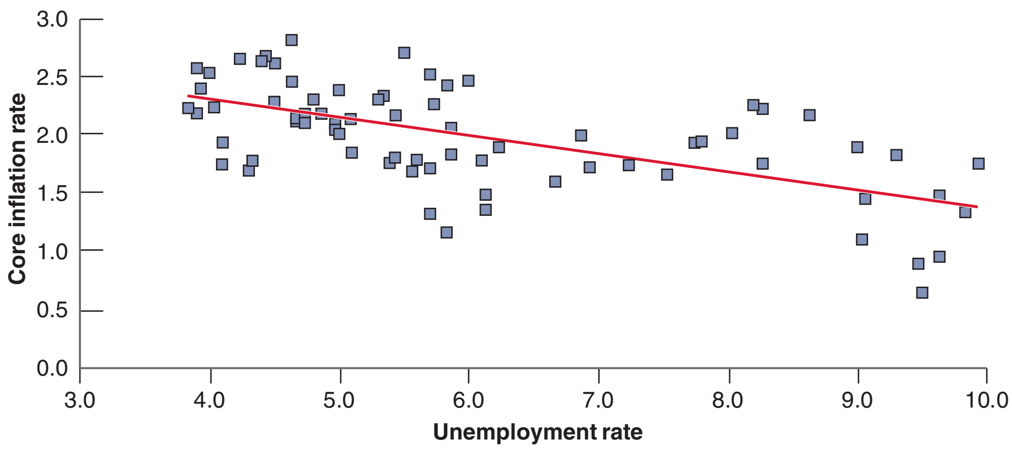


Figure 2-6

Changes in the Inflation Rate versus the Unemployment Rate in the United States, 2000 Q1 To 2018 Q4.

Lower unemployment rate is associated with a higher inflation rate, higher unemployment rate with a lower inflation rate.

Source: FRED. Series GDPC, CPILFESL.

Clearly, a successful economy is an economy that combines high output growth, low unemployment, and low inflation. Can all these objectives be achieved simultaneously? Is low unemployment compatible with low and stable inflation? Do policymakers have the tools to sustain growth, to achieve low unemployment while maintaining low inflation? These are the questions we shall take up as we go through the book. The next two sections give you the road map.

2-5 THE SHORT RUN, THE MEDIUM RUN, AND THE LONG RUN

What determines the level of aggregate output in an economy? Consider three answers:

- Newspaper articles suggest a first answer: Movements in output come from movements in the demand for goods. You probably have read news stories that begin like this: "Production and sales of automobiles were higher last month due to a surge in consumer confidence, which drove consumers to showrooms in record numbers." Stories like these highlight the role demand plays in determining aggregate output; they point to factors that affect demand, ranging from consumer confidence to government spending to interest rates.
- But, surely, no amount of Indian consumers rushing to Indian showrooms can raise India's output to the level of output in the United States. This suggests a second answer: What matters when it comes to aggregate output is the supply side—how much the economy can produce. How much can be produced depends on how advanced the technology of the country is, how much capital it is using, and the size and the skills of its labor force. These factors—not consumer confidence—are the fundamental determinants of a country's level of output.
- The previous argument can be taken one step further: Neither technology, nor capital, nor skills are given. The technological sophistication of a country depends on its ability to innovate and introduce new technologies. The size of its capital stock depends on how much people have saved. The skills of workers depend on the quality of the country's education system. Other factors are also important: If firms are to operate efficiently, for example, they need a clear system of laws under which to operate and an honest government to enforce those laws. This suggests a third answer: The true determinants of output are factors

like a country's education system, its saving rate, and the quality of its government. If we want to understand what determines the level of output, we must look at these factors.

The next three bullet points may be the most important lesson of the book.

You might be wondering at this point, which of the three answers is right? The fact is that all three are right. But each applies over a different time frame:

- In the **short run**, say, a few years, the first answer is the right one. Year-to-year movements in output are primarily driven by movements in demand. Changes in demand, perhaps as a result of changes in consumer confidence or other factors, can lead to a decrease in output (a recession) or an increase in output (an expansion).
- In the **medium run**, say, a decade, the second answer is the right one. Over the medium run, the economy tends to return to the level of output determined by supply factors: the capital stock, the level of technology, and the size of the labor force. And, over a decade or so, these factors move sufficiently slowly that we can take them as given.
- In the **long run**, say, a few decades or more, the third answer is the right one. To understand why China has been able to achieve such a high growth rate since 1980, we must understand why both the capital stock and the level of technology in China are increasing so fast. To do so, we must look at factors like the education system, the saving rate, and the role of the government.

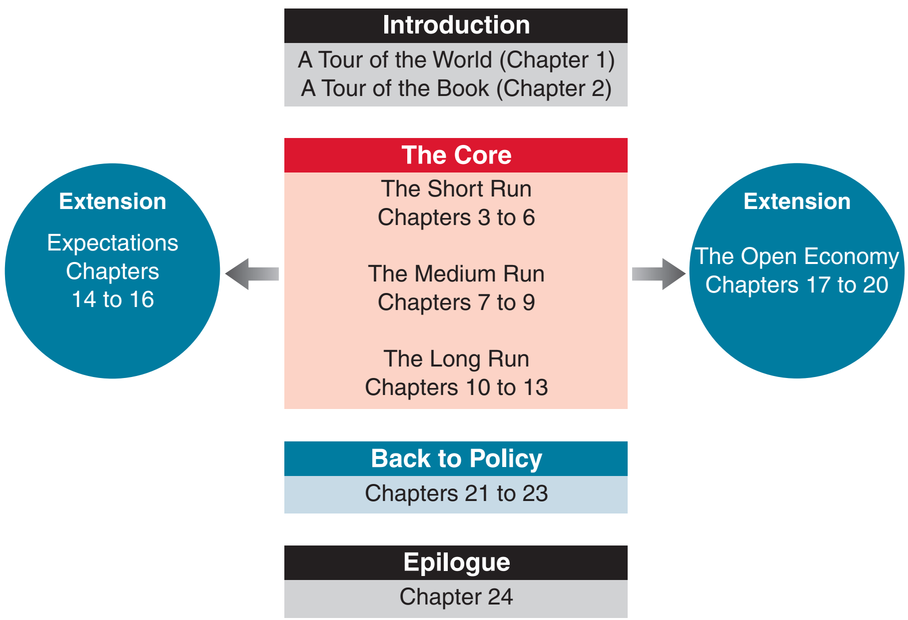
This way of thinking about the determinants of output underlies macroeconomics, and it underlies the organization of this book.

2-6 A TOUR OF THE BOOK

The book is organized in three parts: A core; two extensions; and, finally, a comprehensive look at the role of macroeconomic policy. This organization is shown in Figure 2-7. We now describe it in more detail.

Figure 2-7

The Organization of the Book.



The Core

The core is composed of three parts—the short run, the medium run, and the long run.

- Chapters 3 to 6 look at how output is determined in the short run. To focus on the role of demand, we assume that firms are willing to supply any quantity at a given price. In other words, we ignore supply constraints. Chapter 3 shows how the demand for goods determines output. Chapter 4 shows how monetary policy determines the interest rate. Chapter 5 puts the two together, by allowing demand to depend on the interest rate, and then showing the role of monetary and fiscal policy in determining output. Chapter 6 extends the model by introducing a richer financial system and using it to explain what happened during the financial crisis.
- Chapters 7 to 9 develop the supply side and look at how output is determined in the medium run. Chapter 7 introduces the labor market. Chapter 8 builds on it to derive the relation between inflation and unemployment. Chapter 9 puts all the parts together, and shows the determination of output, unemployment, and inflation in both the short and the medium run.
- Chapters 10 to 13 focus on the long run. Chapter 10 introduces the relevant facts by looking at the growth of output both across countries and over long periods of time. Chapters 11 and 12 discuss how both capital accumulation and technological progress determine growth. Chapter 13 looks at the challenges to growth, from inequality to global warming.

Extensions

The core chapters give you a way of thinking about how output (and unemployment, and inflation) is determined over the short, medium, and long run. However, they leave out several elements, which are explored in two extensions:

- Expectations play an essential role in macroeconomics. Nearly all the economic decisions people and firms make depend on their expectations about future income, future profits, future interest rates, and so on. Fiscal and monetary policies affect economic activity not only through their direct effects, but also through their effects on people's and firms' expectations. Although we touch on these issues in the core, Chapters 14 to 16 offer a more detailed treatment and draw the implications for fiscal and monetary policy.
- The core chapters treat the economy as *closed*, ignoring its interactions with the rest of the world. But the fact is, economies are increasingly *open*, trading goods and services and financial assets with one another. As a result, countries are becoming more and more interdependent. The nature of this interdependence and the implications for fiscal and monetary policy are the topics of Chapters 17 to 20.

Back to Policy

Monetary and fiscal policies are discussed in nearly every chapter of this book. But once the core and the extensions have been covered, it is useful to go back and put things together.

Chapter 21 focuses on general issues of policy, whether macroeconomists know enough about how the economy works to use policy as a stabilization tool at all, and whether policymakers can be trusted to do what is right.

Chapters 22 and 23 return to the role of fiscal and monetary policies.

Epilogue

Macroeconomics is not a fixed body of knowledge. It evolves over time. The final chapter, Chapter 24, looks at the history of macroeconomics and how macroeconomists have come to believe what they believe today. From the outside, macroeconomics sometimes looks like a field divided among schools of economists—"Keynesians," "monetarists," "new classicals," "supply-siders," and so on—hurling arguments at each other. The actual process of research is more orderly and more productive than this image suggests. I identify what I see as the main differences among macroeconomists, and the set of propositions that define the core of macroeconomics today.

SUMMARY

- We can think of GDP, the measure of aggregate output, in three equivalent ways: (1) GDP is the value of the final goods and services produced in the economy during a given period; (2) GDP is the sum of value added in the economy during a given period; and (3) GDP is the sum of incomes in the economy during a given period.
- Nominal GDP is the sum of the quantities of final goods produced times their current prices. This implies that changes in nominal GDP reflect both changes in quantities and changes in prices. Real GDP is a measure of output. Changes in real GDP reflect changes in quantities only.
- A person is classified as unemployed if he or she does not have a job and is looking for one. The unemployment rate is the ratio of the number of people unemployed to the number of people in the labor force. The labor force is the sum of those employed and those unemployed.
- Economists care about unemployment because of the human cost it represents. They also look at unemployment because it sends a signal about how efficiently the economy is using its resources. High unemployment indicates that the country is not using its resources efficiently.
- Inflation is a rise in the general level of prices—the price level. The inflation rate is the rate at which the price level increases. Macroeconomists look at two measures of the price level. The first is the GDP deflator, which is the average price of the goods produced in the economy. The second is the Consumer Price Index (CPI), which is the average price of goods consumed in the economy.
- Inflation leads to changes in income distribution, to distortions, and to increased uncertainty.
- There are two important relations among output, unemployment, and inflation. The first, called Okun's law, is a relation between output growth and the change in unemployment: High output growth typically leads to a decrease in the unemployment rate. The second, called the Phillips curve, is a relation between unemployment and inflation: A lower unemployment rate typically leads to a higher inflation rate.
- Macroeconomists distinguish between the short run (a few years), the medium run (a decade), and the long run (a few decades or more). They think of output as being determined by demand in the short run. They think of output as being determined by the level of technology, the capital stock, and the labor force in the medium run. Finally, they think of output as being determined by factors like education, research, saving, and the quality of government in the long run.

KEY TERMS

national income and product accounts, 20
aggregate output, 20
gross domestic product (GDP), 20
gross national product (GNP), 20
intermediate good, 20
final good, 21
value added, 21
nominal GDP, 22
real GDP, 22
real GDP in chained (2009) dollars, 23
dollar GDP, GDP in current dollars, 24
GDP in terms of goods, GDP in constant dollars, GDP adjusted for inflation, GDP in chained 2012 dollars, GDP in 2012 dollars, 24
real GDP per person, 24
GDP growth, 24
expansions, 24
recessions, 24
hedonic pricing, 25
employment, 25
unemployment, 25
labor force, 25

unemployment rate, 25
Current Population Survey (CPS), 26
not in the labor force, 26
discouraged workers, 26
participation rate, 26
inflation, 27
price level, 27
inflation rate, 27
deflation, 27
GDP deflator, 28
index number, 28
cost of living, 29
Consumer Price Index (CPI), 29
Okun's law, 31
Phillips curve, 32
core inflation, 32
short run, 34
medium run, 34
long run, 34

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. US GDP was 38 times higher in 2018 than it was in 1960.
b. When the unemployment rate is high, the participation rate is also likely to be high.
c. The rate of unemployment tends to fall during expansions and rise during recessions.
d. If the Japanese CPI is currently at 108 and the US CPI is at 104, then the Japanese rate of inflation is higher than the US rate of inflation.
e. The rate of inflation computed using the CPI is a better index of inflation than the rate of inflation computed using the GDP deflator.
f. Okun's law shows that when output growth is lower than normal, the unemployment rate tends to rise.
g. Periods of negative GDP growth are called *recessions*.
h. When the economy is functioning normally, the unemployment rate is zero.
i. The Phillips curve is a relation between the level of prices and the level of unemployment.

2. Suppose you are measuring annual US GDP by adding up the final value of all goods and services produced in the economy. Determine the effect on GDP of each of the following transactions.

a. A seafood restaurant buys $100 worth of fish from a fisherman.
b. A family spends $100 on a fish dinner at a seafood restaurant.
c. Delta Air Lines buys a new jet from Boeing for $200 million.
d. The Greek national airline buys a new jet from Boeing for $200 million.
e. Delta Air Lines sells one of its jets to Jennifer Lawrence for $100 million.

3. During a given year, the following activities occur:

i. A silver mining company pays its workers $200,000 to mine 75 pounds of silver. The silver is then sold to a jewelry manufacturer for $300,000.
ii. The jewelry manufacturer pays its workers $250,000 to make silver necklaces, which the manufacturer sells directly to consumers for $1,000,000.

a. Using the production-of-final-goods approach, what is GDP in this economy?
b. What is the value added at each stage of production? Using the value-added approach, what is GDP?
c. What are the total wages and profits earned? Using the income approach, what is GDP?

4. An economy produces three goods: cars, computers, and oranges. Quantities and prices per unit for years 2012 and 2013 are as follows:

	2012		2013	
	Quantity	Price	Quantity	Price
Cars	10	$2000	12	$3000
Computers	4	$1000	6	$500
Oranges	1,000	$1	1000	$1

a. What is nominal GDP in 2012 and in 2013? By what percentage does nominal GDP change from 2012 to 2013?
b. Using the prices for 2012 as the set of common prices, what is real GDP in 2012 and in 2013? By what percentage does real GDP change from 2012 to 2013?
c. Using the prices for 2013 as the set of common prices, what is real GDP in 2012 and in 2013? By what percentage does real GDP change from 2012 to 2013?
d. Why are the two output growth rates constructed in parts b and c different? Which one is correct? Explain your answer.

5. Consider the economy described in Problem 4.

a. Use the prices for 2012 as the set of common prices to compute real GDP in 2012 and in 2013. Compute the GDP deflator for 2012 and for 2013, and compute the rate of inflation from 2012 to 2013.
b. Use the prices for 2013 as the set of common prices to compute real GDP in 2012 and in 2013. Compute the GDP deflator for 2012 and for 2013 and compute the rate of inflation from 2012 to 2013.
c. Why are the two rates of inflation different? Which one is correct? Explain your answer.

6. Consider the economy described in Problem 4.

a. Construct real GDP for years 2012 and 2013 by using the average price of each good over the two years.
b. By what percentage does real GDP change from 2012 to 2013?
c. What is the GDP deflator in 2012 and 2013? Using the GDP deflator, what is the rate of inflation from 2012 to 2013?
d. Is this an attractive solution to the problems pointed out in Problems 4 and 5 (i.e., two different growth rates and two different inflation rates, depending on which set of prices is used)? (The answer is yes and is the basis for the construction of chained-type deflators. See the appendix to this chapter for more discussion.)

7. The Consumer Price Index

The Consumer Price Index represents the average price of goods that households consume. Many thousands of goods are included in such an index. Here consumers are represented as buying only food (pizza) and gas as their basket of goods. Here is a representation of the kind of data the Bureau of Economic Analysis collects to construct a consumer price index. In the base year, 2012, both the prices of goods purchased and the quantity of goods purchased are collected. In subsequent years, only prices are collected. Each year, the agency collects the price of that good and constructs an index of prices that represents two exactly equivalent concepts: How much more money does it take to buy the same basket of goods in the current year than in the base year? How much has the purchasing power of money declined, measured in baskets of goods, in the current year, from the base year?

The data: In an average week in 2012, the Bureau of Economic Analysis surveys many consumers and determines that the average consumer purchases 2 pizzas and 6 gallons of gas in a week. The prices per pizza and per gallon in subsequent years are shown below.

Year	Price of Pizzas	Price of Gas
2012	$10	$3
2013	$11	$3.30
2014	$11.55	$3.47
2015	$11.55	$3.50
2016	$11.55	$2.50
2017	$11.55	$3.47

a. What is the cost of the consumer price basket in 2012?
b. What is the cost of the consumer price basket in 2013 and in subsequent years?
c. Represent the annual cost of the consumer price basket as an index number. Set the value of the index number equal to 100 in 2012.
d. Calculate the annual rate of inflation using the percent change in the value of the index number between each year from 2013 on.

You may find it helpful to fill in the table below:

Year	Consumer Price Index 2012 = 100	Inflation rate
2012	100	
2013		
2014		
2015		
2016		
2017		

e. Is there a year in which inflation is negative? Why does this happen?
f. What is the source of inflation in the year 2015? How is that different from inflation in 2013 and 2014?
g. How many baskets of goods can I buy with $100 in 2012? How many baskets can I buy with that money in 2017? What is the percentage decline in the purchasing power of my money? How does that decline relate to the change in the value of the price index between 2012 and 2017?
h. From 2013 to 2015, the price of a pizza remains the same. The price of gas rises. How might consumers respond to such a change? In 2016, the price of gas falls. What are the implications of such changes in relative prices for the construction of the Consumer Price Index?
i. Suppose the Bureau of Economic Analysis determines that in 2017, the average consumer buys 2 pizzas and 7 gallons of gas in a week. Use a spreadsheet to calculate the Consumer Price Index set equal to 100 in 2017 and use the 2017 basket in the years from 2012 to 2017. Fill in the table below:

Year	Consumer Price Index 2017 = 100	Inflation rate
2012		
2013		
2014		
2015		
2016		
2017	100	

Why are the inflation rates (slightly) different in part d and part i?

8. Using macroeconomic relations:

a. Okun's law states that when output growth is higher than usual, the unemployment rate tends to fall. Explain why usual output growth is positive.

b. In which year—a year in which output growth is 2% or a year in which it is –2%—will the unemployment rate rise more?
c. The Phillips curve is a relation between the inflation rate and the unemployment rate. Using the Phillips curve, is the unemployment rate zero when the rate of inflation is 2%?
d. The Phillips curve is often portrayed as a line with a negative slope. In Figure 2-6, the slope is about −0.17. In your opinion, will the economy be "better" if the line has a large slope, say –0.5, or a smaller slope, say −0.1?

DIG DEEPER

9. Hedonic pricing

As the first Focus box in this chapter explains, it is difficult to measure the true increase in prices of goods whose characteristics change over time. For such goods, part of any price increase can be attributed to an increase in quality. Hedonic pricing offers a method to compute the quality-adjusted increase in prices.

a. Consider the case of a routine medical check-up. Name some reasons you might want to use hedonic pricing to measure the change in the price of this service.

Now consider the case of a medical check-up for a pregnant woman. Suppose that a new ultrasound method is introduced. In the first year that this method is available, half of doctors offer the new method, and half offer the old method. A check-up using the new method costs 10% more than a check-up using the old method.

b. In percentage terms, how much of a quality increase does the new method represent over the old method? (Hint: Consider the fact that some women *choose* to see a doctor offering the new method when they could have chosen to see a doctor offering the old method.)

Now, in addition, suppose that in the first year the new ultrasound method is available, the price of check-ups using the new method is 15% higher than the price of check-ups in the previous year (when everyone used the old method).

c. How much of the higher price for check-ups using the new method (as compared to check-ups in the previous year) reflects a true price increase of check-ups and how much represents a quality increase? In other words, how much higher is the quality-adjusted price of check-ups using the new method as compared to the price of check-ups in the previous year?

In many cases, the kind of information we used in parts (b) and (c) is not available. For example, suppose that in the year the new ultrasound method is introduced, all doctors adopt the new method, so the old method is no longer used. In addition, continue to assume that the price of check-ups in the year the new method is introduced is 15% higher than the price of check-ups in the previous year (when everyone used the old method). Thus, we observe a 15% price increase in check-ups, but we realize that the quality of check-ups has increased.

d. Under these assumptions, what information required to compute the quality-adjusted price increase of check-ups is lacking? Even without this information, can we say anything about the quality-adjusted price increase of check-ups? Is it more than 15%? Less than 15%? Explain.

10. Measured and true GDP

Suppose that instead of cooking dinner for an hour, you decide to work an extra hour, earning an additional $12. You then purchase some (takeout) Chinese food, which costs you $10.

a. By how much does measured GDP increase?
b. Do you think the increase in measured GDP accurately reflects the effect on output of your decision to work? Explain.

EXPLORE FURTHER

11. Comparing the recessions of 2000 and 2008.

One very easy to use source of data is the FRED database. The series that measures real GDP is GDPC1, real GDP in each quarter of the year expressed at a seasonally adjusted annual rate (denoted SAAR). The monthly series for the unemployment rate is UNRATE. You can download these series in a variety of ways from this database.

a. Look at the data on quarterly real GDP growth from 1999 through 2001 and then from 2007 through 2009. Which recession has larger negative values for GDP growth, the recession centered on 2000 or the recession centered on 2008?
b. The unemployment rate is series UNRATE. Is the unemployment rate higher in the 2001 recession or the 2009 recession?
c. The National Bureau of Economic Research (NBER), which dates recessions, identified a recession beginning in March 2001 and ending in November 2001. The equivalent dates for the next, longer recession were December 2007 ending June 2009. In other words, according to the NBER, the economy began a recovery in November 2001 and in June 2009. Given your answers to parts a and b, do you think the labor market recovered as quickly as GDP? Explain.

For more on NBER recession dating, visit www.nber.org. This site provides a history of recession dates and some discussion of NBER's methodology.

FURTHER READINGS

- In 1995, the US Senate set up a commission to study the construction of the CPI and make recommendations about potential changes. The commission concluded that the rate of inflation computed using the CPI was on average about 1% too high. If this conclusion is correct, this implies in particular that real wages (nominal wages divided by the CPI) have grown 1% more per year than is currently being reported. For more on the conclusions of the commission and some of the exchanges that followed, read "Consumer Prices, the Consumer Price Index, and the Cost of Living," by Michael Boskin et al., *Journal of Economic Perspectives*, 1998, 12(1): pp. 3–26.

- For a short history of the construction of the National Income Accounts, read *GDP: One of the Great Inventions of the 20th Century*, Survey of Current Business, January 2000, 1–9 (www.bea.gov/scb/pdf/BEAWIDE/2000/0100od.pdf).
- For a discussion of some of the problems involved in measuring activity, read Katherine Abraham, "What We Don't Know Could Hurt Us; Some Reflections on the Measurement of Economic Activity," *Journal of Economic Perspectives*, 2005, 19(3): pp. 3–18.
- To see why it is hard to measure the price level and output correctly, read "Viagra and the Wealth of Nations" by Paul Krugman, *New York Times*, August 23, 1998 (www.pkarchive.org/theory/viagra.html). (Paul Krugman is a Nobel Prize–winning economist and a columnist at the *New York Times*. His columns are opinionated, insightful, and fun to read.)
- Data underlying the construction of the CPI are typically collected monthly, from visits to stores. With the advent of the internet and big data methods, the process seems archaic. And, indeed, the Billion Prices Project (www.thebillionpricesproject.com/) now constructs daily price indexes, based on 15 million products from 900 retailers in 20 countries. The project turned out to particularly useful in Argentina in the late 2000s when the government started manipulating official numbers in order to minimize reported inflation. The project was able to show that the true rate of inflation was more than double the rate reported by the government.

APPENDIX: The Construction of Real GDP and Chain-Type Indexes

The example we used in the chapter had only one final good—cars—so constructing real GDP was easy. But how do we construct real GDP when there is more than one final good? This appendix gives the answer.

To understand how real GDP in an economy with many final goods is constructed, all you need to do is look at an economy where there are just two final goods. What works for two goods works just as well for millions of goods.

Suppose that an economy produces two final goods, say wine and potatoes:

- In year 0, it produces 10 pounds of potatoes at a price of $1 a pound, and 5 bottles of wine at a price of $2 a bottle.
- In year 1, it produces 15 pounds of potatoes at a price of $1 a pound, and 5 bottles of wine at a price of $3 a bottle.
- Nominal GDP in year 0 is therefore equal to $20. Nominal GDP in year 1 is equal to $30.
- This information is summarized in the following table.

Nominal GDP in Year 0 and in Year 1.

	Year 0		
	Quantity	$ Price	$ Value
Potatoes (pounds)	10	1	10
Wine (bottles)	5	2	10
Nominal GDP			20
	Year 1		
	Quantity	$ Price	$ Value
Potatoes (pounds)	15	1	15
Wine (bottles)	5	3	15
Nominal GDP			30

The rate of growth of nominal GDP from year 0 to year 1 is equal to $(\$30 - \$20)/\$20 = 50\%$. But what is the rate of growth of real GDP?

Answering this question requires constructing real GDP for each of the two years. The basic idea behind constructing real GDP is to evaluate the quantities in each year using the *same set of prices*.

Suppose we choose, for example, the prices in year 0. Year 0 is then called the **base year**. In this case, the computation is as follows:

- Real GDP in year 0 is the sum of the quantity in year 0 times the price in year 0 for both goods: $(10 \times \$1) + (5 \times \$2) = \$20$.
- Real GDP in year 1 is the sum of the quantity in year 1 times the price in year 0 for both goods: $(15 \times \$1) + (5 \times \$2) = \$25$.
- The rate of growth of real GDP from year 0 to year 1 is then $(\$25 - \$20)/\$20$, or 25%.

This answer raises, however, an obvious issue: Instead of using year 0 as the base year, we could have used year 1, or any other year. If, for example, we had used year 1 as the base year, then:

- Real GDP in year 0 would be equal to $(10 \times \$1 + 5 \times \$3) = \$25$.
- Real GDP in year 1 would be equal to $(15 \times \$1 + 5 \times \$3) = \$30$.
- The rate of growth of real GDP from year 0 to year 1 would be equal to $\$5/\$25$, or 20%.

The answer using year 1 as the base year would therefore be different from the answer using year 0 as the base year. So if the choice of the base year affects the constructed percentage rate of change in real output, which base year should one choose?

Until the mid-1990s in the United States—and still in most countries today—the practice was to choose a base year and change it infrequently, say, every five years or so. For example, in the United States, 1987 was the base year used from December 1991 to December 1995. That is, measures of real GDP published, for example, in 1994 for both 1994 and for all earlier years were constructed using 1987 prices. In December 1995, national income accounts shifted to 1992 as a base year; measures of real GDP for all earlier years were recalculated using 1992 prices.

This practice was logically unappealing. Every time the base year was changed and a new set of prices was used, all past real GDP numbers—and all past real GDP growth rates—were recomputed: Economic history was, in effect, rewritten every five years! Starting in December 1995, the US Bureau of Economic Analysis (BEA)—the government office that produces the GDP numbers—shifted to a new method that does not suffer from this problem.

The method requires four steps:

1. Constructing the rate of change of real GDP from year t to year $t + 1$ in two different ways: First using the prices from year t as the set of common prices; second, using the prices from year $t + 1$ as the set of common prices. For example, the rate of change of GDP from 2017 to 2018 is computed by:
   - Constructing real GDP for 2017 and real GDP for 2018 using 2017 prices as the set of common prices, and computing a first measure of the rate of growth of GDP from 2017 to 2018.
   - Constructing real GDP for 2017 and 2018 using 2018 prices as the set of common prices, and computing a second measure of the rate of growth of GDP from 2017 to 2018.
2. Constructing the rate of change of real GDP as the average of these two rates of change.
3. Constructing an index for the level of real GDP by *linking*—or *chaining*—the constructed rates of change for each year. The index is set equal to 1 in some arbitrary year. At the time this book is being written, the arbitrary year is 2012. Given that the constructed rate of change from 2012 to 2013 by the BEA is 1.7%, the index for 2013 equals $(1 + 1.7\%) = 1.017$. The index for 2014 is then obtained by multiplying the index for 2013 by the rate of change from 2013 to 2014, and so on.
4. Multiplying this index by nominal GDP in 2012 to derive *real GDP in chained (2012) dollars*. As the index is 1 in 2012, this implies that real GDP in 2012 equals nominal GDP in 2012.

   Chained refers to the chaining of rates of change described previously. *(2012)* refers to the year where, by construction, real GDP is equal to nominal GDP.

   This index is more complicated to construct than the indexes used before 1995. (To make sure you understand the steps, construct real GDP in chained (year 0) dollars for year 1 in the earlier example.)* But it is clearly better conceptually: The prices used to evaluate real GDP in two adjacent years are the right prices, namely the average prices for those two years. And, because the rate of change from one year to the next is constructed using the prices in those two years rather than the set of prices in an arbitrary base year, history will not be rewritten every five years—as it used to be when, under the previous method for constructing real GDP, the base year was changed every five years.

* What happens to measured real GDP when the base year is changed? What happens to real GDP growth?

(For more details, go to www.bea.gov/scb/pdf/national/nipa/1995/0795od.pdf.)

Key Term

base year, 40

This page is intentionally left blank

The Short Run

In the short run, demand determines output. Many factors affect demand, from consumer confidence to the state of the financial system, to fiscal and monetary policy.

Chapter 3

Chapter 3 looks at equilibrium in the goods market and the determination of output. It focuses on the interaction between demand, production, and income. It shows how fiscal policy affects output.

Chapter 4

Chapter 4 looks at equilibrium in financial markets and the determination of the interest rate. It shows how monetary policy determines the interest rate.

Chapter 5

Chapter 5 looks at the goods market and financial markets together. It shows what determines output and the interest rate in the short run. It looks at the role of fiscal and monetary policy.

Chapter 6

Chapter 6 extends the model by introducing a richer financial system and uses it to explain what happened during the financial crisis.

This page is intentionally left blank

The Goods Market

3

When economists think about year-to-year movements in economic activity, they focus on the interactions among *demand, production, and income*:

- **Changes in the demand for goods lead to changes in production.**
- **Changes in production lead to changes in income.**
- **Changes in income lead to changes in the demand for goods.**

Nothing makes the point better than this cartoon:

PHILIP J., JULIANO/Pearson Education

This chapter looks at these interactions and their implications.

Section 3-1 looks at the composition of GDP and the different sources of the demand for goods.

Section 3-2 looks at the determinants of the demand for goods.

Section 3-3 shows how equilibrium output is determined by the condition that the production of goods must be equal to the demand for goods.

Section 3-4 gives an alternative way of thinking about the equilibrium, based on the equality of investment and saving.

Section 3-5 takes a first pass at the effects of fiscal policy on equilibrium output.

If you remember one basic message from this chapter, it should be: In the short run, demand determines output. ▶▶▶

3-1 THE COMPOSITION OF GDP

The terms *output* and *production* are synonymous. There is no rule for using one or the other. Use the one that sounds better.

The purchase of a machine by a firm, the decision to go to a restaurant by a consumer, and the purchase of combat airplanes by the federal government are clearly different decisions and depend on different factors. So, if we want to understand what determines the demand for goods, it makes sense to decompose aggregate output (GDP) from the point of view of the different goods being produced, and from the point of view of the different buyers for these goods.

The decomposition of GDP (which we shall denote by the letter Y when we use algebra throughout this book) typically used by macroeconomists is shown in Table3-1 (a more detailed version, with precise definitions, appears in Appendix 1 at the end of the book).

Warning! To most people, the term *investment* refers to the purchase of assets like gold or shares of General Motors. Economists use *investment* to refer to the purchase of *new capital goods*, such as (new) machines, (new) buildings, or (new) houses. When economists refer to the purchase of gold, or shares of General Motors, or other financial assets, they use the term *financial investment*.

- First comes **consumption** (**C**). These are the goods and services purchased by consumers, ranging from food to airline tickets, to new cars, and so on. Consumption is by far the largest component of GDP. In 2018, it accounted for 68% of GDP.
- Second comes **investment** (**I**), sometimes called **fixed investment** to distinguish it from inventory investment (which we will discuss later). Investment is the sum of **nonresidential investment**, the purchase by firms of new plants or new machines (from turbines to computers), and **residential investment**, the purchase by people of new houses or apartments.

  Nonresidential investment and residential investment, and the decisions behind them, have more in common than might first appear. Firms buy machines or plants to produce output in the future. People buy houses or apartments to get *housing*

Table 3-1 The Composition of US GDP, 2018

		Billions of Dollars	Percent of GDP
	GDP (*Y*)	20,500	100.0
1	Consumption (*C*)	13,951	68.0
2	Investment (*I*)	3,595	17.5
	Nonresidential	2,800	13.6
	Residential	795	3.8
3	Government spending (*G*)	3,522	17.2
4	Net exports	−625	−3.0
	Exports (*X*)	2,550	12.4
	Imports (*IM*)	−3,156	−15.4
5	Inventory investment	56	0.2

Source: Survey of Current Business, February 2019, Table 1-1-5 *services* in the future. In both cases, the decision to buy depends on the services these goods will yield in the future, so it makes sense to treat them together. Together, nonresidential and residential investment accounted for 17.5% of GDP in 2018.

- Third comes **government spending (*G*)**. This represents the purchases of goods and services by the federal, state, and local governments. The goods range from airplanes to office equipment. The services include services provided by government employees: In effect, the national income accounts treat the government as buying the services provided by government employees—and then providing these services to the public, free of charge.

  Note that *G* does not include **government transfers**, like Medicare or Social Security payments, nor interest payments on the government debt. Although these are clearly government expenditures, they are not purchases of goods and services. That is why the number for government spending on goods and services in Table 3-1, 17.2% of GDP, is smaller than the number for total government spending including transfers and interest payments. That number, in 2018, was approximately 33.0% of GDP when transfers and interest payments of federal, state, and local governments are combined.

- The sum of lines 1, 2, and 3 gives the *purchases of goods and services by US consumers, US firms, and the US government.* To determine the *purchases of US goods and services*, two more steps are needed:

  First, we must add **exports (*X*)**, the purchases of US goods and services by foreigners.

  Second, we must subtract **imports (*IM*)**, the purchases of foreign goods and services by US consumers, US firms, and the US government.

  The difference between exports and imports is called **net exports** $(X - IM)$, or the **trade balance**. If exports exceed imports, the country is said to run a **trade surplus**. If exports are less than imports, the country is said to run a **trade deficit**. In 2018, US exports accounted for 12.4% of GDP. US imports were equal to 15.4% of GDP, so the United States was running a trade deficit equal to 3.0% of GDP.

- So far we have looked at various sources of purchases (sales) of US goods and services in 2018. To determine US production in 2018, we need to take one last step:

  In any given year, production and sales need not be equal. Some of the goods produced in a given year are not sold in that year but in later years. And some of the goods sold in a given year may have been produced in a previous year. The difference between goods produced and goods sold in a given year—the difference between production and sales, in other words—is called **inventory investment**.

If production exceeds sales and firms accumulate inventories as a result, then inventory investment is said to be positive. If production is less than sales and firms' inventories fall, then inventory investment is said to be negative. Inventory investment is typically small—positive in some years and negative in others. In 2018, inventory investment was positive, equal to just $56 billion. Put another way, production was higher than sales by an amount equal to $56 billion.

We now have what we need to develop our first model of output determination.

Exports > imports
⇔ trade surplus
Imports > exports
⇔ trade deficit

Although it is called "inventory investment," the word *investment* is slightly misleading. In contrast to fixed investment, which represents decisions by firms, inventory investment is partly involuntary, reflecting the fact that firms did not anticipate sales accurately in making production plans.

Make sure you understand each of these three equivalent ways of stating the relations among production, sales, and inventory investment:

Inventory Investment = Production − Sales
Production = Sales + Inventory Investment
Sales = Production − Inventory Investment

3-2 THE DEMAND FOR GOODS

Denote the total demand for goods by *Z*. Using the decomposition of GDP we saw in Section 3-1, we can write *Z* as

$$Z \equiv C + I + G + X - IM$$

Recall that inventory investment is not part of demand. ▶

This equation is an **identity** (which is why it is written using the symbol "≡" rather than an equals sign). It *defines* Z as the sum of consumption, plus investment, plus government spending, plus exports, minus imports.

We now need to think about the determinants of Z. To make the task easier, let's first make a number of simplifications:

A model nearly always starts with "Assume" (or "Suppose"). This is an indication that reality is about to be simplified to focus on the issue at hand. ▶

- Assume that all firms produce the same good, which can then be used by consumers for consumption, by firms for investment, or by the government. With this (big) simplification, we need to look at only one market—the market for "the" good—and think about what determines supply and demand in that market.
- Assume that firms are willing to supply any amount of the good at a given price level (P). This assumption allows us to focus on the role demand plays in the determination of output. As we shall see, this assumption is valid only in the short run. When we move to the study of the medium run (starting in Chapter 7), we shall abandon it. But for the moment, it will simplify our discussion.
- Assume that the economy is *closed*—that it does not trade with the rest of the world: Both exports and imports are zero. This assumption clearly goes against the facts: Modern economies trade with the rest of the world. Later on (starting in Chapter 17), we will abandon this assumption as well and look at what happens when the economy is open. But, for the moment, this assumption will also simplify our discussion because we won't have to think about what determines exports and imports.

Under the assumption that the economy is closed, $X = IM = 0$, so the demand for goods Z is simply the sum of consumption, investment, and government spending:

$$Z \equiv C + I + G$$

Let's discuss each of these three components in turn.

Consumption (C)

Consumption decisions depend on many factors. But the main one is surely income, or, more precisely, **disposable income** (Y_D), the income that remains once consumers have received transfers from the government and paid their taxes. When their disposable income goes up, people buy more goods; when it goes down, they buy fewer goods.

We can then write:

$$C = \underset{(+)}{C(Y_D)} \tag{3.1}$$

This is a formal way of stating that consumption C is a function of disposable income Y_D. The function $C(Y_D)$ is called the **consumption function**. The positive sign below Y_D reflects the fact that when disposable income increases, so does consumption. Economists call such an equation a **behavioral equation** to indicate that the equation captures some aspect of behavior—in this case, the behavior of consumers.

We will use functions in this book as a way of representing relations between variables. What you need to know about functions—which is very little—is described in Appendix 2 at the end of the book. This appendix develops the mathematics you need to go through this book. Not to worry: I shall always describe a function in words when I introduce it for the first time.

It is often useful to be more specific about the form of the function. Here is such a case. It is reasonable to assume that the relation between consumption and disposable income is given by the simpler relation:

$$C = c_0 + c_1 Y_D \tag{3.2}$$

In other words, it is reasonable to assume that the function is a **linear relation**. The relation between consumption and disposable income is then characterized by two **parameters**, c_0 and c_1:

Think about your own consumption behavior. What are your values of c_0 and c_1?

- The parameter c_1 is called the **propensity to consume**. (It is also called the *marginal propensity to consume*. I will drop the word *marginal* for simplicity.) It gives the effect an additional dollar of disposable income has on consumption. If c_1 is equal to 0.6, then an additional dollar of disposable income increases consumption by $\$1 \times 0.6 = 60$ cents.

  A natural restriction on c_1 is that it be positive: An increase in disposable income is likely to lead to an increase in consumption. Another natural restriction is that c_1 be less than 1: People are likely to consume only part of any increase in disposable income and save the rest.
- The parameter c_0 has a literal interpretation. It is what people would consume if their disposable income in the current year were equal to zero: If Y_D equals zero in equation (3.2), $C = c_0$. If we use this interpretation, a natural restriction is that, if current income were equal to zero, consumption would still be positive: With or without income, people still need to eat! This implies that c_0 is positive. How can people have positive consumption if their income is equal to zero? Answer: They dissave. They consume either by selling some of their assets or by borrowing.
- The parameter c_0 has a less literal and more frequently used interpretation. Changes in c_0 reflect changes in consumption for a given level of disposable income. Increases in c_0 reflect an increase in consumption given income, decreases in c_0 a decrease. There are many reasons why people may decide to consume more or less, given their disposable income. They may, for example, find it easier or more difficult to borrow, or may become more or less optimistic about the future. An example of a decrease in c_0 is given in the Focus Box, "The Lehman Bankruptcy, Fears of Another Great Depression, and Shifts in the Consumption Function."

The relation between consumption and disposable income shown in equation (3.2) is drawn in Figure 3-1. Because it is a linear relation, it is represented by a straight line. Its intercept with the vertical axis is c_0; its slope is c_1. Because c_1 is less than 1, the slope of the line is less than 1: Equivalently, the line is flatter than a 45-degree line. If the value of c_0 increases, then the line shifts up by the same amount. (A refresher on graphs, slopes, and intercepts is given in Appendix 2 at the end of the book.)

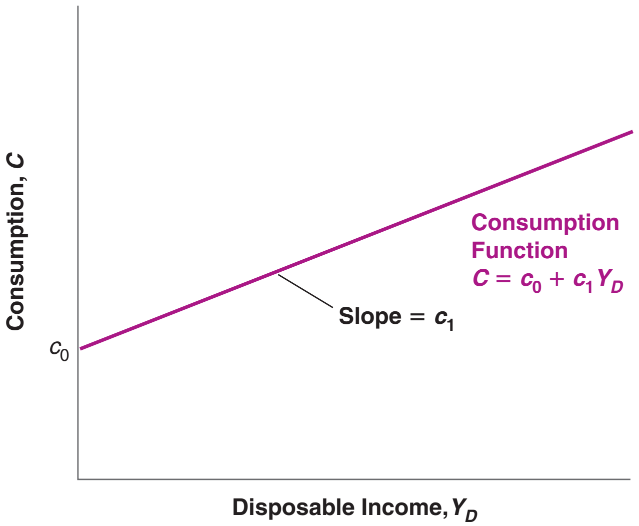


Figure 3-1

Consumption and Disposable Income

Consumption increases with disposable income but less than one for one. A lower value of c_0 will shift the entire line down.

In the United States, the two major taxes paid by individuals are income taxes and Social Security contributions. The main government transfers are Social Security benefits, Medicare (health care for retirees), and Medicaid (health care for the poor). In 2018, taxes and social contributions paid by individuals were \$2,990 billion, and government transfers to individuals were \$3,000 billion.

Next we need to define disposable income Y_D. Disposable income is given by

$$Y_D \equiv Y - T$$

where Y is income and T is taxes paid minus government transfers received by consumers. For short, we will refer to T simply as taxes—but remember that it is equal to taxes minus transfers. Note that the equation is an identity, indicated by " $\equiv$ ".

Replacing Y_D in equation (3.2) gives

$$C = c_0 + c_1(Y - T) \tag{3.3}$$

Equation (3.3) tells us that consumption C is a function of income Y and taxes T. Higher income increases consumption, but less than one for one. Higher taxes decrease consumption, also less than one for one.

Investment (*I*)

Models have two types of variables. Some variables depend on other variables in the model and are therefore explained within the model. Variables like these are called **endogenous variables**. This was the case for consumption given previously. Other variables are not explained within the model but are instead taken as given. Variables like these are called **exogenous variables**. This is how we will treat investment here. We will take investment as given and write:

$$I = \bar{I} \tag{3.4}$$

Endogenous variables: explained within the model. Exogenous variables: taken as given.

Putting a bar on investment is a simple typographical way to remind us that we take investment as given.

We take investment as given to keep our model simple. But the assumption is not innocuous. It implies that, when we later look at the effects of changes in production, we will assume that investment does not respond to changes in production. It is not hard to see that this implication may be a bad description of reality: Firms that experience an increase in production might well decide they need more machines and increase their investment as a result. For now, though, we will leave this mechanism out of the model. In Chapter 5 we will introduce a more realistic treatment of investment.

Government Spending (*G*)

Recall: Taxes means taxes minus government transfers.

The third component of demand in our model is government spending, G. Together with taxes T, G describes **fiscal policy**—the choice of taxes and spending by the government. Just as we did for investment, we will take G and T as exogenous. But the reason why we assume G and T are exogenous is different from the reason we assumed investment is exogenous. It is based on two distinct arguments:

- First, governments do not behave with the same regularity as consumers or firms, so there is no reliable rule we could write for G or T corresponding to the rule we wrote, for example, for consumption. (This argument is not airtight, though. Even if governments do not follow simple behavioral rules as consumers do, a good part of their behavior is predictable. We will look at these issues later, in particular in Chapters 22 and 23. Until then, I shall set them aside.)
- Second, and more importantly, one of the tasks of macroeconomists is to think about the implications of alternative spending and tax decisions. We want to be able to say, "If the government chose these values for G and T, this is what would happen." The approach in this book will typically treat G and T as variables chosen by the government and will not try to explain them within the model.

Because we will (nearly always) take G and T as exogenous, I won't use a bar to denote their values. This will keep the notation lighter.

3-3 THE DETERMINATION OF EQUILIBRIUM OUTPUT

Let's put together the pieces we have introduced so far.

Assuming that exports and imports are both equal to zero, the demand for goods is the sum of consumption, investment, and government spending:

$$Z \equiv C + I + G$$

Replacing C and I from equations (3.3) and (3.4), we get

$$Z = c_0 + c_1(Y - T) + \bar{I} + G \qquad (3.5)$$

The demand for goods Z depends on income Y, taxes T, investment $\bar{I}$, and government spending G.

Let's now turn to **equilibrium** in the goods market, and the relation between production and demand. If firms hold inventories, then production need not be equal to demand: For example, firms can satisfy an increase in demand by drawing upon their inventories—by having negative inventory investment. They can respond to a decrease in demand by continuing to produce and accumulating inventories—by having positive inventory investment. Let's first ignore this complication, though, and begin by assuming that firms do not hold inventories. In this case, inventory investment is always equal to zero, and **equilibrium in the goods market** requires that production Y be equal to the demand for goods Z:

◀ Think of an economy that produces only haircuts. There cannot be inventories of haircuts—haircuts produced but not sold?—so production must always be equal to demand.

$$Y = Z \qquad (3.6)$$

This equation is called an **equilibrium condition**. Models include three types of equations: identities, behavioral equations, and equilibrium conditions. You now have seen examples of each: The equation defining disposable income is an identity, the consumption function is a behavioral equation, and the condition that production equals demand is an equilibrium condition.

Replacing demand Z in (3.6) by its expression from equation (3.5) gives

◀ There are three types of equations:
Identities
Behavioral equations
Equilibrium conditions

$$Y = c_0 + c_1(Y - T) + \bar{I} + G \qquad (3.7)$$

Equation (3.7) represents algebraically what we stated informally at the beginning of this chapter:

In equilibrium, production Y (the left side of the equation), is equal to demand (the right side). Demand in turn depends on income Y, which is itself equal to production.

◀ Can you relate this statement to the cartoon at the start of the chapter?

Note that we are using the same symbol Y for production and income. This is no accident! As you saw in Chapter 2, we can look at GDP either from the production side or from the income side. Production and income are identically equal.

Having constructed a model, we can solve it to look at what determines the level of output—how output changes in response to, say, a change in government spending. Solving a model means not only solving it algebraically but also understanding why the results are what they are. In this book, solving a model will also mean characterizing the results using graphs—sometimes skipping the algebra altogether—and describing the results and the mechanisms in words. Macroeconomists always use these three tools:

1. Algebra to make sure that the logic is correct,
2. Graphs to build the intuition, and
3. Words to explain the results.

Make it a habit to do the same.

Using Algebra

Rewrite the equilibrium equation (3.7):

$$Y = c_0 + c_1 Y - c_1 T + \bar{I} + G$$

Move $c_1 Y$ to the left side and reorganize the right side:

$$(1 - c_1)Y = c_0 + \bar{I} + G - c_1 T$$

Divide both sides by $(1 - c_1)$:

$$Y = \frac{1}{1 - c_1}[c_0 + \bar{I} + G - c_1 T] \qquad (3.8)$$

Autonomous means independent—in this case, independent of output. ►

Equation (3.8) characterizes equilibrium output, i.e. the level of output such that production equals demand. Let's look at both terms on the right, beginning with the term in brackets.

- The term $[c_0 + \bar{I} + G - c_1 T]$ is the part of the demand for goods that does not depend on output. For this reason, it is called **autonomous spending**.

  Can we be sure that autonomous spending is positive? We cannot, but it is very likely to be. The first two terms in brackets, c_0 and $\bar{I}$, are positive. What about the last two, $G - c_1 T$? Suppose the government is running a **balanced budget**—taxes equal government spending. If $T = G$, and the propensity to consume (c_1) is less than 1 (as we have assumed), then $(G - c_1 T)$ is positive and so is autonomous spending. Only if the government were running a very large budget surplus—if taxes were much larger than government spending—could autonomous spending be negative. We can safely ignore that case here.

If $T = G$, then $(G - c_1 T) = (T - c_1 T) = (1 - c_1)T > 0$ ►

- Turn to the first term, $1/(1 - c_1)$. Because the propensity to consume (c_1) is between zero and 1, $1/(1 - c_1)$ is a number greater than one. For this reason, this number, which *multiplies* autonomous spending, is called the **multiplier**. The closer c_1 is to 1, the larger the multiplier.

What does the multiplier imply? Suppose that, for a given level of income, consumers decide to consume more. More concretely, assume that c_0 in equation (3.3) increases by \$1 billion. Equation (3.8) tells us that output will increase by more than \$1 billion. For example, if c_1 equals 0.6, the multiplier equals $1/(1 - 0.6) = 1/0.4 = 2.5$, so that output increases by $2.5 \times \$1\text{billion} = \$2.5\text{billion}$

We have looked at an increase in consumption, but equation (3.8) makes it clear that any change in autonomous spending—from a change in investment, to a change in government spending, to a change in taxes—will have the same qualitative effect: It will change output by more than its direct effect on autonomous spending.

Where does the multiplier effect come from? Looking back at equation (3.7) gives us the clue: An increase in c_0 increases demand. The increase in demand then leads to an increase in production. The increase in production leads to an equivalent increase in income (remember the two are identically equal). The increase in income further increases consumption, which further increases demand, and so on. The best way to describe this mechanism is to represent the equilibrium using a graph. Let's do that.

Using a Graph

Let's characterize the equilibrium graphically.

- First, plot production as a function of income.

  In Figure 3-2, measure production on the vertical axis. Measure income on the horizontal axis. Plotting production as a function of income is straightforward: Recall that production and income are identically equal. Thus, the relation between them is the 45-degree line, the line with a slope equal to 1.
- Second, plot demand as a function of income.

  The relation between demand and income is given by equation (3.5). Let's rewrite it here for convenience, regrouping the terms for autonomous spending together in the term in parentheses:

$$Z = (c_0 + \bar{I} + G - c_1 T) + c_1 Y \tag{3.9}$$

  Demand depends on autonomous spending and on income—via its effect on consumption. The relation between demand and income is drawn as ZZ in the graph. The intercept with the vertical axis—the value of demand when income is equal to zero—equals autonomous spending. The slope of the line is the propensity to consume, c_1: When income increases by 1, demand increases by c_1 Under the restriction that c_1 is positive but less than 1, the line is upward sloping but has a slope of less than 1.
- In equilibrium, production equals demand.

  Equilibrium output, Y, therefore occurs at the intersection of the 45-degree line and the demand function. This is at point A. To the left of A, demand exceeds production; to the right of A, production exceeds demand. Only at A are demand and production equal.

  Suppose that the economy is at the initial equilibrium, represented by point A in the graph, with production equal to Y.

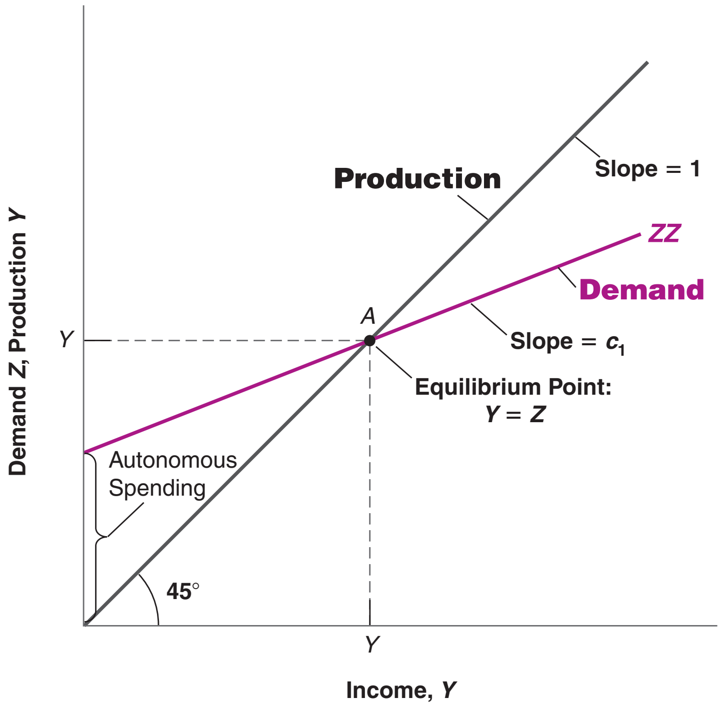


Figure 3-2

Equilibrium in the Goods Market

Equilibrium output is determined by the condition that production is equal to demand.

This makes use of the second interpretation of changes in c_0. How much more consumers are willing to spend at a given level of income.

Now suppose c_0 increases by \$1 billion. At the initial level of income (the level of disposable income associated with point A since T is unchanged in this example), consumers increase their consumption by \$1 billion. What happens is shown in Figure 3-3, which builds on Figure 3-2.

Equation (3.9) tells us that, for any value of income, if c_0 is higher by \$1 billion, demand is higher by \$1 billion. Before the increase in c_0 the relation between demand and income was given by the line ZZ. After the increase in c_0 by \$1 billion, the relation between demand and income is given by the line ZZ', which is parallel to ZZ but higher by \$1 billion. In other words, the demand curve shifts up by \$1 billion. The new equilibrium is at the intersection of the 45-degree line and the new demand relation, at point A'.

Look at the vertical axis. The distance between Y and Y' on the vertical axis is larger than the distance between A and B—which is equal to \$1 billion.

Equilibrium output increases from Y to Y'. The increase in output, $(Y' - Y)$, which we can measure either on the horizontal or the vertical axis, is larger than the initial increase in consumption of \$1 billion. This is the multiplier effect.

With the help of the graph, it becomes easier to tell how and why the economy moves from A to A'. The initial increase in consumption leads to an increase in demand of \$1 billion. At the initial level of income, Y, the level of demand is shown by point B:

Demand is \$1 billion higher. To satisfy this higher level of demand, firms increase production by \$1 billion. This increase in production of \$1 billion implies that income increases by \$1 billion (recall: income = production), so the economy moves to point C. (In other words, both production and income are higher by \$1 billion.) But this is not the end of the story. The increase in income leads to a further increase in demand. Demand is now shown by point D. Point D leads to a higher level of production, and so on, until the economy is at A', where production and demand are again equal. This is therefore the new equilibrium.

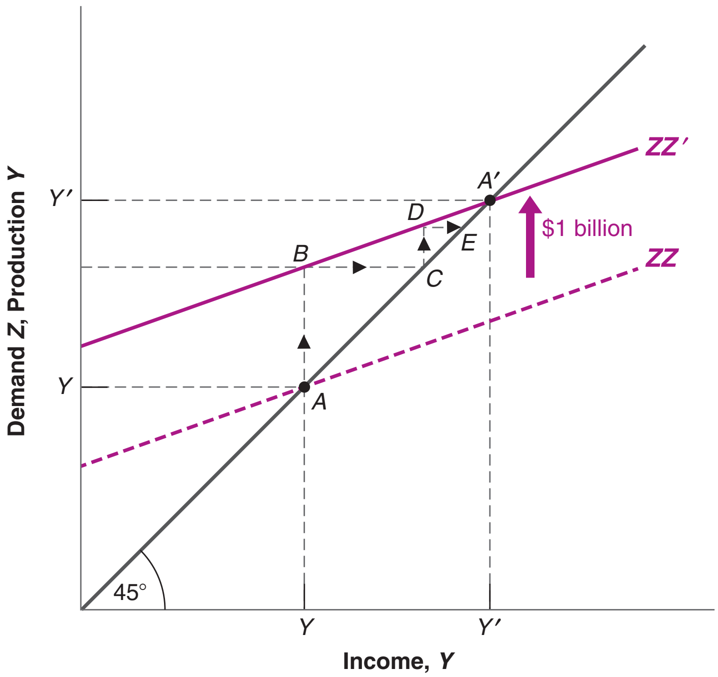
We can pursue this line of explanation a bit more, which will give us another way to think about the multiplier.

- The first-round increase in demand, shown by the distance AB in Figure 3-3—equals \$1 billion.

Figure 3-3

The Effects of an Increase in Autonomous Spending on Output

An increase in autonomous spending has a more than one-for-one effect on equilibrium output.



- This first-round increase in demand leads to an equal increase in production, or \$1 billion, which is also shown by the distance *AB*.
- This first-round increase in production leads to an equal increase in income, shown by the distance *BC*, also equal to \$1 billion.
- The second-round increase in demand, shown by the distance *CD*, equals \$1 billion (the increase in income in the first round) times the propensity to consume, c_1—hence, $\$c_1$ billion.
- This second-round increase in demand leads to an equal increase in production, also shown by the distance *CD*, and thus an equal increase in income, shown by the distance *DE*.
- The third-round increase in demand equals $\$c_1$ billion (the increase in income in the second round) times c_1, the marginal propensity to consume; it is equal to $\$c_1 \times c_1 = \$c_1^2$ billion, and so on.

Following this logic, the total increase in production after, say, $n + 1$ rounds equals \$1 billion times the sum:

$$1 + c_1 + c_1^2 + \cdots + c_1^n$$

Such a sum is called a **geometric series**. Geometric series will frequently appear in this book. A refresher is given in Appendix 2 at the end of the book. One property of geometric series is that, when c_1 is less than one (as it is here) and as n gets larger and larger, the sum keeps increasing but approaches a limit. That limit is $1/(1 - c_1)$, making the eventual increase in output $\$1/(1 - c_1)$ billion.

The expression $1/(1 - c_1)$ should be familiar: It is the multiplier, derived another way. This gives us an equivalent but more intuitive way of thinking about the multiplier. We can think of the original increase in demand as triggering successive increases in production, with each increase in production leading to an increase in income, which leads to an increase in demand, which leads to a further increase in production, and so on. The multiplier is the sum of all these successive increases in production.

Trick question: Think about the multiplier as the result of these successive rounds. What would happen in each successive round if c_1, the propensity to consume, was larger than one?

Using Words

How can we summarize our findings in words?

Production depends on demand, which depends on income, which is itself equal to production. An increase in demand, such as an increase in government spending, leads to an increase in production and a corresponding increase in income. This increase in income leads to a further increase in demand, which leads to a further increase in production, and so on. The end result is an increase in output that is larger than the initial shift in demand, by a factor equal to the multiplier.

The size of the multiplier is directly related to the value of the propensity to consume: The higher the propensity to consume, the higher the multiplier. What is the value of the propensity to consume in the United States today? To answer this question, and more generally to estimate behavioral equations and their parameters, economists use **econometrics**, the set of statistical methods used in economics. To give you a sense of what econometrics is and how it is used, read Appendix 3 at the end of the book. This appendix gives you a quick introduction, along with an application estimating the propensity to consume. A reasonable estimate of the propensity to consume in the United States today is around 0.6 (the regressions in Appendix 3 yield two estimates, 0.5 and 0.8). In other words, an additional dollar of disposable income leads on average to an increase in consumption of 60 cents. This implies that the multiplier is equal to $1/(1 - c_1) = 1/(1 - 0.6) = 2.5$.

The empirical evidence suggests that multipliers are typically smaller than that. This is because the simple model developed in this chapter leaves out a number of important mechanisms, for example, the reaction of monetary policy to changes in spending, or the fact that some of the demand falls on foreign goods. We shall come back to the issue as we go through the book.

How Long Does It Take for Output to Adjust?

Let's return to our example one last time. Suppose that c_0 increases by \$1 billion. We know that output will increase by an amount equal to the multiplier $1/(1 - c_1)$ times \$1 billion. But how long will it take for output to reach this higher value?

In the model we saw previously, we ruled out this possibility by assuming firms did not hold inventories, and so could not rely on drawing down inventories to satisfy an increase in demand.

Under the assumptions we have made so far, the answer is: Right away! In writing the equilibrium condition (3.6), I have assumed that production is always equal to demand. In other words, I have assumed that production responds to demand instantaneously. In writing the consumption function (3.2) as I did, I have assumed that consumption responds to changes in disposable income instantaneously. Under these two assumptions, the economy goes instantaneously from point A to point A' in Figure 3-3: The increase in demand leads to an immediate increase in production, the increase in income associated with the increase in production leads to an immediate increase in demand, and so on. There is nothing wrong in thinking about the adjustment in terms of successive rounds as we did previously, even though the equations indicate that all these rounds happen at once.

This instantaneous adjustment isn't really plausible: A firm that faces an increase in demand might well decide to wait before adjusting its production, meanwhile drawing down its inventories to satisfy demand. A worker who gets a pay raise might not adjust her consumption right away. These delays imply that the adjustment of output will take time.

Formally describing this adjustment of output over time—that is, writing the equations for what economists call the **dynamics** of adjustment, and solving this more complicated model—would be too hard to do here. But it is easy to do it informally in words:

- Suppose, for example, that firms make decisions about their production levels at the beginning of each quarter. Once their decisions are made, production cannot be adjusted for the rest of the quarter. If purchases by consumers are higher than production, firms draw down their inventories to satisfy the purchases. On the other hand, if purchases are lower than production, firms accumulate inventories.
- Now suppose consumers decide to spend more, that they increase c_0. During the quarter in which this happens, demand increases, but production—because we assumed it was set at the beginning of the quarter—doesn't yet change. Therefore, income doesn't change either.
- Having observed an increase in demand, firms are likely to set a higher level of production in the following quarter. This increase in production leads to a corresponding increase in income and a further increase in demand. If purchases still exceed production, firms further increase production in the following quarter, and so on.
- In short, in response to an increase in consumer spending, output does not jump to the new equilibrium, but rather increases over time from Y to Y'.

How long the adjustment takes depends on how fast consumers respond to changes in income and how fast firms respond to changes in sales. If firms adjust their production schedules more frequently in response to past increases in purchases, the adjustment will occur faster.

We will often do in this book what I just did here. After we have looked at changes in equilibrium output, we will then describe informally how the economy moves from one equilibrium to the other. This will not only make the description of what happens in the economy feel more realistic, but it will often reinforce your intuition about why the equilibrium changes.

We have focused in this section on increases in demand. But the mechanism, of course, works both ways: Decreases in demand lead to decreases in output. The recession of 2008–2009 was the result of two of the four components of autonomous spending dropping by a large amount at the same time. Recall that autonomous spending is given by $[c_0 + I + G - c_1 T]$. The Focus Box "The Lehman Bankruptcy, Fears of Another Great Depression, and Shifts in the Consumption Function" shows how, when the crisis started,

FOCUS

The Lehman Bankruptcy, Fears of Another Great Depression, and Shifts in the Consumption Function

Why would consumers decrease consumption if their disposable income has not changed? Or, in terms of equation (3.2), why might c_0 decrease—leading in turn to a decrease in demand, output, and so on?

One of the first reasons that come to mind is that, even if their current income has not changed, consumers start worrying about the future and decide to save more. This is precisely what happened at the start of the crisis, in late 2008 and early 2009. The basic facts are shown in Figure 1 below. The figure plots, from the first quarter of 2008 to the third quarter of 2009, the behavior of three variables: disposable income, total consumption, and consumption of durables—the part of consumption that falls on goods such as cars, computers, and so on (Appendix 1 at the end of the book gives a more precise definition). To make things visually simple, all three variables are normalized to equal 1 in the first quarter of 2008.

Note two things about the figure. First, even though the crisis led to a large fall in GDP, during that period, disposable income did not initially move much. It even increased in the first quarter of 2008. But consumption was unchanged from the first to the second quarter of 2008 and then fell before disposable income fell. It fell by 3 percentage points in 2009 relative to 2008, more than the decrease in disposable income. In terms of Figure 1, the distance between the line for disposable income and the line for consumption increased. Second, during the third and especially the fourth quarters of 2008, the consumption of durables dropped sharply. By the fourth quarter of 2008, it was down 10% relative to the first quarter, before recovering briefly in early 2009 and then decreasing again.

Why did consumption, and especially consumption of durables, decrease at the end of 2008 despite relatively small changes in disposable income? Several factors were at play, but the main one was the psychological fallout of the financial crisis. Recall from Chapter 1 that, on September 15, 2008, Lehman Brothers, a very large bank, went bankrupt, and that, in the ensuing weeks, it appeared that many more banks might follow suit and the financial system might collapse. For most people, the main sign of trouble was what they read in newspapers: Even though they still had their job and received their monthly income checks, the events reminded them of the stories of the Great Depression and the pain that came with it. One way to see this is to look at the Google Trends series that gives the number of searches for "Great Depression," from January 2008 to September 2009, and is plotted in Figure 2. The series is normalized so its average value is 1 over the two years. Note how sharply the series

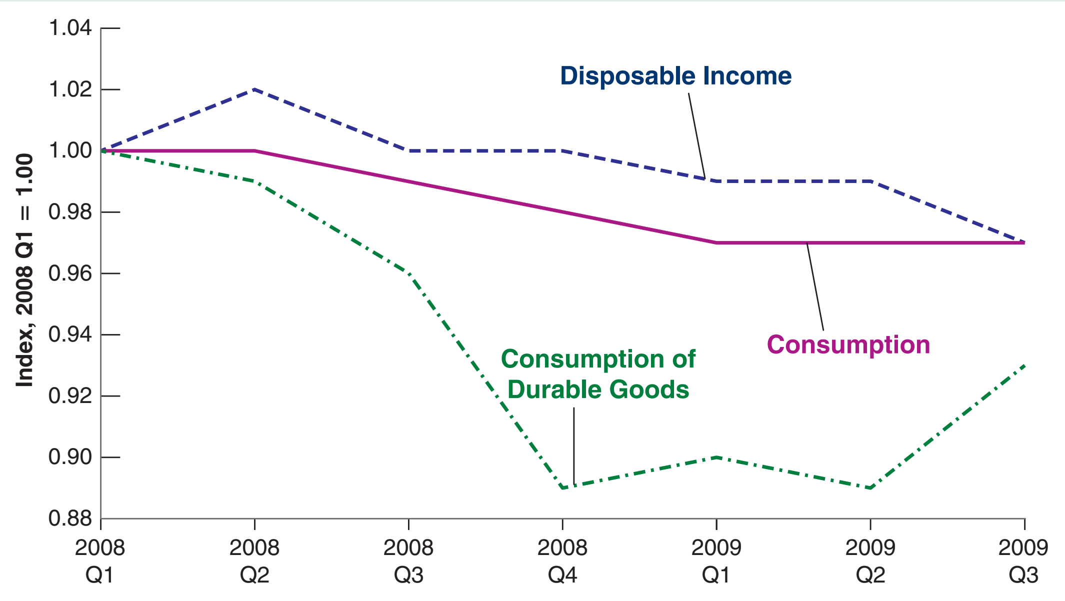


Figure 1

Disposable Income, Consumption, and Consumption of Durables in the United States, 2008:1 to 2009:3

Source: FRED: DPIC96, PCECC96, PCDGCC96.

peaked in October 2008 and then slowly decreased over the course of 2009 as it became clear that, while the crisis was a serious one, policy makers were going to do whatever they could to avoid a repeat of the Great Depression.

If you felt that the economy might go into another Great Depression, what would you do? Worried that you might become unemployed or that your income might decline in the future, you would probably cut consumption, even if your disposable income had not yet changed. And, given the uncertainty about what was going on, you might also delay the purchases you could afford to delay; for example, the purchase of a new car or a new TV. As Figure 1 in this box shows, this is exactly what consumers did in late 2008: Total consumption decreased, and consumption of durables collapsed. In 2009, as the smoke slowly cleared, and the worst scenarios became increasingly unlikely, consumption of durables picked up. But by then, many other factors were contributing to the crisis.

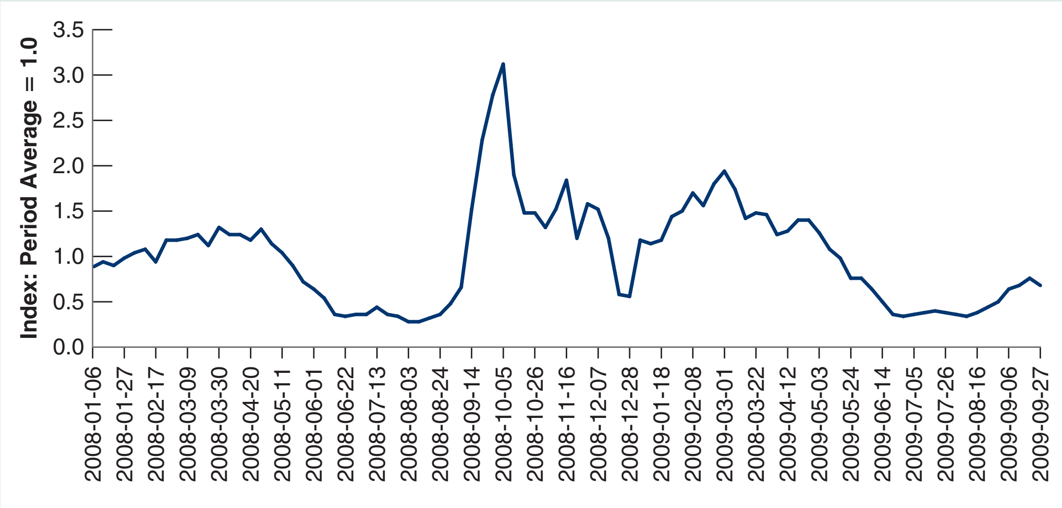


Figure 2

Google Search Volume for "Great Depression," January 2008 to September 2009

Source: Google Trends, "Great Depression."

worries about the future led consumers to cut their spending despite the fact that their disposable income had not yet declined; that is, c_0 decreased sharply. As house prices fell, building new homes became much less desirable. New homes are part of autonomous investment spending, so I also fell sharply. As autonomous spending decreased, the total demand for goods fell, and so did output. We shall return at many points in the book to the factors and the mechanisms behind the crisis and steadily enrich our story line. But this effect on autonomous spending will remain a central element of the story.

3-4 INVESTMENT EQUALS SAVING: AN ALTERNATIVE WAY OF THINKING ABOUT GOODS-MARKET EQUILIBRIUM

Thus far, we have been thinking of equilibrium in the goods market in terms of the equality of the production and the demand for goods. An alternative—but, it turns out, equivalent—way of thinking about equilibrium focuses instead on investment and saving. This is how John Maynard Keynes first articulated this model in 1936, in *The General Theory of Employment, Interest and Money.*

Let's start by looking at saving. Saving is the sum of private saving and public saving.

- By definition, **private saving** (S) (i.e., saving by consumers) is equal to their disposable income minus their consumption:

$$S \equiv Y_D - C$$

Using the definition of disposable income, we can rewrite private saving as income minus taxes minus consumption:

$$S \equiv Y - T - C$$

Private saving is also done by firms, which do not distribute all of their profits and use those retained earnings to finance investment. For simplicity, we ignore saving by firms here. But the bottom line, namely the equality of investment and saving in equation (3.10), does not depend on this simplification.

- By definition, **public saving** ($T - G$) is equal to taxes (net of transfers) minus government spending. If taxes exceed government spending, the government is running a **budget surplus**, so public saving is positive. If taxes are less than government spending, the government is running a **budget deficit**, so public saving is negative.

Public saving ⇔ Budget surplus

- Now return to the equation for equilibrium in the goods market that we derived previously. Production must be equal to demand, which, in turn, is the sum of consumption, investment, and government spending:

$$Y = C + I + G$$

Subtract taxes (T) from both sides and move consumption to the left side:

$$Y - T - C = I + G - T$$

The left side of this equation is simply private saving (S), so

$$S = I + G - T$$

Or, equivalently,

$$I = S + (T - G) \tag{3.10}$$

On the left is investment. On the right is saving, the sum of *private saving* and *public saving*.

Equation (3.10) gives us another way of thinking about equilibrium in the goods market: It says that equilibrium in the goods market requires that investment equal **saving**—the sum of private and public saving. This way of looking at equilibrium explains why the equilibrium condition for the goods market is called the ***IS* relation**, which stands for "**I**nvestment equals **S**aving": What firms want to invest must be equal to what people and the government want to save.

To understand equation (3.10), imagine an economy with only one person who has to decide how much to consume, invest, and save—a "Robinson Crusoe" economy, for example. For Robinson Crusoe, the saving and the investment decisions are one and the same: What he invests (say, by keeping rabbits for breeding rather than having them for dinner), he automatically saves. In a modern economy, however, investment decisions are made by firms, whereas saving decisions are made by consumers and the government. In equilibrium, equation (3.10) tells us, all these decisions have to be consistent: Investment must equal saving.

To summarize: There are two equivalent ways of stating the condition for equilibrium in the goods market:

$$\text{Production} = \text{Demand}$$
$$\text{Investment} = \text{Saving}$$

We characterized the equilibrium using the first condition, equation (3.6). We now do the same using the second condition, equation (3.10). The results will be the same, but the derivation will give you another way of thinking about the equilibrium.

- Note first that *consumption and saving decisions are one and the same:* Given their disposable income, once consumers have chosen consumption, their saving is determined, and vice versa. The way we specified consumption behavior implies that private saving is given by:

$$S = Y - T - C$$
$$= Y - T - c_0 - c_1(Y - T)$$

Rearranging, we get

$$S = -c_0 + (1 - c_1)(Y - T) \tag{3.11}$$

- In the same way that we called c_1 the propensity to consume, we can call $(1 - c_1)$ the **propensity to save**. The propensity to save tells us how much of an additional unit of income people save. The assumption we made previously—that the propensity to consume (c_1) is between zero and one implies that the propensity to save $(1 - c_1)$ is also between zero and one. Private saving increases with disposable income, but by less than one dollar for each additional dollar of disposable income.

In equilibrium, investment must be equal to saving, the sum of private and public saving. Replacing private saving in equation (3.10) by its expression in equation (3.11),

$$I = -c_0 + (1 - c_1)(Y - T) + (T - G)$$

Solving for output,

$$Y = \frac{1}{1 - c_1}[c_0 + \bar{I} + G - c_1 T] \tag{3.12}$$

Equation (3.12) is exactly the same as equation (3.8). This should come as no surprise. We are looking at the same equilibrium condition, just in a different way. This alternative way will prove useful in various applications later in the book. The Focus Box "The Paradox of Saving" looks at such an application, which was first emphasized by Keynes and is often called the paradox of saving.

3-5 IS THE GOVERNMENT OMNIPOTENT? A WARNING

Equation (3.8) implies that the government, by choosing the level of spending (G) or the level of taxes (T) can choose the level of output it wants. If it wants output to be higher by, say, \$1 billion, all it needs to do is to increase G by $\$(1 - c_1)$ billion. This increase in government spending, in theory, will lead to an output increase of $\$(1 - c_1)$ billion times the multiplier $1/(1 - c_1)$, or \$1 billion.

Can governments really achieve the level of output they want? Obviously not: If they could, and it was as easy as it sounds in the previous paragraph, why would the US government have allowed growth to stall in 2008 and output to actually fall in 2009? Why wouldn't the government increase the growth rate now, so as to decrease unemployment more rapidly? There are many aspects of reality that we have not yet incorporated in our model, and all of them complicate the government's task. We shall introduce them in due time. But it is useful to list them briefly here:

For a glimpse at the longer list, go to Section 22-1, "What You Have Learned," in Chapter 22.

- Changing government spending or taxes is not easy. Getting the US Congress to pass bills always takes time. Implementing what is in the bills takes even more time (Chapters 21 and 22).

FOCUS

The Paradox of Saving

As we grow up, we are told about the virtues of thrift. Those who spend all their income are condemned to end up poor. Those who save are promised a happy life. Similarly, governments tell us, an economy that saves is an economy that will grow strong and prosper! The model we have seen in this chapter, however, tells a different and surprising story.

Suppose that, at a given level of disposable income, consumers decide to save more. In other words, suppose consumers decrease c_0, therefore decreasing consumption and increasing saving at a given level of disposable income. What happens to output and to saving?

Equation (3.12) makes it clear that equilibrium output decreases: As people save more at their initial level of income, they decrease their consumption. But this decreased consumption decreases demand, which decreases production.

Can we tell what happens to saving? Let's return to the equation for private saving, equation (3.11) (recall that we assume no change in public saving, so saving and private saving move together):

$$S = -c_0 + (1 - c_1)(Y - T)$$

On the one hand, $-\mathbf{c_0}$ is higher (less negative): Consumers are saving more at any level of income; this tends to increase saving. But, on the other hand, their income Y is lower (because production is lower): This decreases saving. The net effect would seem to be ambiguous. In fact, we can tell which way it goes. To see how, go back to equation (3.10), the equilibrium condition that investment and saving must be equal:

$$I = S + (T - G)$$

By assumption, investment does not change: $I = \bar{I}$. Nor do T or G. So the equilibrium condition tells us that in equilibrium, private saving S cannot change either. Although people want to save more at a given level of income, their income decreases by an amount such that their saving is unchanged.

This means that as people attempt to save more, the result is both a decline in output and unchanged saving, not a good outcome. This surprising pair of results is known as the *paradox of saving* (or the *paradox of thrift*). Note that the same result would obtain if we looked at public rather than private saving: A decrease in the budget deficit would also lead to a lower output and unchanged overall (public and private) saving. And note that, if we extended our model to allow investment to decrease with output (we shall do this in Chapter 5) rather than assuming it is constant, the result would be even more dramatic: An attempt to save more, either by consumers or by the government, would lead to lower output, lower investment, and by implication lower saving!

So should you forget the old wisdom? Should the government tell people to be less thrifty? No. The results of this simple model are of much relevance in the short run. The desire of consumers to save more is an important factor in many of the US recessions, including, as we saw in the previous Focus box, the financial crisis. But—as we will see later when we look at the medium run and the long run—other mechanisms come into play over time, and an increase in the saving rate is likely to lead over time to higher saving and higher income. A warning remains, however: Policies that encourage saving might be good in the medium run and in the long run, but they can lead to a reduction in demand and in output, and perhaps even a recession, in the short run.

- We have assumed that investment remained constant. But investment is also likely to respond in a variety of ways. So are imports: Some of the increased demand by consumers and firms will not be for domestic goods but for foreign goods. The exchange rate may change. All these responses are likely to be associated with complex, dynamic effects, making it hard for governments to assess the effects of their policies with much certainty (Chapters 5 and 9, and 18 to 20).
- Expectations are likely to matter. For example, the reaction of consumers to a tax cut is likely to depend on whether they think of the tax cut as transitory or permanent. The more they perceive the tax cut as permanent, the larger will be their consumption response. Similarly, the reaction of consumers to an increase in spending is likely to depend on when they think the government will raise taxes to pay for the spending (Chapters 14 to 16).
- Achieving a given level of output can come with unpleasant side effects. Trying to achieve too high a level of output can, for example, lead to increasing inflation and, for that reason, be unsustainable in the medium run (Chapter 9).
- Cutting taxes or increasing government spending, as attractive as it may seem in the short run, can lead to large budget deficits and an accumulation of public debt. A large debt has adverse effects in the long run. This is a hot issue today in almost every advanced country in the world (Chapters 9, 11, 16, and 22).

In short, the proposition that, by using fiscal policy, the government can affect demand and output in the short run is an important and correct proposition. But as we refine our analysis, we will see that the role of the government in general, and the successful use of fiscal policy in particular, become increasingly difficult: Governments will never have it so good as they have had it in this chapter.

SUMMARY

What you should remember about the components of GDP:

- GDP is the sum of consumption, investment, government spending, inventory investment, and exports minus imports.
- Consumption (C) is the purchase of goods and services by consumers. Consumption is the largest component of demand.
- Investment (I) is the sum of nonresidential investment—the purchase of new plants and new machines by firms—and of residential investment—the purchase of new houses or apartments by people.
- Government spending (G) is the purchase of goods and services by federal, state, and local governments.
- Exports (X) are purchases of US goods by foreigners. Imports (IM) are purchases of foreign goods by US consumers, US firms, and the US government.
- Inventory investment is the difference between production and purchases. It can be positive or negative.

What you should remember about our first model of output determination:

- In the short run, demand determines production. Production is equal to income. Income in turn affects demand.
- The consumption function shows how consumption depends on disposable income. The propensity to consume describes how much consumption increases for a given increase in disposable income.
- Equilibrium output is the level of output at which production equals demand. In equilibrium, output equals autonomous spending times the multiplier. Autonomous spending is that part of demand that does not depend on income. The multiplier is equal to $1/(1 - c_1)$, where c_1 is the propensity to consume.
- Increases in consumer confidence, investment demand, government spending, or decreases in taxes all increase equilibrium output in the short run.
- An alternative way of stating the goods-market equilibrium condition is that investment must be equal to saving—the sum of private and public saving. For this reason, the equilibrium condition is called the *IS* relation (*I* for investment, *S* for saving).

KEY TERMS

consumption (C), 46
investment (I), 46
fixed investment, 46
nonresidential investment, 46
residential investment, 46
government spending (G), 47
government transfers, 47
exports (X), 47
imports (IM), 47
net exports ($X - IM$), 47
trade balance, 47
trade surplus, 47
trade deficit, 47
inventory investment, 47
identity, 48
disposable income (Y_D), 48
consumption function, 48
behavioral equation, 48
linear relation, 49
parameter, 49
propensity to consume (c_1), 49
endogenous variables, 50
exogenous variables, 50
fiscal policy, 50
equilibrium, 51
equilibrium in the goods market, 51
equilibrium condition, 51
autonomous spending, 52
balanced budget, 52
multiplier, 52
geometric series, 55
econometrics, 55
dynamics, 56
private saving (S), 59
public saving ($T - G$), 59
budget surplus, 59
budget deficit, 59
saving, 59
IS relation, 59
propensity to save, 60
paradox of saving, 61

QUESTIONS AND PROBLEMS

QUICK CHECK

1. *Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.*
   a. The largest component of GDP is consumption.
   b. Government spending, including transfers, was equal to 17.4% of GDP in 2018.
   c. The propensity to consume has to be positive, but otherwise it can take on any positive value.
   d. One factor in the 2009 recession was a drop in the value of the parameter c_0.
   e. Fiscal policy describes the choice of government spending and taxes and is treated as exogenous in our goods market model.
   f. The equilibrium condition for the goods market states that consumption equals output.
   g. An increase of one unit in government spending leads to an increase of one unit in equilibrium output.
   h. An increase in the propensity to consume leads to a decrease in output.

2. *Suppose that the economy is characterized by the following behavioral equations:*

$$C = 160 + 0.6Y_D$$
$$I = 150$$
$$G = 150$$
$$T = 100$$

   Solve for the following variables.
   a. Equilibrium GDP (Y)
   b. Disposable income (Y_D)
   c. Consumption spending (C)

3. *Use the economy described in Problem 2.*
   a. Solve for equilibrium output. Compute total demand. Is it equal to production? Explain.
   b. Assume that G is now equal to 110. Solve for equilibrium output. Compute total demand. Is it equal to production? Explain.
   c. Assume that G is equal to 110, so output is given by your answer to part b. Compute private plus public saving. Is the sum of private and public saving equal to investment? Explain.

DIG DEEPER

4. *The balanced budget multiplier*

   For both political and macroeconomic reasons, governments are often reluctant to run budget deficits. Here, we examine whether policy changes in G and T that maintain a balanced budget are macroeconomically neutral. Put another way, we examine whether it is possible to affect output through changes in G and T so that the government budget remains balanced.

   Start from equation (3.8).
   a. By how much does Y increase when G increases by one unit?
   b. By how much does Y decrease when T increases by one unit?
   c. Why are your answers to parts a and b different?

   Suppose that the economy starts with a balanced budget: G = T. If the increase in G is equal to the increase in T, then the budget remains in balance. Let us now compute the balanced budget multiplier.
   d. Suppose that G and T increase by one unit each. Using your answers to parts a and b, what is the change in equilibrium GDP? Are balanced budget changes in G and T macroeconomically neutral?
   e. How does the specific value of the propensity to consume affect your answer to part a? Why?

5. *Automatic stabilizers*

   In this chapter we have assumed that the fiscal policy variables G and T are independent of the level of income. In the real world, however, this is not the case. Taxes typically depend on the level of income and so tend to be higher when income is higher. In this problem, we examine how this automatic response of taxes can help reduce the impact of changes in autonomous spending on output.

   Consider the following behavioral equations:

$$C = c_0 + c_1 Y_D$$
$$T = t_0 + t_1 Y$$
$$Y_D = Y - T$$

   G and I are both constant. Assume that t_1 is between 0 and 1.
   a. Solve for equilibrium output.
   b. What is the multiplier? Does the economy respond more to changes in autonomous spending when t_1 is 0 or when t_1 is positive? Explain.
   c. Why is fiscal policy in this case called an automatic stabilizer?

6. *Balanced budget versus automatic stabilizers*

   It is often argued that a balanced budget amendment would actually be destabilizing. To understand this argument, consider the economy in Problem 5.
   a. Solve for equilibrium output.
   b. Solve for taxes in equilibrium.

   Suppose that the government starts with a balanced budget and that there is a drop in c_0.
   c. What happens to Y? What happens to taxes?
   d. Suppose that the government cuts spending in order to keep the budget balanced. What will be the effect on Y? Does the cut in spending required to balance the budget counteract or reinforce the effect of the drop in c_0 on output? (Don't do the algebra. Use your intuition and give the answer in words.)

7. *Taxes and transfers*

   Recall that we define taxes, T, as net of transfers. In other words,

$$T = \textit{Taxes} - \textit{Transfer Payments}$$

   a. Suppose that the government increases transfer payments to private households, but these transfer payments are not financed by tax increases. Instead, the government borrows to pay for the transfer payments. Show in a diagram (similar to Figure 3-2) how this policy affects equilibrium output. Explain.

b. Suppose instead that the government pays for the increase in transfer payments with an equivalent increase in taxes. How does the increase in transfer payments affect equilibrium output in this case?

c. Now suppose that the population includes two kinds of people: those with high propensity to consume and those with low propensity to consume. Suppose the transfer policy increases taxes on those with low propensity to consume to pay for transfers to people with high propensity to consume. How does this policy affect equilibrium output?

d. How do you think the propensity to consume might vary across individuals according to income? In other words, how do you think the propensity to consume compares for people with high income and people with low income? Explain. Given your answer, do you think tax cuts will be more effective at stimulating output when they are directed toward high-income or toward low-income taxpayers?

8. Investment and income

This problem examines the implications of allowing investment to depend on output. Chapter 5 carries this analysis much further and introduces an essential relation—the effect of the interest rate on investment—not examined in this problem.

a. Suppose the economy is characterized by the following behavioral equations:

$$C = c_0 + c_1 Y_D$$
$$Y_D = Y - T$$
$$I = b_0 + b_1 Y$$

Government spending and taxes are constant. Note that investment now increases with output. (Chapter 5 discusses the reasons for this relation.) Solve for equilibrium output.

b. What is the value of the multiplier? How does the relation between investment and output affect the value of the multiplier? For the multiplier to be positive, what condition must $(c_1 + b_1)$ satisfy? Explain your answers.

c. What would happen if $(c_1 + b_1) > 1$? (Trick question. Think about what happens in each round of spending).

d. Suppose that the parameter b_0, sometimes called *business confidence*, increases. How will equilibrium output be affected? Will investment change by more or less than the change in b_0? Why? What will happen to national saving?

EXPLORE FURTHER

9. The paradox of saving revisited

You should be able to complete this question without doing any algebra, although you may find making a diagram helpful for part a. For this problem, you do not need to calculate the magnitudes of changes in economic variables—only the direction of change.

a. Consider the economy described in Problem 8. Suppose that consumers decide to consume less (and therefore to save more) for any given amount of disposable income. Specifically, assume that consumer confidence (c_0) falls. What will happen to output?

b. As a result of the effect on output you determined in part a, what will happen to investment? What will happen to public saving? What will happen to private saving? Explain. (*Hint*: Consider the saving-equals-investment characterization of equilibrium.) What is the effect on consumption?

c. Suppose that consumers had decided to increase consumption expenditure, so that c_0 had increased. What would have been the effect on output, investment, and private saving in this case? Explain. What would have been the effect on consumption?

d. Comment on the following logic: "When output is too low, what is needed is an increase in demand for goods and services. Investment is one component of demand, and saving equals investment. Therefore, if the government could just convince households to attempt to save more, then investment, and output, would increase."

Output is not the only variable that affects investment. As we develop our model of the economy, we will revisit the paradox of saving in future chapter problems.

10. Using fiscal policy to avoid the recession of 2009.

GDP in 2009 was roughly $15,000 billion. In 2009, GDP fell by approximately 3 percentage points in 2009.

a. How many billion dollars is 3 percentage points of $15,000 billion?

b. If the propensity to consume were 0.5, by how much would government spending have to have increased to prevent a decrease in output?

c. If the propensity to consume were 0.5, by how much would taxes have been cut to prevent any decrease in output?

d. Suppose Congress had chosen to both increase government spending and raise taxes by the same amount in 2009. What increase in government spending and taxes would have been required to prevent the decline in output in 2009?

11. The "exit strategy" problem

In fighting the recession associated with the crisis, taxes were cut and government spending was increased. The result was a large government deficit. To reduce that deficit, taxes must be increased or government spending must be cut. This is the "exit strategy" from the large deficit.

a. How will reducing the deficit in either way affect the equilibrium level of output in the short run?

b. Which will change equilibrium output more: (i) cutting G by $100 billion (ii) raising T by $100 billion?

c. How does your answer to part b depend on the value of the marginal propensity to consume?

d. You hear the argument that a reduction in the deficit will increase consumer and business confidence and thus reduce the decline in output that would otherwise occur with deficit reduction. Is this argument valid?

12. Fiscal policy and the boom of 2018 in the simplest model

The Tax Cut and Jobs Act was passed by Congress in December 2017. GDP grew from $18,000 billion (2012 dollars) in 2017 to $18,500 billion (2012 dollars) in 2018.

a. By what percentage did real GDP grow from 2017 to 2018?

b. Estimates from the Congressional Budget Office suggest the tax cut in 2018 associated with the Tax Cut and Jobs Act was approximately $150 B (measured in current dollars). The GDP deflator in 2018 (2012=100) was 110. How large was the tax cut measured in 2012 dollars?

c. If the marginal propensity to consume is 0.6, how large was the increase in real GDP attributable to the tax cut?

d. What proportion of real GDP growth from 2017 to 2018 would be accounted for by the tax cut with a marginal propensity to consume of 0.6?

e. You hear the argument that a tax cut will increase consumer and business confidence and cause a larger than usual increase in demand. If normal growth in US real GDP is 2%, does the effect of the tax cut in 2018 support this argument? How does your answer change if normal growth in US real GDP is 1%? We saw in Chapter 1 that normal growth in labor productivity has been closer to 1% in recent years.

This page is intentionally left blank

Financial Markets I

4

Financial markets are intimidating. They involve a maze of players, from banks, to money market funds, mutual funds, investment funds, and hedge funds. Trading involves bonds, stocks, and other financial claims with exotic names, such as swaps and options. The financial pages of newspapers quote interest rates on many government bonds, on many corporate bonds, on short-term bonds, on long-term bonds, and it is easy to get confused. But financial markets play an essential role in the economy. They determine the cost of funds for firms, for households, and for the government, and in turn affect their spending decisions. To understand their role we must proceed in steps.

In this chapter, we focus on the role of the central bank in affecting these interest rates. To do so, we drastically simplify reality and think of the economy as having only *two* financial assets, namely money, which does not pay interest, and bonds, which do. This will allow us to understand how the interest rate on bonds is determined, and the role of the central bank (in the United States, the **Fed,** short for **Federal Reserve Bank**) in this determination.

In the next chapter, Chapter 5, we shall combine the model of the goods market we developed in the previous chapter with the model of financial markets we develop in this chapter, and have another look at equilibrium output. Having done so, we shall return to financial markets in Chapter 6, allowing for more financial assets and more interest rates, and exploring further the role of banks and other financial institutions. This will give us a richer model and allow us to better understand what happened in the Great Financial Crisis.

The chapter has four sections:

Section 4-1 looks at the demand for money.

Section 4-2 assumes that the central bank directly controls the supply of money and shows how the interest rate is determined by the condition that the demand for money be equal to the supply of money.

Section 4-3 introduces banks as suppliers of money, revisits the determination of the interest rate, and describes the role of the central bank in that context.

Section 4-4 looks at the case of the liquidity trap, the constraint on monetary policy coming from the fact that the interest rate on bonds cannot be negative, a constraint that played an important role in the crisis.

If you remember one basic message from this chapter, it should be: In the short run, the central bank determines the interest rate. ▶▶▶

4-1 THE DEMAND FOR MONEY

This section looks at the determinants of *the demand for money.* A warning before we start: Words such as *money* or *wealth* have specific meanings in economics, often not the same meanings as in everyday conversation. The purpose of the Focus Box "Semantic Traps: Money, Income, and Wealth" is to help you avoid some of these traps. Read it carefully and refer back to it once in a while.

Make sure you see the difference between the decision about how much to save (a decision that determines how your wealth changes over time) and the decision about how to allocate a given stock of wealth between money and bonds.

Suppose, as a result of having steadily saved part of your income in the past, your financial wealth today is $50,000. You may intend to keep saving in the future and increase your wealth further, but its value today is given. Suppose also that you have the choice between only two assets, money and bonds:

- *Money*, which you can use for transactions, pays no interest. In the real world, there are two types of money: **currency**, coins and bills, and **checkable deposits**, the bank deposits on which you can write checks or use a debit card. The distinction between the two will be important when we look at the supply of money. For the moment, however, the distinction does not matter and we can ignore it. Just think *currency*.
- **Bonds** pay a positive interest rate, i, but they cannot be used for transactions. In the real world, there are many types of bonds and other financial assets, each associated with a specific interest rate. For the time being, we also ignore this aspect of reality and assume that there is just one type of bond and that it pays the rate of interest i.

Assume that buying or selling bonds implies some cost; for example, a phone call to your broker and the payment of a transaction fee. How much of your $50,000 should you hold in money, and how much in bonds? On the one hand, holding all your wealth in the form of money is clearly very convenient. You won't ever need to call a broker or pay transaction fees. But it also means you will receive no interest income. On the other hand, if you hold all your wealth in the form of bonds, you will earn interest on the full amount, but you will have to call your broker frequently—whenever you need money to take the subway, pay for a cup of coffee, and so on. This is a rather inconvenient way of going through life.

You may want to pay by credit card and avoid carrying currency. But you still must have money in your checking account when you pay the credit card company.

Therefore, it is clear that you should hold both money and bonds. But in what proportions? This will depend mainly on two variables:

- Your *level of transactions.* You will want to have enough money on hand to avoid having to sell bonds whenever you need money. Say, for example, that you typically spend $3,000 a month. In this case, you might want to have, on average, say, two months' worth of spending on hand, or $6,000 in money, and the rest, $50,000 − $6,000 = $44,000, in bonds. If, instead, you typically spend $4,000 a month, you might want to have, say, $8,000 in money and only $42,000 in bonds.
- The *interest rate on bonds*. The only reason to hold any of your wealth in bonds is that they pay interest. The higher the interest rate, the more you will be willing to deal with the hassle and costs associated with buying and selling bonds. If the interest rate is very high, you might even decide to squeeze your money holdings to an average of only two weeks' worth of spending, or $1,500 (assuming your monthly spending is $3,000). This way, you will be able to keep, on average, $48,500 in bonds and earn more interest as a result.

Let's make this last point more concrete. Most of you probably do not hold bonds; my guess is that few of you have a broker. However, some of you hold bonds indirectly if you have a money market account with a financial institution. **Money market funds** (the full name is *money market mutual funds*) pool together the funds of many people. The

Semantic Traps: Money, Income, and Wealth

FOCUS

In everyday conversation, we use "money" to denote many different things. We use it as a synonym for income: "making money." We use it as a synonym for wealth: "She has a lot of money." In economics, you must be more careful. Here is a basic guide to some terms and their precise meanings in economics.

Money is what can be used to pay for transactions. Money is currency and checkable deposits at banks.

Income is what you earn from working plus what you receive in interest and dividends. It is a **flow**—something expressed in units of time: weekly income, monthly income, or yearly income, for example. It is said that J. Paul Getty was once asked what his income was. Getty answered: "$1,000." He meant, but did not say: $1,000 per minute!

Saving is that part of after-tax income that you do not spend. It is also a flow. If you save 10% of your income, and your income is $3,000 per month, then you save $300 per month. **Savings** (plural) is sometimes used as a synonym for wealth—the value of what you have accumulated over time. To avoid confusion, I shall not use the term *savings* in this book.

Your **financial wealth**, or wealth for short, is the value of all your financial assets minus all your financial liabilities. In contrast to income or saving, which are flow variables, financial wealth is a **stock** variable. It is the value of wealth at a given moment in time.

At a given moment in time, you cannot change the total amount of your financial wealth. It can only change over time as you save or dissave, or as the value of your assets and liabilities change. But, at any time, you can change the composition of your wealth; you can, for example, decide to repay part of your mortgage by writing a check against your checking account. This leads to a decrease in your liabilities (a smaller mortgage) and a corresponding decrease in your assets (a smaller checking account balance); but, at that moment, it does not change your wealth.

Financial assets that can be used directly to buy goods are called *money*. Someone who is wealthy might have only small money holdings—say, $1,000,000 in stocks but only $500 in a checking account. It is also possible for a person to have a large income but only small money holdings—say, a monthly income of $10,000 but only $1,000 in the checking account.

Investment is a term economists reserve for the purchase of new capital goods, from machines to plants to office buildings. When you want to talk about the purchase of shares or other financial assets, you should refer to them as a **financial investment**.

Learn how to be economically correct:

Do not say "Mary is making a lot of money"; say "Mary has a high income."

Do not say "Joe has a lot of money"; say "Joe is very wealthy."

funds are then used to buy bonds—typically government bonds. Money market funds pay an interest rate close to but slightly below the interest rate on the bonds they hold—the difference coming from the administrative costs of running the funds and from their profit margins.

When the interest rate on these funds reached 14% per year in the early 1980s (a very high interest rate by today's standards), people who had previously kept all of their wealth in their checking accounts (which paid little or no interest) realized how much interest they could earn by moving some of it into money market accounts instead. Now that interest rates are much lower, people are less careful about putting as much as they can in money market funds. Put another way, for a given level of transactions, people now keep more of their wealth in money than they did in the early 1980s.

Deriving the Demand for Money

Let's go from this discussion to an equation describing the demand for money.

Denote the amount of money people and firms want to hold—their *demand for money*—by M^d (the superscript d stands for *demand*). The demand for money in the economy as a whole is just the sum of all the individual demands for money by the people and firms in the economy. Therefore, it depends on the overall level of transactions in the economy and on the interest rate. The overall level of transactions in the economy is hard to measure, but it is likely to be roughly proportional to nominal income (income measured in dollars): If nominal income were to increase by 10%, it is reasonable to think that the dollar value of

Revisit Chapter 2's example of an economy composed of a steel company and a car company. Calculate the total value of transactions in that economy. If the steel and the car companies doubled in size, what would happen to transactions and to GDP?

transactions in the economy would also increase by roughly 10%. So we can write the relation between the demand for money, nominal income, and the interest rate as:

$$M^d = \$Y \underset{(-)}{L(i)} \qquad (4.1)$$

Where $\$Y$ denotes nominal income. Read this equation in the following way: *The demand for money M^d is equal to nominal income $\$Y$ times a decreasing function of the interest rate, with the function denoted by $L(i)$*. The minus sign under i in $L(i)$ captures the fact that the interest rate has a negative effect on money demand: An increase in the interest rate *decreases* the demand for money, as people put more of their wealth into bonds.

What matters here is nominal income—income in dollars, not real income. If real income does not change but prices double, leading to a doubling of nominal income, people will need to hold twice as much money to buy the same consumption basket.

Equation (4.1) summarizes what we have discussed so far:

- First, the demand for money increases in proportion to nominal income. If nominal income doubles, increasing from $\$Y$ to $\$2Y$, then the demand for money also doubles, increasing from $\$Y\,L(i)$ to $\$2Y\,L(i)$.
- Second, the demand for money depends negatively on the interest rate. This is captured by the function $L(i)$ and the negative sign underneath: An increase in the interest rate decreases the demand for money.

The relation between the demand for money, nominal income, and the interest rate implied by equation (4.1) is shown in Figure 4-1. The interest rate, i, is measured on the vertical axis. Money, M, is measured on the horizontal axis.

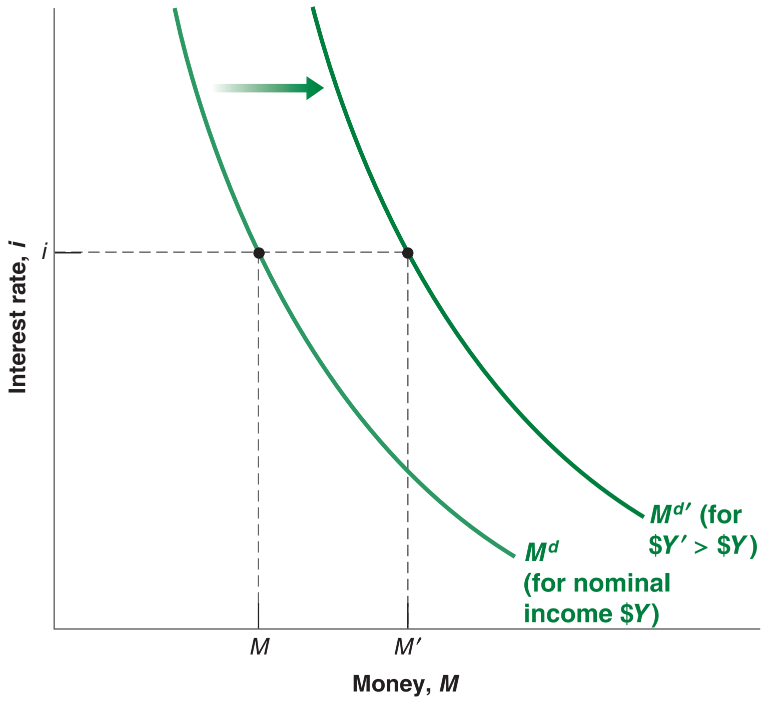
The relation between the demand for money and the interest rate *for a given level of nominal income* $\$Y$ is represented by the M^d curve. The curve is downward sloping: The lower the interest rate (the lower i), the higher the amount of money people want to hold (the higher M).

For a given interest rate, an increase in nominal income increases the demand for money. In other words, an increase in nominal income shifts the demand for money to the right, from M^d to $M^{d\prime}$. For example, at interest rate i, an increase in nominal income from $\$Y$ to $\$Y'$ increases the demand for money from M to M'.

Figure 4-1

The Demand for Money.

For a given level of nominal income, a lower interest rate increases the demand for money. At a given interest rate, an increase in nominal income shifts the demand for money to the right.



FOCUS

Who Holds US Currency?

According to a 2006 household survey, the average US household held $1,600 in currency (dollar bills and coins). Multiplying by the number of households in the US economy at the time (about 110 million), this implies that the total amount of currency held by US households was around $170 billion.

But according to the Federal Reserve Board—which issues the dollar bills and therefore knows how much is in circulation—the amount of currency in circulation was actually a much higher $750 billion. Here lies the puzzle: If it was not held by households, where was all this currency?

Clearly some currency was held by firms rather than by households. And some was held by those involved in the underground economy or in illegal activities. When dealing with drugs, for example, dollar bills (and, now, bitcoin?), not checks, are the way to settle accounts. Surveys of firms and IRS estimates of the underground economy suggest, however, that this can only account for another $80 billion at most. This leaves $500 billion, or 66% of the total, unaccounted for. Where was it? The answer: Abroad, held by foreigners.

A few countries, Ecuador and El Salvador among them, have actually adopted the US dollar as their own currency. So people in these countries use dollar bills for transactions. But these countries are just too small to explain the puzzle.

In a number of countries that have suffered from high inflation in the past, people have learned that their domestic currency may quickly become worthless and they see dollars as a safe and convenient asset. This is, for example, the case for Argentina and Russia. Estimates by the US Treasury suggest that Argentina holds more than $50 billion in dollar bills, Russia more than $80 billion—together, close to the holdings of US households.

In yet other countries, people who have emigrated to the United States send home US dollar bills; or tourists pay some transactions in dollars, and the dollar bills stay in the country. This is, for example, the case for Mexico or Thailand.

The fact that foreigners hold such a high proportion of the dollar bills in circulation has two main macroeconomic implications. First, the rest of the world, by being willing to hold US currency, is making in effect an interest-free loan to the United States to the tune of $500 billion. Second, while we shall think of money demand (which includes both currency and checkable deposits) as being determined by the interest rate and the level of transactions in the country, it is clear that US money demand also depends on other factors. Can you guess, for example, what would happen to US money demand if the degree of civil unrest increased in the rest of the world?

4-2 DETERMINING THE INTEREST RATE: I

Having looked at the demand for money, we now look at the supply of money and then at the equilibrium.

In the real world, there are two types of money: checkable deposits, which are supplied by banks, and currency, which is supplied by the central bank. In this section, we shall assume that the only money in the economy is currency, central bank money. This is clearly not realistic, but it will make the basic mechanisms most transparent. We shall reintroduce checkable deposits and look at the role banks play in the next section.

Money Demand, Money Supply, and the Equilibrium Interest Rate

Suppose the central bank decides to supply an amount of money equal to M, so

$$M^s = M$$

The superscript s stands for *supply*. (Let's disregard, for the moment, the issue of how exactly the central bank supplies this amount of money. We shall return to it in a few paragraphs.)

Throughout this section, the term *money* means central bank money, or currency.

Equilibrium in financial markets requires that money supply be equal to money demand, that $M^s = M^d$. Then, using $M^s = M$, and equation (4.1) for money demand, the equilibrium condition is

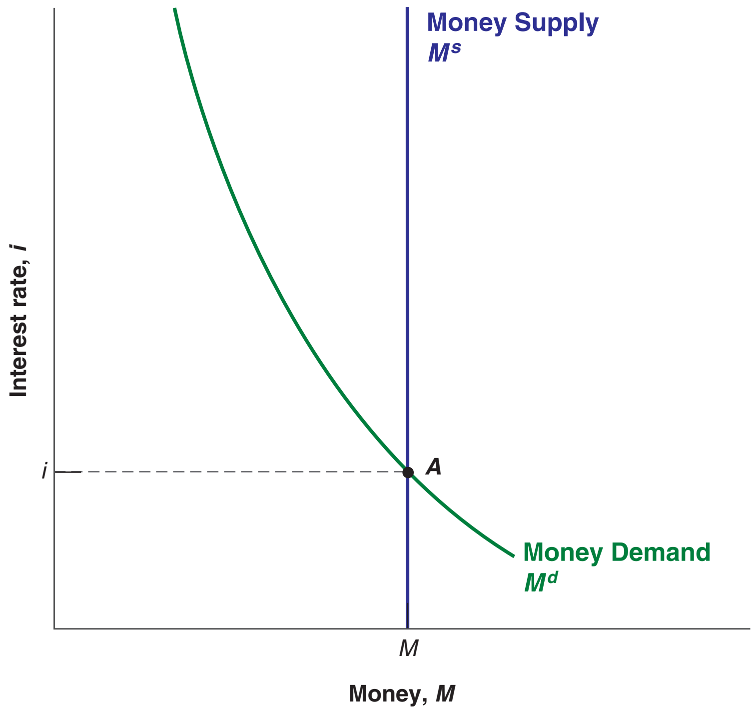
$$\text{Money supply} = \text{Money demand}$$

$$M = \$Y\,L(i) \tag{4.2}$$

Figure 4-2

The Determination of the Interest Rate.

The interest rate must be such that the supply of money is equal to the demand for money.



This equation tells us that the interest rate i must be such that, given their income $\$Y$, people are willing to hold an amount of money equal to the existing money supply M.

This equilibrium condition is represented graphically in Figure 4-2. As in Figure 4-1, money is measured on the horizontal axis and the interest rate is measured on the vertical axis. The demand for money, M^d, drawn for a given level of nominal income, $\$Y$, is downward sloping: A lower interest rate implies a higher demand for money. The supply of money is drawn as the vertical line denoted M^s: The money supply equals M and is independent of the interest rate. Equilibrium occurs at point A, and the equilibrium interest rate is given by i.

Now that we have characterized the equilibrium, we can look at how changes in the money supply by the central bank or changes in nominal income affect the equilibrium interest rate.

- Figure 4-3 shows the effects of an increase in the money supply on the interest rate.

  The initial equilibrium is at point A, with interest rate i. An increase in the money supply, from $M^s = M$ to $M^{s\prime} = M'$, leads to a shift of the money supply curve to the right, from M^s to $M^{s\prime}$. The equilibrium moves from A down to A'; the interest rate decreases from i to i'.

  In words: *an increase in the supply of money by the central bank leads to a decrease in the interest rate.* The decrease in the interest rate increases the demand for money so it equals the now larger money supply.
- Figure 4-4 shows the effects of an increase in nominal income on the interest rate.

  The figure replicates Figure 4-2, and the initial equilibrium is at point A. An increase in nominal income from $\$Y$ to $\$Y'$ increases the level of transactions, which increases the demand for money at any interest rate. The money demand curve *shifts* to the right, from M^d to $M^{d\prime}$. The equilibrium moves from A up to A', and the equilibrium interest rate increases from i to i'.

  In words: For a given money supply, *an increase in nominal income leads to an increase in the interest rate*. The reason: At the initial interest rate, the demand for money exceeds the supply. The increase in the interest rate decreases the amount of money people want to hold and reestablishes equilibrium.

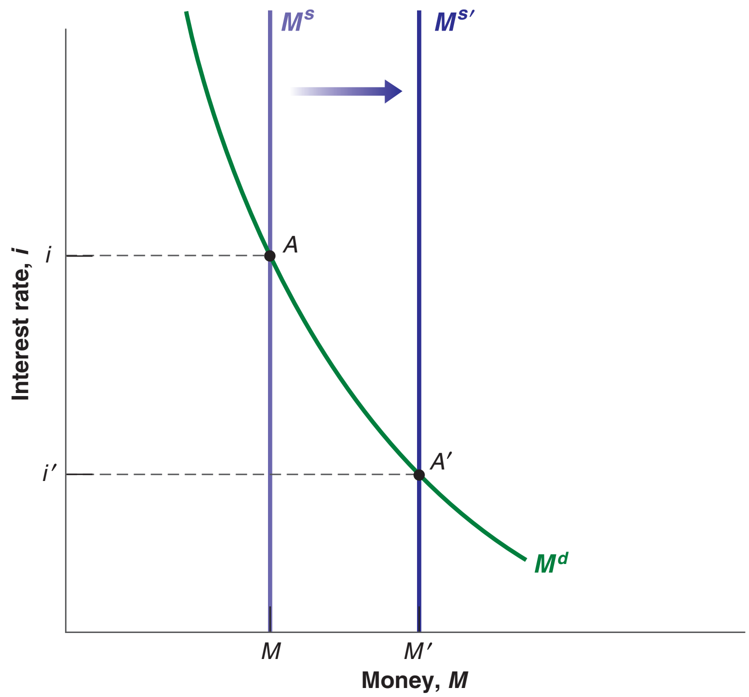


Figure 4-3

The Effects of an Increase in the Money Supply on the Interest Rate.

An increase in the supply of money leads to a decrease in the interest rate.

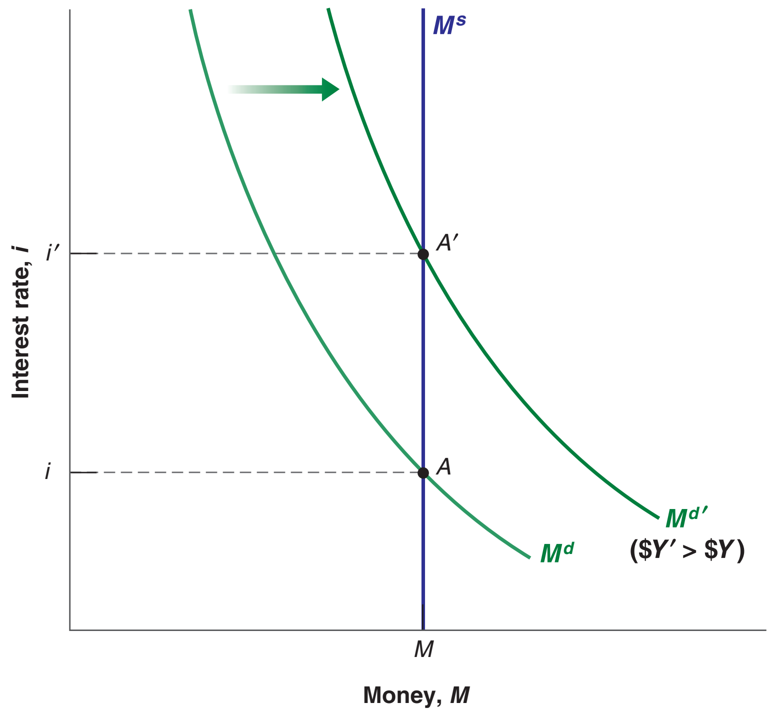


Figure 4-4

The Effects of an Increase in Nominal Income on the Interest Rate.

Given the money supply, an increase in nominal income leads to an increase in the interest rate.

Monetary Policy and Open Market Operations

We can get a better understanding of the results in Figures 4-3 and 4-4 by looking more closely at how the central bank actually changes the money supply, and what happens when it does so.

The way central banks typically change the supply of money is by buying or selling bonds in the bond market. If a central bank wants to increase the amount of money in

Central banks buy and sell other assets, or engage in lending to or borrowing from banks. But leave this aside for the time being.

the economy, it buys bonds and pays for them by creating money. If it wants to decrease the amount of money in the economy, it sells bonds and removes from circulation the money it receives in exchange for the bonds. These actions are called **open market operations** because they take place in the "open market" for bonds.

The Balance Sheet of the Central Bank

To understand what open market operations do, it is useful to start with the balance sheet of the central bank, given in Figure 4-5. The assets of the central bank are the bonds it holds in its portfolio. Its liabilities are the stock of money in the economy. Open market operations lead to equal changes in assets and liabilities.

The balance sheet of a bank (or firm, or individual) is a list of its assets and liabilities at a point in time. The assets are the sum of what the bank owns and what is owed to the bank by others. The liabilities are what the bank owes to others. It goes without saying that Figure 4-5 gives a much simplified version of an actual central bank balance sheet, but it will do for our purposes.

If the central bank buys, say, \$1 million worth of bonds, the amount of bonds it holds is higher by \$1 million, and so is the amount of money in the economy. Such an operation is called an **expansionary open market operation**, because the central bank increases (*expands*) the supply of money.

If the central bank sells \$1 million worth of bonds, both the amount of bonds held by the central bank and the amount of money in the economy are lower by \$1 million. Such an operation is called a **contractionary open market operation**, because the central bank decreases (*contracts*) the supply of money.

Bond Prices and Bond Yields

We have focused so far on the interest rate on bonds. In fact, what is determined in bond markets are not interest rates, but bond *prices.* The two are, however, directly related. Understanding the relation between the two will prove useful both here and later in this book.

- Suppose the bonds are one-year bonds—bonds that promise a payment of a given number of dollars, say \$100, a year from now. In the United States, bonds issued by the government promising payment in a year or less are called **Treasury bills** or **T-bills**. Let the price of a bond today be $\$P_B$, where the subscript B stands for "bond." If you buy the bond today and hold it for a year, the rate of return on holding the bond for a year is $(\$100 - \$P_B)/\$P_B$. Therefore, the interest rate on the bond is given by

$$i = \frac{\$100 - \$P_B}{\$P_B}$$

Figure 4-5

The Balance Sheet of the Central Bank and the Effects of an Expansionary Open Market Operation.

The assets of the central bank are the bonds it holds. The liabilities are the stock of money in the economy. An open market operation in which the central bank buys bonds and issues money increases both assets and liabilities by the same amount.

Central Bank Balance Sheet

Assets	Liabilities
Bonds	Money (currency)

The Effects of an Expansionary Open Market Operation

Assets	Liabilities
Change in bond holdings: +\$1 million	Change in money stock: +\$1 million

If $\$P_B$ is \$99, the interest rate equals $\$1/\$99 = 0.010$, or 1.0% per year. If $\$P_B$ is \$90, the interest rate is $\$10/\$90 = 11.1\%$ per year. *The higher the price of the bond, the lower the interest rate*.

- If we are given the interest rate, we can figure out the price of the bond using the same formula. Reorganizing the formula above, the price today of a one-year bond paying \$100 a year from today is given by

$$\$P_B = \frac{100}{1 + i}$$

The price of the bond today is equal to the final payment divided by 1 plus the interest rate. If the interest rate is positive, the price of the bond is less than the final payment. *The higher the interest rate, the lower the price today*. You may read or hear that "bond markets went up today." This means that *the prices of bonds went up*, and therefore that *interest rates went down.*

The interest rate is what you get for the bond a year from now (\$100) minus what you pay for the bond today ($\$P_B$), divided by the price of the bond today ($\$P_B$).

Back to Open Market Operations

We are now ready to return to the effects of an open market operation and its effect on equilibrium in the money market.

Consider first an expansionary open market operation, in which the central bank buys bonds in the bond market and pays for them by creating money. As the central bank buys bonds, the demand for bonds goes up, increasing their price. Conversely, the interest rate on bonds goes down. Note that by buying the bonds in exchange for money that it created, the central bank has increased the money supply.

Consider instead a contractionary open market operation, in which the central bank decreases the supply of money. This leads to a decrease in the bond price. Conversely, the interest rate goes up. Note that by selling the bonds in exchange for money previously held by households, the central bank has reduced the money supply.

This way of describing how monetary policy affects interest rates is pretty intuitive. By buying or selling bonds in exchange for money, the central bank affects the price of bonds, and by implication, the interest rate on bonds.

Let's summarize what we have learned in the first two sections:

- The interest rate is determined by the equilibrium condition that the supply of money equals the demand for money.
- By changing the supply of money, the central bank can affect the interest rate.
- The central bank changes the supply of money through open market operations, which are purchases or sales of bonds for money.
- Open market operations in which the central bank increases the money supply by buying bonds lead to an increase in the price of bonds and a decrease in the interest rate. In Figure 4-2, the purchase of bonds by the central bank shifts the money supply to the right.
- Open market operations in which the central bank decreases the money supply by selling bonds lead to a decrease in the price of bonds and an increase in the interest rate. In Figure 4-2, the sale of bonds by the central bank shifts the money supply to the left.

Choosing Money or Choosing the Interest Rate?

Let me take up one more issue before moving on. I have described the central bank as choosing the money supply and letting the interest rate be determined at the point where money supply equals money demand. Instead, I could have described the central bank as choosing the interest rate and then adjusting the money supply so as to achieve the interest rate it has chosen.

To see this, return to Figure 4-3. Figure 4-3 showed the effect of a decision by the central bank to increase the money supply from M^s to $M^{s\prime}$ causing the interest rate to fall from i to i'. However, we could have described the figure in terms of the central bank decision to lower the interest rate from i to i' by increasing the money supply from M^s to $M^{s\prime}$.

Or return to Figure 4-4. Figure 4-4 showed the effect of an increase in nominal income, which led to a shift in the money demand curve from M^d to $M^{d\prime}$, causing the interest rate to increase from i to i'. Suppose, however, that the central bank did not want the interest rate to increase. Then, it could shift the money supply to the right, until the equilibrium interest rate remained unchanged and equal to i. In this case, the increase in money demand would be fully reflected in an equal increase in the money supply.

Why is it useful to think about the central bank as choosing the interest rate? Because this is what modern central banks, including the Fed, typically do. They typically think about the interest rate they want to achieve, and then move the money supply to achieve it. This is why, when you listen to the news, you do not hear: "The Fed decided to decrease the money supply today." Instead you hear: "The Fed decided to increase the interest rate today." The way the Fed did it was by decreasing the money supply appropriately.

4-3 DETERMINING THE INTEREST RATE: II

We took a shortcut in Section 4–2 in assuming that all money in the economy consisted of currency supplied by the central bank. In the real world, money includes not only currency but also checkable deposits. Checkable deposits are supplied not by the central bank but by (private) banks. In this section, we reintroduce checkable deposits and examine how this changes our conclusions. Let me give you the bottom line: Even in this more complicated case, by changing the amount of central bank money the central bank still can, and does, control the interest rate.

To understand what determines the interest rate in an economy with both currency and checkable deposits, we must first look at what banks do.

What Banks Do

Modern economies are characterized by the existence of many types of **financial intermediaries**—institutions that receive funds from people and firms and use these funds to buy financial assets or to make loans to other people and firms. The assets of these institutions are the financial assets they own and the loans they have made. Their liabilities are what they owe to the people and firms from whom they have received funds.

Banks are one type of financial intermediary. What makes banks special—and the reason we focus on banks here rather than on financial intermediaries in general—is that their liabilities are money: People can pay for transactions by writing checks up to the amount of their account balance. Let's look more closely at what they do.

Banks have other types of liabilities in addition to checkable deposits, and they are engaged in more activities than just holding bonds or making loans. Ignore these complications for the moment. We consider them in Chapter 6.

The balance sheet of banks is shown in Figure 4-6(b).

- Banks receive funds from people and firms who either deposit funds directly or have funds sent to their checking account (via direct deposit of their paycheck, for example). At any point in time, people and firms can write checks, use a debit card, or

(a) **Central Bank**

Assets	Liabilities
Bonds	Central Bank Money = Reserves + Currency

(b) **Banks**

Assets	Liabilities
Reserves Loans Bonds	Checkable deposits

Figure 4-6

The Balance Sheet of Banks, and the Balance Sheet of the Central Bank Revisited.

withdraw funds, up to the full amount of their account balance. The liabilities of the banks are therefore equal to the value of these *checkable deposits.*

- Banks keep as **reserves** some of the funds they receive. They are held partly in cash and partly in an account the banks have at the central bank, which they can draw on when they need to. Banks hold reserves for four reasons:

  On any given day, some depositors withdraw cash from their checking accounts, and others deposit cash into their accounts. There is no reason for the inflows and outflows of cash to be equal, so the bank must keep some cash on hand.

  In the same way, on any given day, people with accounts at the bank write checks to people with accounts at other banks, and people with accounts at other banks write checks to people with accounts at the bank. As a result of these transactions, what the bank owes the other banks can be larger or smaller than what the other banks owe to it. For this reason also, the bank needs to keep reserves.

  The first two reasons imply that the banks would want to keep some reserves even if they were not required to do so. But, in addition, banks are typically subject to reserve requirements, which require them to hold reserves in some proportion of their checkable deposits. In the United States, reserve requirements are set by the Fed. In the US, banks are required to hold at least 10% of the value of the checkable deposits. They can use the rest to make loans or buy bonds.

  The last reason is that, in many countries including the United States, the central bank now pays interest on reserves. The higher the interest rate on reserves, the less unattractive it is for banks to hold reserves rather than to buy bonds or make loans, thus the higher the reserves they are willing to hold. In this chapter, for simplicity, I shall assume that reserves do not pay interest. But the interest rate on reserves is becoming a more and more important monetary policy instrument, and I shall return to the issue in Chapter 23.
- Loans represent roughly 70% of banks' nonreserve assets. Bonds account for the rest, 30%. The distinction between bonds and loans is unimportant for our purposes in this chapter on how the money supply is determined. For this reason, I shall assume that banks do not make loans, that they hold only reserves and bonds as assets.

◀ The distinction between loans and bonds is important for other purposes, from the possibility of "bank runs" to the role of federal deposit insurance. More on this in Chapter 6.

Figure 4-6(a) returns to the balance sheet of the central bank, in an economy in which there are banks. It is similar to the balance sheet of the central bank we saw in Figure 4-5. The asset side is the same as before: The assets of the central bank are the bonds it holds. The liabilities of the central bank are the money it has issued,

central bank money. The new feature, relative to Figure 4-5, is that not all central bank money is held as currency by the public. Some of it is held as reserves by banks.

The Demand for and Supply of Central Bank Money

How do we think about the equilibrium in this more realistic setting? Very much in the same way as before, in terms of the *demand for and supply of central bank money.*

- The demand for central bank money is now equal to the demand for currency by people plus the demand for reserves by banks.
- The supply of central bank money is under the direct control of the central bank.
- The equilibrium interest rate is such that the demand and the supply for central bank money are equal.

The Demand for Central Bank Money

The demand for central bank money now has two components. The first is the demand for currency by people, the second is the demand for reserves by banks. To make the algebra simple, I shall assume that people want to hold money only in the form of checkable deposits, and do not hold any currency. Relaxing this assumption would involve more algebra, but yield the same basic conclusions.

In this case, the demand for central bank money is simply the demand for reserves by banks. This demand in turn depends on the demand for checkable deposits by people. So let's start there. Under our assumption that people hold no currency, the demand for checkable deposits is equal to the demand for money by people. So, to describe the demand for checkable deposits, we can use the same equation as we used before (equation (4.1)):

$$M^d = \$Y \underset{(-)}{L(i)} \tag{4.3}$$

People want to hold more checkable deposits the higher their level of transactions and the lower the interest rate on bonds.

Now turn to the demand for reserves by banks. The larger the amount of checkable deposits, the larger the amount of reserves the banks must hold, for both precautionary and regulatory reasons. Let θ (the Greek lowercase letter theta) be the **reserve ratio**, the amount of reserves banks hold per dollar of checkable deposits. Under the assumption that reserves do not pay interest, banks will not want to hold more than what they are required to hold, so you can think of θ as the required reserve ratio.

Then, using equation (4.3), the demand for reserves by banks, call it H^d is given by:

The use of the letter *H* comes from the fact that central bank money is sometimes called **high-powered money**, to reflect its role in determining the equilibrium interest rate. Another name for central bank money is the **monetary base**.

$$H^d = \theta M^d = \theta \$Y\ L(i) \tag{4.4}$$

The first equality reflects the fact that the demand for reserves is proportional to the demand for checkable deposits. The second equality reflects the fact that the demand for checkable deposits depends on nominal income and on the interest rate. So, the demand for central bank money, equivalent to the demand for reserves by banks, is equal to θ times the demand for money by people.

Equilibrium in the Market for Central Bank Money

As before, the supply of central bank money—equivalently the supply of reserves by the central bank—is under the control of the central bank. And as before, the central bank can change the amount of central bank money H through open market operations. The

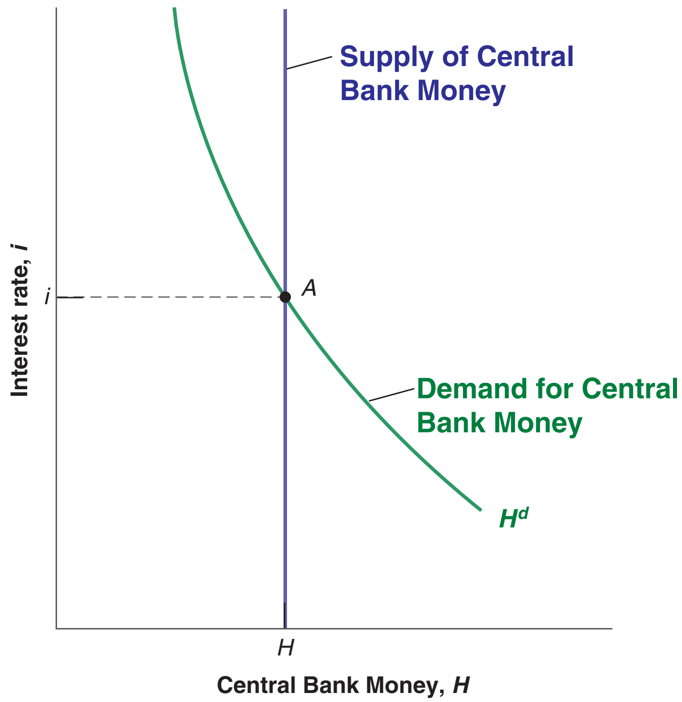


Figure 4-7

Equilibrium in the Market for Central Bank Money and the Determination of the Interest Rate.

The equilibrium interest rate is such that the supply of central bank money is equal to the demand for central bank money.

equilibrium condition is that the supply of central bank money be equal to the demand for central bank money:

$$H = H^d \tag{4.5}$$

Or, using equation (4.4):

$$H = \theta \$Y\, L(i) \tag{4.6}$$

We can represent the equilibrium condition, equation (4.6), graphically, and we do this in Figure 4-7. The figure looks the same as Figure 4-2, but with central bank money (H) rather than money (M) on the horizontal axis. The interest rate is measured on the vertical axis. The demand for central bank money, H^d, is drawn for a given level of nominal income. A higher interest rate implies a lower demand for central bank money as the demand for checkable deposits by people, and thus the demand for reserves by banks, goes down. The supply of money is fixed and is represented by a vertical line at H. Equilibrium is at point A, with interest rate i.

The effects of either changes in nominal income or changes in the supply of central bank money are qualitatively the same as in the previous section. In particular, an increase in the supply of central bank money leads to a shift in the vertical supply line to the right. This leads to a lower interest rate. Conversely, a decrease in the supply of central bank money leads to an increase in the interest rate. So, the basic conclusion is the same as in Section 4–2: By controlling the supply of central bank money, the central bank can determine the interest rate on bonds.

The Federal Funds Market and the Federal Funds Rate

You may wonder whether there is an actual market in which the demand and the supply of reserves determine the interest rate. And, indeed, in the United States, there is an actual market for bank reserves, where the interest rate adjusts to balance the supply and demand for reserves. This market is called the **federal funds market**. The interest

FOCUS

Will Bitcoins Replace Dollars?

Bitcoins are virtual assets that can be used in transactions. Their use relies on a technology that allows transactions to take place without any third party being involved. As of December 2018, the total number of bitcoins in circulation was 17.3 million. One bitcoin was worth $3,900, so the total value of the bitcoin stock was $67 billion. This may sound big, but it is very small compared to the total value of the stock of US currency, $1.5 trillion. Despite this, some bitcoin supporters argue that, one day, bitcoins will replace dollars (and euros and yen) as the currency used in most transactions. If this were to happen, the supply of currency would no longer be determined by the Fed (or other central banks), but by the mechanical rule that determines the creation of new bitcoins over time. Monetary policy as we know it would disappear (and much of my textbook would have to be rewritten).

Is it at all conceivable? One can indeed imagine a world in which all prices were quoted in bitcoins, all interest rates were quoted in bitcoins, and people and firms did all their transactions in bitcoins. Then, there would be no reason to hold dollars, and dollars would indeed disappear.

Is it likely to happen? No. For three reasons:

First, for economic reasons. So long as most prices are quoted in dollars, transactions in bitcoins will involve what is known as "price risk." If the price of bitcoins in terms of dollars were stable, the risk would be limited. This is not the case. As shown in Figure 1 below, since its introduction, the price of bitcoins has fluctuated widely. From about $1,000 at the beginning of 2017, it soared to $19,300 at the end of 2017, only to fall back to $3,900 at the end of 2018. Holding bitcoins is a very risky proposition and is likely to remain so.

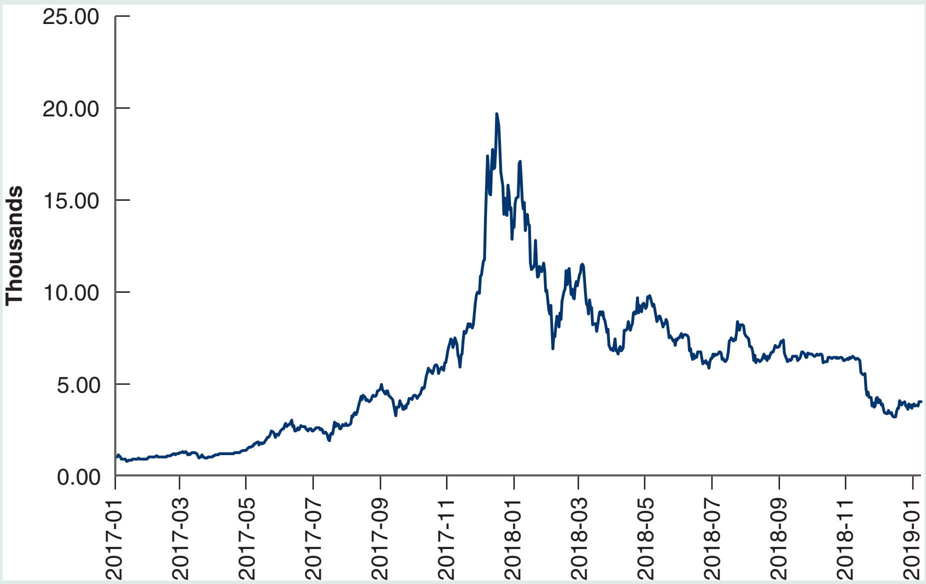
Second, for technical reasons. The technology of bitcoins is such that the verification process and the creation of bitcoins are extremely expensive. The electricity used today by computers for Bitcoin-related operations would be enough to supply electricity to 4 million US households. Were the use of bitcoins increase dramatically, the costs would most likely follow.

Third, for political reasons. Countries want to keep control of their monetary policy and will not relinquish it to private actors. Moving to an equilibrium where bitcoins become the main or the only currency would require a large coordination among price setters: You will typically not want to quote your prices in bitcoins if nearly everybody else quotes theirs in dollars. You will only do it if nearly everybody else quotes theirs in bitcoins. Such coordination to move from one equilibrium to the other can be achieved only with the help of the state. And the states will not willingly give up their currency.

Figure 1

Price of Bitcoins in Dollars, 2017–2018.

For more on Bitcoin: Rainer Böhme, Nicolas Christin, Benjamin Edelman, Tyler Moore, "Bitcoin: Economics, Technology, and Governance," *Journal of Economic Perspectives*, 2015, 29(2): pp. 213–238.



rate determined in this market is called the **federal funds rate**. Because the Fed can in effect choose the federal funds rate by changing the supply of central bank money, H, the federal funds rate is typically thought of as the main indicator of US monetary policy. This is why so much attention is focused on it, and why changes in the federal funds rate typically make front page news.

4-4 THE LIQUIDITY TRAP

The main conclusion from the first three sections was that the central bank can, by choosing the supply of central bank money, choose the interest rate that it wants. If it wants to increase the interest rate, it decreases the amount of central bank money. If it wants to decrease the interest rate, it increases the amount of central bank money. This section shows that this conclusion comes with an important caveat: The interest rate cannot go below zero, a constraint known as the **zero lower bound**. When the interest rate is down to zero, monetary policy cannot decrease it further. Monetary policy no longer works, and the economy is said to be in a **liquidity trap**.

The concept of a liquidity trap (i.e., a situation in which increasing the amount of money ["liquidity"] does not have an effect on the interest rate [the liquidity is "trapped"]) was developed by John Maynard Keynes in the 1930s, although the expression itself came later.

Fifteen years ago, the zero lower bound was seen as a mostly irrelevant issue. Most economists believed that there would never be a need for negative interest rates, so the constraint did not matter. The Great Financial Crisis, however, changed those perceptions. Many central banks decreased interest rates down to zero and would have liked to go down even further. But the zero lower bound stood in the way and turned out to be a serious constraint on policy.

Let's look at the argument more closely. When we derived the demand for money in Section 4–1, we did not ask what happens when the interest rate is down to zero. Now we must ask the question. The answer: When the interest rate on bonds is down to zero, and once people hold enough money for transaction purposes, they are then indifferent between holding the rest of their financial wealth in the form of money or in the form of bonds. The reason they are indifferent is that both money and bonds pay the same interest rate, namely zero. Thus, the demand for money is as shown in Figure 4-8.

If you look at Figure 4-1, you will see that I avoided the issue by not drawing the demand for money when interest rates are close to zero.

In fact, because of the inconvenience and the dangers of holding currency in very large amounts, people and firms are willing to hold some bonds even when the interest rate is a bit negative. This is the case in Germany today (look at German bond rates on the internet), another complication we ignore here.

- As the interest rate decreases, people want to hold more money (and thus fewer bonds): The demand for money increases.
- As the interest rate becomes equal to zero, people want to hold an amount of money at least equal to the distance OB. This is what they need for transaction purposes. But they are willing to hold even more money (and therefore hold fewer bonds) because they are indifferent between money and bonds. Both pay zero interest. Therefore, the demand for money (M^d) becomes horizontal beyond point B.

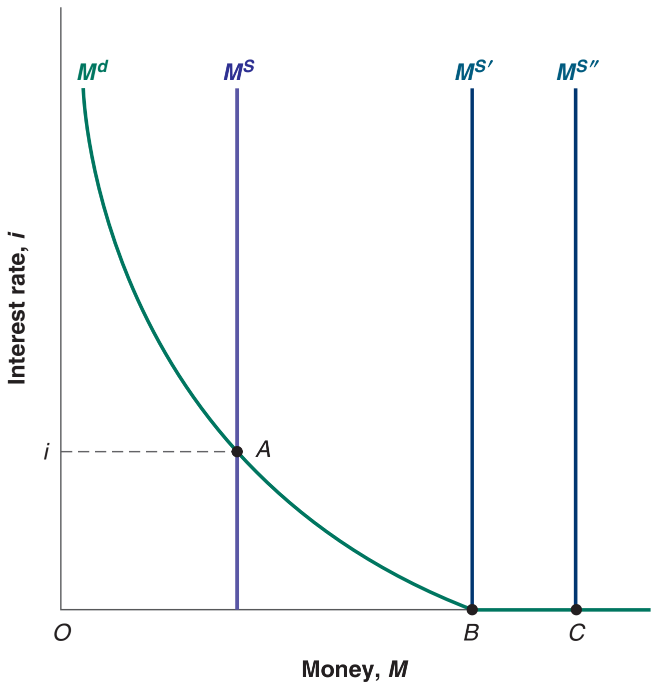


Figure 4-8

Money Demand, Money Supply, and the Liquidity Trap.

When the interest rate is equal to zero, and once people have enough money for transaction purposes, they become indifferent between holding money and holding bonds. The demand for money becomes horizontal. When the interest rate is equal to zero, increases in the money supply have no effect on the interest rate and it remains equal to zero.

FOCUS

The Liquidity Trap in Action

You saw in Chapter 1 how, when the financial crisis began, the Fed decreased the federal funds rate from 5% in mid-2007 to 0% by the end of 2008, when it hit the zero lower bound. It remained at zero for seven years, before starting to slowly increase again.

What did the Fed do? From the end of 2008 on, it continued to increase the money supply through open market operations, buying bonds in exchange for money. The analysis in the text suggests that, despite the interest rate remaining equal to zero, there should have been an increase in checkable deposits by households, and an increase in reserves by banks. And, indeed, as Figure 1 shows, this is exactly what happened. Checkable deposits of both households and firms—which were decreasing before 2007, reflecting the increasing use of credit cards—increased from $620 billion in mid-2008 to $1,700 billion at the end of 2015. Bank reserves rose dramatically from $10 billion in mid-2008 to $700 billion at the end of 2008 and to $2,500 billion by the end of 2015. In other words, the very large increase in the supply of central bank money was willingly absorbed by households and by banks with no change in the interest rate, which remained equal to zero.

This brief description, together with Figure 1, should, however, lead you to ask two questions:

Why did the Fed continue to increase the money supply, despite the fact that it had no effect on the federal funds rate, which remained equal to zero? You will see the reason in Chapter 6: In effect, in an economy with more than one type of bond, open market operations can affect interest rates on other bonds and affect the economy.

Why were banks willing to keep high levels of reserves when, starting at the end of 2015, the interest rate started to increase? You will see the reason in Chapter 22: In order to induce banks to keep high reserve levels, the Fed increased the interest rate on reserves, and this policy instrument is becoming steadily more important.

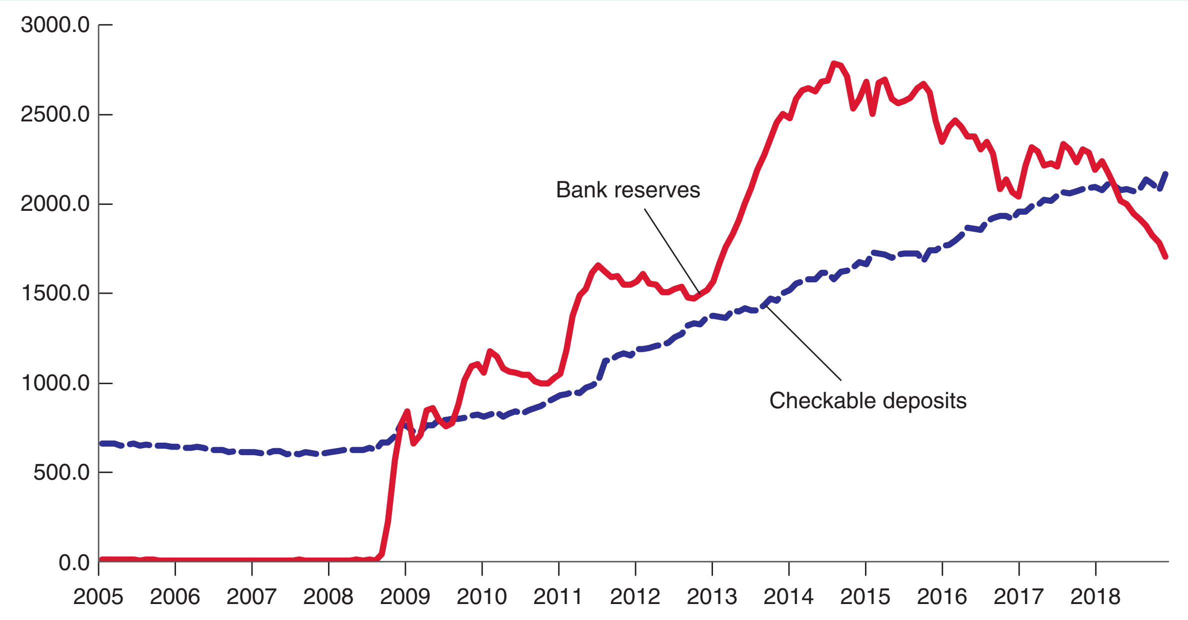


Figure 1

Checkable Deposits and Bank Reserves, 2005–2018. (billions).

Source: FRED: TCP, WRESBAL.

Now consider the effects of an increase in the money supply. (Let's ignore banks for the time being, and assume, as in Section 4–2, that all money is currency, so we can use the same diagram as in Figure 4-2 extended to allow for the horizontal portion of money demand. We come back to banks and bank money below.)

- Consider the case where the money supply is M^s, so the interest rate consistent with financial market equilibrium is positive and equal to i. (This is the case we considered in Section 4–2.) Starting from that equilibrium, an increase in the money supply—a shift of the M^s line to the right—leads to a decrease in the interest rate.
- Now consider the case where the money supply is $M^{s\prime}$, so the equilibrium is at point B; or the case where the money supply is $M^{s\prime\prime}$, so the equilibrium is given by point C. In either case, the initial interest rate is zero. And, in either case, an increase in the money supply has no effect on the interest rate. Think of it this way:

  Suppose the central bank increases the money supply. It does so through an open market operation in which it buys bonds and pays for them by creating money. As the interest rate is zero, people are indifferent to how much money or bonds they hold, so they are willing to hold fewer bonds and more money at the same interest rate, namely zero. The money supply increases, but with no effect on the interest rate—which remains equal to zero.

  What happens when we reintroduce checkable deposits and a role for banks, along the lines of Section 4–3? Everything we just said still applies to the demand for money by people: If the interest rate is zero, they are indifferent to whether they hold money or bonds: Both pay zero interest. But, now a similar argument also applies to banks and their decision whether to hold reserves or buy bonds. If the interest rate is equal to zero, they will also be indifferent as to whether to hold reserves and to buy bonds: Both pay zero interest. Thus, when the interest rate is down to zero, and the central bank increases the money supply, we are likely to see an increase in checkable deposits and an increase in bank reserves, with the interest rate remaining at zero. As the Focus Box "The Liquidity Trap in Action" shows, this is exactly what we saw during the crisis. As the Fed decreased the interest rate to zero, and continued to expand the money supply, both checkable deposits by people and reserves by banks steadily increased.

SUMMARY

- The demand for money depends positively on the level of transactions in the economy and negatively on the interest rate.
- The interest rate is determined by the equilibrium condition that the supply of money be equal to the demand for money.
- For a given supply of money, an increase in income leads to an increase in the demand for money and an increase in the interest rate. An increase in the supply of money for a given income leads to a decrease in the interest rate.
- The way the central bank changes the supply of money is through open market operations.
- Expansionary open market operations, in which the central bank increases the money supply by buying bonds, lead to an increase in the price of bonds and a decrease in the interest rate.
- Contractionary open market operations, in which the central bank decreases the money supply by selling bonds, lead to a decrease in the price of bonds and an increase in the interest rate.
- When money includes both currency and checkable deposits, we can think of the interest rate as being determined by the condition that the supply of central bank money be equal to the demand for central bank money.
- The supply of central bank money is under the control of the central bank. In the special case where people hold only checkable deposits, the demand for central bank money is equal to the demand for reserves by banks, which is itself equal to the overall demand for money times the reserve ratio chosen by banks.
- The market for bank reserves is called the *federal funds market*. The interest rate determined in that market is called the *federal funds rate*.
- The interest rate chosen by the central bank cannot go below zero. When the interest rate is equal to zero, people and banks become indifferent to holding money or bonds. An increase in the money supply leads to an increase in money demand, an increase in reserves by banks, and no change in the interest rate. This case is known as the liquidity trap. In the liquidity trap, monetary policy no longer affects the interest rate.

KEY TERMS

Federal Reserve Bank (Fed), 67
currency, 68
checkable deposits, 68
bonds, 68
money market funds, 68
money, 69
income, 69
flow, 69
saving, 69
savings, 69
financial wealth, 69
stock, 69
investment, 69
financial investment, 69
open market operation, 74
expansionary open market operation, 74
contractionary open market operation, 74
Treasury bill (T-bill), 74
financial intermediaries, 76
(bank) reserves, 77
central bank money, 78
reserve ratio, 78
high-powered money, 78
monetary base, 78
federal funds market, 79
federal funds rate, 80
zero lower bound, 81
liquidity trap, 81

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. Income and financial wealth are both examples of stock variables.
b. The term *investment*, as used by economists, refers to the purchase of bonds and shares of stock.
c. The demand for money does not depend on the interest rate because only bonds earn interest.
d. A large proportion of US currency appears to be held outside the United States.
e. The central bank can increase the supply of money by selling bonds.
f. The Federal Reserve can determine the money supply, but it cannot change interest rates.
g. Bond prices and interest rates always move in opposite directions.
h. An increase in income (GDP) will always be accompanied by an increase in interest rates when the money supply is not increased.
i. Once the interest rate is zero, the Fed has no further policy options.

2. Suppose that a person's yearly income is \$60,000. Also suppose that this person's money demand function is given by

$$M^d = \$Y(0.35 - i)$$

a. What is this person's demand for money when the interest rate is 5%? 10%?
b. Explain how the interest rate affects money demand.
c. Suppose that the interest rate is 10%. In percentage terms, what happens to this person's demand for money if the yearly income is reduced by 50%?
d. Suppose that the interest rate is 5%. In percentage terms, what happens to this person's demand for money if the yearly income is reduced by 50%?
e. Summarize the effect of income on money demand. In percentage terms, how does this effect depend on the interest rate?

3. Consider a bond that promises to pay \$100 in one year.

a. What is the interest rate on the bond if its price today is \$75? \$85? \$95?
b. What is the relation between the price of the bond and the interest rate?
c. If the interest rate is 8%, what is the price of the bond today?

4. Suppose that money demand is given by

$$M^d = \$Y(0.25 - i)$$

where \$Y is \$100. Also, suppose that the supply of money is \$20.

a. What is the equilibrium interest rate?
b. If the Federal Reserve Bank wants to increase the equilibrium interest rate *i* by 10 percentage points from its value in part a, at what level should it set the supply of money?

DIG DEEPER

5. Suppose that a person's wealth is \$50,000 and that her yearly income is \$60,000. Also suppose that her money demand function is given by

$$M^d = \$Y(0.35 - i)$$

a. Derive the demand for bonds. Suppose the interest rate increases by 10 percentage points. What is the effect on her demand for bonds?
b. What are the effects of an increase in wealth on her demand for money and her demand for bonds? Explain in words.
c. What are the effects of an increase in income on her demand for money and her demand for bonds? Explain in words.
d. Consider the statement "When people earn more money, they obviously will hold more bonds." What is wrong with this statement?

6. The demand for bonds

In this chapter, you learned that an increase in the interest rate makes bonds more attractive, so it leads people to hold more of their wealth in bonds as opposed to money. However, you also learned that an increase in the interest rate reduces the price of bonds.

How can an increase in the interest rate make bonds more attractive and reduce their price?

7. ATMs and credit cards

This problem examines the effect of the introduction of ATMs and credit cards on money demand. For simplicity, let's examine a person's demand for money over a period of four days. Suppose that before ATMs and credit cards, this person goes to the bank once at the beginning of each four-day period and withdraws from her savings account all the money she needs for four days. Assume that she needs $4 per day.

a. How much does this person withdraw each time she goes to the bank? Compute this person's money holdings for days 1 through 4 (in the morning, before she needs any of the money she withdraws).

b. What is the amount of money this person holds, on average?

Suppose now that with the advent of ATMs, this person withdraws money once every two days.

c. Recompute your answer to part a.

d. Recompute your answer to part b.

Finally, with the advent of credit cards, this person pays for all her purchases using her card. She withdraws no money until the fourth day, when she withdraws the whole amount necessary to pay for her credit card purchases over the previous four days.

e. Recompute your answer to part a.

f. Recompute your answer to part b.

g. Based on your previous answers, what do you think has been the effect of ATMs and credit cards on money demand?

8. Money and the banking system

I described a monetary system that included simple banks in Section 4-3. Assume the following:

i. *The public holds no currency.*

ii. *The ratio of reserves to deposits is 0.1.*

iii. *The demand for money is given by*

$$M^d = \$Y(0.8 - 4i)$$

Initially, the supply of central bank money is $100 billion, and nominal income is $5 trillion.

a. What is the demand for central bank money?

b. Find the equilibrium interest rate by setting the demand for central bank money equal to the supply of central bank money.

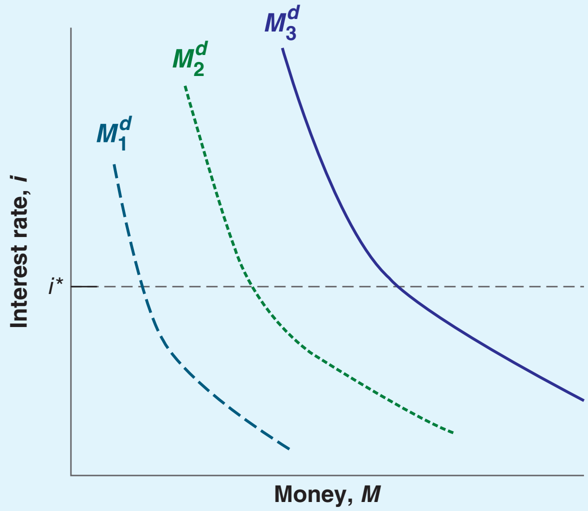
c. What is the overall supply of money? Is it equal to the overall demand for money at the interest rate you found in part b?

d. What is the effect on the interest rate if central bank money is increased to $300 billion?

e. If the overall money supply increases to $3,000 billion, what will be the effect on *i*? [*Hint*: Use what you discovered in part c.]

9. Understanding the Fed's actions that are needed to stabilize the interest rate

The diagram below shows three different money demand curves and a target interest rate *i**



Fill in the table below using the first entry as an example.

Initial Money Demand Curve for real income *Y* and price level *P*				Final Money Demand Curve for real income *Y* and price level *P*				Action required by the Fed to maintain an interest rate at *i**
Initial M^d curve	*Y*	*P*	*$Y*	Final M^d curve	*Y*	*P*	*$Y*	Explanation
M^d_2	250	100	250	M^d_3	300	105	315	The Fed must increase the money supply as nominal income rises — both real income and prices rise
M^d_2	200	80			250	100		
M^d_2	250	100			300	100		
M^d_2	250	100			200	95		
M^d_2	250	100			275	80		

10. Choosing the quantity of money or the interest rate

Suppose that money demand is given by

$$M^d = \$Y(0.25 - i)$$

where $\$Y$ is $\$100$.

a. If the Federal Reserve Bank sets an interest rate target of 5%, what is the money supply the Federal Reserve must create?
b. If the Federal Reserve Bank wants to increase i from 5% to 10%, what is the new level of the money supply the Federal Reserve must set?
c. What is the effect on the Federal Reserve's balance sheet of the increase in the interest rate from 5% to 10%?

11. Monetary policy in a liquidity trap

Suppose that money demand is given by

$$M^d = \$Y(0.25 - i)$$

as long as interest rates are positive. The questions below then refer to situations where the interest rate is zero.

a. What is the demand for money when interest rates are zero and $\$Y = 80$?
b. If $\$Y = 80$, what is the smallest value of the money supply at which the interest rate is zero?
c. Once the interest rate is zero, can the central bank continue to increase the money supply?
d. The United States experienced a long period of zero interest rates after 2008. Can you find evidence in the text that the money supply continued to increase over this period?
e. Go to the database at the Federal Reserve Bank of St. Louis known as FRED. Find the series BOGGMBASE (the central bank money, also called the monetary base) and look at its behavior from 2010 to 2015. What happened to the monetary base? What happened to the federal funds rate in the same period?

EXPLORE FURTHER

12. Current monetary policy

Go to the website for the Federal Reserve Board of Governors (www.federalreserve.gov) and download the most recent monetary policy press release of the Federal Open Market Committee (FOMC), the body that makes monetary policy decisions. Make sure you get the most recent FOMC press release and not simply the most recent Fed press release.

a. What is the current stance of monetary policy? (Note that policy will be described in terms of increasing or decreasing the federal funds rate as opposed to increasing or decreasing the money supply or the monetary base.)
b. Find a press release announcing a change in the federal funds rate. How did the Federal Reserve explain the need for that change in monetary policy?

Finally, visit the Fed's website and find various statements explaining the Fed's current policy on interest rates. These statements set the stage for the analysis in Chapter 5. Some parts of these statements will be clearer once you have read Chapter 5.

FURTHER READINGS

- While we shall return to many aspects of the financial system throughout the book, you may want to dig deeper and read a textbook on money and banking. Here are four of them:
  Money, Banking, and Financial Markets, by Laurence Ball (Worth, 2017)
  Money, Banking, and Financial Markets, by Stephen Cecchetti and Kermit Schoenholtz (McGraw-Hill/Irwin, 2017)
  Money, the Financial System and the Economy, by R. Glenn Hubbard (Addison-Wesley, 2013)
  The Economics of Money, Banking, and the Financial System, by Frederic Mishkin (Pearson, 2018)
- The Fed maintains a useful website with data on financial markets, information on what the Fed does, recent testimonies by the Fed Chair, and so on (www.federalreserve.gov).

5

Goods and Financial Markets: The IS-LM Model

In Chapter 3, we looked at the goods market. In Chapter 4, we looked at financial markets. We now look at goods and financial markets together. By the end of this chapter you will have a framework to think about how output and the interest rate are determined in the short run.

In developing this framework, we follow a path first traced by two economists, John Hicks and Alvin Hansen in the late 1930s and early 1940s. When the economist John Maynard Keynes published his *General Theory* in 1936, there was much agreement that his book was both fundamental and nearly impenetrable. (Try to read it, and you will agree.) There were (and still are) many debates about what Keynes "really meant." In 1937, John Hicks summarized what he saw as one of Keynes's main contributions: the joint description of goods and financial markets. His analysis was later extended by Alvin Hansen. Hicks and Hansen called their formalization the *IS-LM* (investment-saving, liquidity-money) model.

Macroeconomics has made substantial progress since the early 1940s. This is why the IS-LM model is treated in this and the next chapter rather than in Chapter 24 of this book. (If you had taken this course 40 years ago, you would be nearly done!) But to most economists, the IS-LM model still represents an essential building block—one that, despite its simplicity, captures much of what happens in the economy in *the short run.* This is why the IS-LM model is still taught and used today.

The version of the IS-LM model presented in this book is a bit different (and, you will be happy to know, simpler) than the model developed by Hicks and Hansen. This reflects a change in the way central banks now conduct monetary policy, with a shift in focus from controlling the money stock to controlling the interest rate, as explained in Chapter 4.

This chapter develops the basic version of the IS-LM model.

Section 5-1 looks at equilibrium in the goods market and derives the IS relation.

Section 5-2 looks at equilibrium in financial markets and derives the LM relation.

Sections 5-3 and 5-4 put the IS and LM relations together and use the resulting IS-LM model to study the effects of fiscal and monetary policy—first separately, then together.

Section 5-5 introduces dynamics and explores how the IS-LM model captures what happens in the economy in the short run.

If you remember one basic message from this chapter, it should be: In the short run, output is determined by equilibrium in the goods and financial markets. ▶▶▶

5-1 THE GOODS MARKET AND THE IS RELATION

Let's first summarize what we learned in Chapter 3:

- We characterized equilibrium in the goods market as the condition that production, Y, be equal to the demand for goods, Z. We called this condition the IS relation.
- We defined demand as the sum of consumption, investment, and government spending. We assumed that consumption was a function of disposable income (income minus taxes), and took investment spending, government spending, and taxes as given:

$$Z = C(Y - T) + \bar{I} + G$$

(In Chapter 3, we assumed, to simplify the algebra, that the relation between consumption, C, and disposable income, $Y-T$, was linear. Here, we shall not make this assumption but use the more general form $C = C(Y-T)$ instead.)

- The equilibrium condition was thus given by

$$Y = C(Y - T) + \bar{I} + G$$

Using this equilibrium condition, we then looked at the factors that moved equilibrium output. We looked in particular at the effects of changes in government spending and of shifts in consumption demand.

Much more on the effects of interest rates on both consumption and investment in Chapter 15.

The main simplification of this first model was that the interest rate did not affect the demand for goods. Our first task in this chapter is to abandon this simplification and introduce the interest rate in our model of equilibrium in the goods market. For the time being, we focus only on the effect of the interest rate on investment and leave the discussion of its effects on the other components of demand until later.

Investment, Sales, and the Interest Rate

In Chapter 3, investment was assumed to be constant. This was for simplicity. Investment is in fact far from constant and depends primarily on two factors:

- The level of sales. Consider a firm facing an increase in sales and needing to increase production. To do so, it may need to buy additional machines or build an additional plant. In other words, it needs to invest. A firm facing low sales will feel no such need and will spend little, if anything, on investment.
- The interest rate. Consider a firm deciding whether to buy a new machine. Suppose that to buy the new machine, the firm must borrow. The higher the interest rate, the less attractive it is to borrow and buy the machine. (For the moment, and to keep things simple, we make two simplifications. First, we assume that all firms can borrow at the same interest rate—namely, the interest rate on bonds as determined in Chapter 4. In fact, many firms borrow from banks, possibly at a different rate. We also leave aside the distinction between the nominal interest rate—the interest rate in terms of dollars—and the real interest rate—the interest rate in terms of goods. We return to both issues in Chapter 6.) At a high enough interest rate, the additional profits from using the new machine will not cover interest payments, and the new machine will not be worth buying.

The argument still holds if the firm uses its own funds: The higher the interest rate, the more attractive it is to lend the funds rather than to use them to buy the new machine.

To capture these two effects, we write the investment relation as follows:

$$I = \underset{(+,-)}{I(Y, i)} \qquad (5.1)$$

Equation (5.1) states that investment I depends on production Y and the interest rate i. (We continue to assume that inventory investment is equal to zero, so sales and production are always equal. As a result, Y denotes both sales and production.) The positive sign under Y indicates that an increase in production (equivalently, an increase in sales) leads to an increase in investment. The negative sign under the interest rate i indicates that an increase in the interest rate leads to a decrease in investment.

An increase in output leads to an increase in investment. An increase in the interest rate leads to a decrease in investment.

Determining Output

Taking into account the investment relation (5.1), the condition for equilibrium in the goods market becomes

$$Y = C(Y - T) + I(Y, i) + G \tag{5.2}$$

Production (the left side of the equation) must be equal to the demand for goods (the right side). Equation (5.2) is our expanded IS relation. We can now look at what happens to output when the interest rate changes.

Start with Figure 5-1. Measure the demand for goods (Z) on the vertical axis. Measure output (Y) on the horizontal axis. For a given value of the interest rate i, demand is an increasing function of output, for two reasons:

- An increase in output leads to an increase in income and thus to an increase in disposable income. The increase in disposable income leads to an increase in consumption. We studied this relation in Chapter 3.
- An increase in output also leads to an increase in investment. This is the relation between investment and sales that we have introduced in this chapter.

In short, an increase in output leads, through its effects on both consumption and investment, to an increase in the demand for goods. This relation between demand and output, for a given interest rate, is represented by the upward-sloping curve ZZ.

Note two characteristics of ZZ in Figure 5-1:

- Because we have not assumed that the consumption and investment relations in equation (5.2) are linear, ZZ is in general a curve rather than a line, as shown in Figure 5-1. All the arguments that follow would apply if we assumed

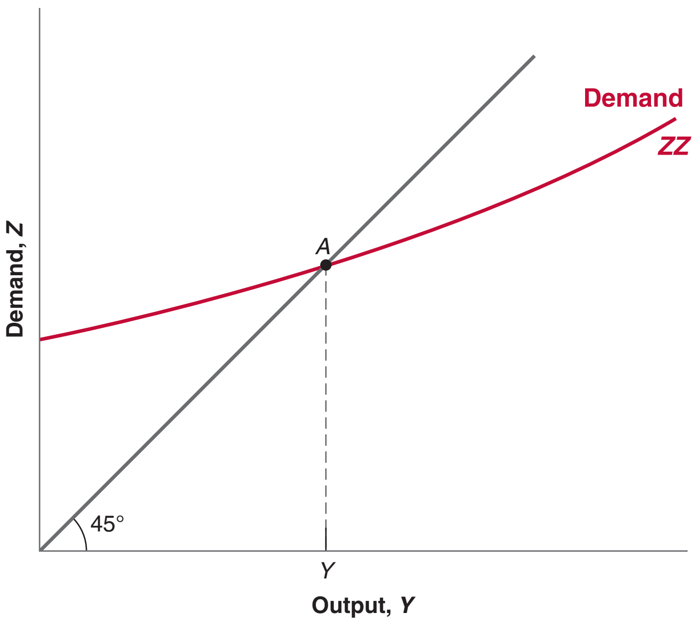


Figure 5-1

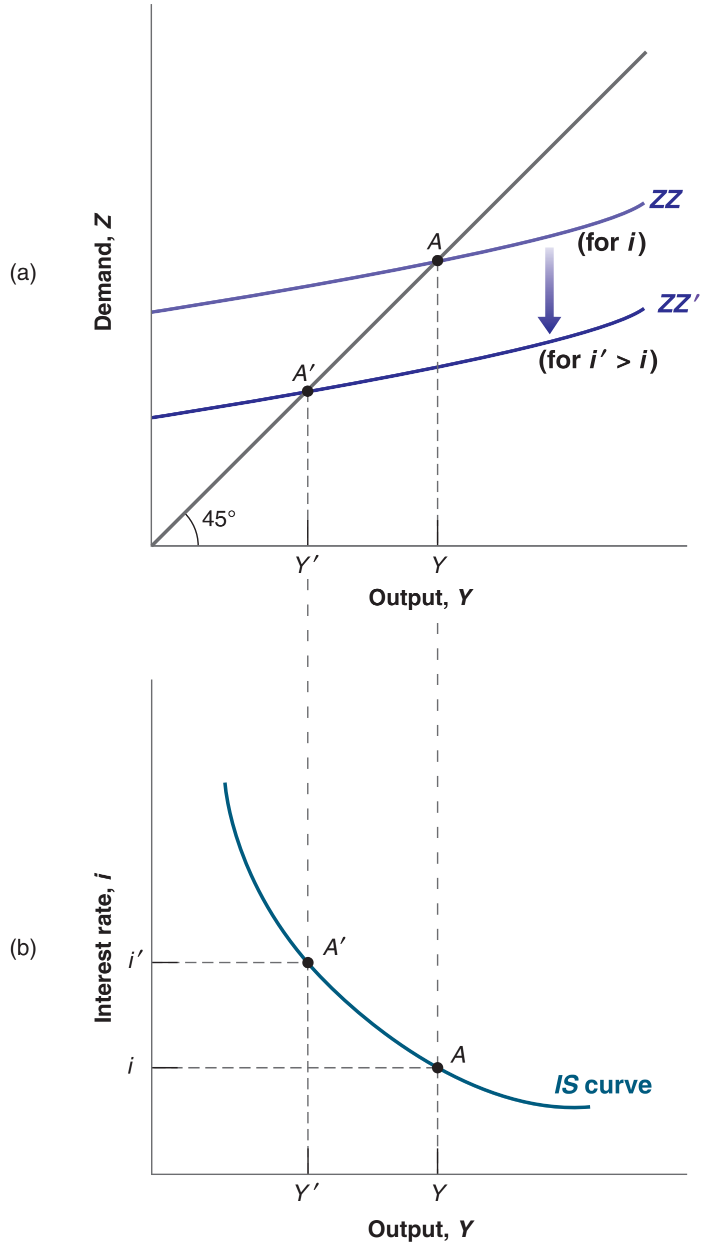
Equilibrium in the Goods Market

The demand for goods is an increasing function of output. Equilibrium requires that the demand for goods be equal to output.

Figure 5-2

The IS Curve

(a) An increase in the interest rate decreases the demand for goods at any level of output, leading to a decrease in the equilibrium level of output.
(b) Equilibrium in the goods market implies that an increase in the interest rate leads to a decrease in output. The IS curve, which gives the relation between the interest rate and output, is therefore downward sloping.



that the consumption and investment relations were linear and that ZZ were a straight line.

- We have drawn ZZ so that it is flatter than the 45-degree line. Put another, but equivalent, way, we have assumed that an increase in output leads to a less than one-for-one increase in demand. In Chapter 3, where investment was constant, this restriction naturally followed from the assumption that consumers spend only part of their additional income on consumption. But now that we allow investment to respond to production, this restriction may no longer hold. In response to an increase in output, the sum of the increase in consumption and the increase in investment could exceed the initial increase in output. Although this is a theoretical possibility, the empirical evidence suggests that it is not the case in reality. That's why we shall assume that the response of demand to output is less than one for one and draw ZZ flatter than the 45-degree line.

Make sure you understand why the two statements mean the same thing. ▶

Equilibrium in the goods market is reached at the point where the demand for goods equals output; that is, at point A, the intersection of ZZ and the 45-degree line. The equilibrium level of output is given by Y.

So far, what we have done is extend, in straightforward fashion, the analysis of Chapter 3. We are now ready to derive the IS curve.

Deriving the IS Curve

We have drawn the demand relation, ZZ, in Figure 5-1 for a given value of the interest rate. Let's now derive in Figure 5-2 what happens if the interest rate changes.

Suppose that, in Figure 5-2(a), the demand curve is given by ZZ and the initial equilibrium is at point A. Suppose now that the interest rate increases from its initial value i to i'. At any level of output, the higher interest rate leads to lower investment and lower demand. The demand curve ZZ shifts down to ZZ': At a given level of output, demand is lower. The new equilibrium is at the intersection of the lower demand curve ZZ' and the 45-degree line, at point A'. The equilibrium level of output is now equal to Y'.

In words: The increase in the interest rate decreases investment. The decrease in investment leads to a decrease in output, which further decreases consumption and investment, through the multiplier effect.

Can you show graphically the size of the multiplier? (*Hint*: Look at the ratio of the decrease in equilibrium output to the initial decrease in investment.)

Using Figure 5-2(a), we can find the equilibrium value of output associated with *any* value of the interest rate. The resulting relation between equilibrium output and the interest rate is drawn in Figure 5-2(b).

Figure 5-2(b) plots equilibrium output Y on the horizontal axis against the interest rate on the vertical axis. Point A in Figure 5-2(b) corresponds to point A in Figure 5-2(a), and point A' in Figure 5-2(b) corresponds to A' in Figure 5-2(a). The higher interest rate is associated with a lower level of output.

This relation between the interest rate and output is represented by the downward-sloping curve in Figure 5-2(b). This curve is called the **IS curve**.

Equilibrium in the goods market implies that an increase in the interest rate leads to a decrease in output. This relation is represented by the downward-sloping IS curve.

Shifts of the IS Curve

We have drawn the IS curve in Figure 5-2, taking as given the values of taxes, T, and government spending, G. Changes in either T or G will shift the IS curve.

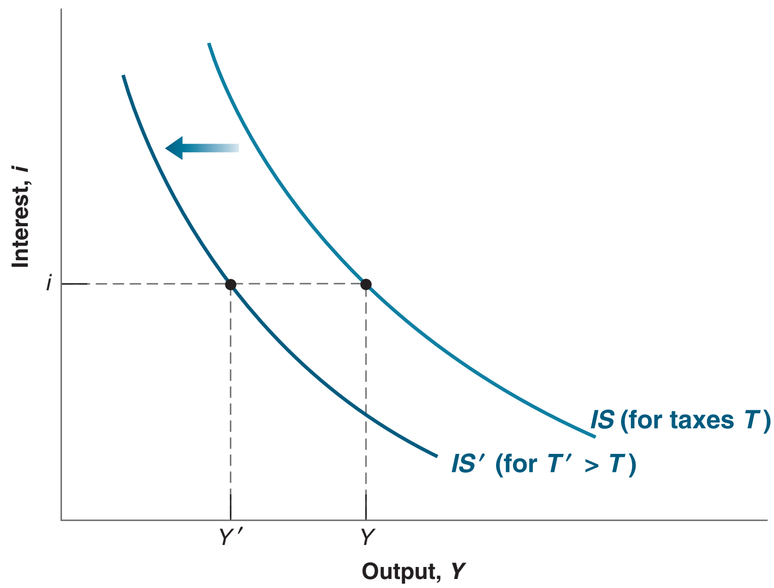


Figure 5-3

Shifts of the IS Curve

An increase in taxes shifts the IS curve to the left.

For a given interest rate, an increase in taxes leads to a decrease in output. In graphic terms: An increase in taxes shifts the IS curve to the left. ▶

To see how, consider Figure 5-3. The IS curve gives the equilibrium level of output as a function of the interest rate. It is drawn for given values of taxes and spending. Now consider an increase in taxes, from T to T'. At a given interest rate, i, disposable income decreases, leading to a decrease in consumption, leading in turn to a decrease in the demand for goods and a decrease in equilibrium output, from Y to Y'. The IS curve shifts to the left: At a given interest rate, the equilibrium level of output is lower than it was before the increase in taxes.

More generally, any factor that, for a given interest rate, decreases the equilibrium level of output causes the IS curve to shift to the left. We have looked at an increase in taxes. But the same would hold for a decrease in government spending or in consumer confidence (which decreases consumption given disposable income). Symmetrically, any factor that, for a given interest rate, increases the equilibrium level of output—a decrease in taxes, an increase in government spending, an increase in consumer confidence—causes the IS curve to shift to the right.

Suppose that the government announces that ▶ the Social Security system is in trouble and it may have to cut retirement benefits in the future. How are consumers likely to react? What is then likely to happen to demand and output?

Let's summarize:

- Equilibrium in the goods market implies that an increase in the interest rate leads to a decrease in output. This relation is represented by the downward-sloping IS curve.
- Changes in factors that decrease the demand for goods, given the interest rate, shift the IS curve to the left. Changes in factors that increase the demand for goods, given the interest rate, shift the IS curve to the right.

5-2 FINANCIAL MARKETS AND THE LM RELATION

Let's now turn to financial markets. We saw in Chapter 4 that the interest rate is determined by the equality of the supply of and demand for money:

$$M = \$Y\,L(i)$$

The variable M on the left side is the nominal money stock. We shall ignore here the details of the money-supply process that we saw in Section 4-3, and simply think of the central bank as controlling M directly.

The right side gives the demand for money, which is a function of nominal income, $\$Y$, and of the nominal interest rate, i. As we saw in Section 4-1, an increase in nominal income increases the demand for money; an increase in the interest rate decreases the demand for money. Equilibrium requires that money supply (the left side of the equation) be equal to money demand (the right side of the equation).

Real Money, Real Income, and the Interest Rate

From Chapter 2: Nominal GDP = Real GDP multiplied by the GDP deflator $\$Y = YP$. Equivalently: ▶ Real GDP = Nominal GDP divided by the GDP deflator $\$Y/P = Y$.

The equation $M = \$Y\,L(i)$ gives a relation between money, nominal income, and the interest rate. It will be more convenient here to rewrite it as a relation among real money (that is, money in terms of goods), real income (that is, income in terms of goods), and the interest rate.

Recall that nominal income divided by the price level equals real income, Y. Dividing both sides of the equation by the price level P gives

$$\frac{M}{P} = Y\,L(i) \qquad (5.3)$$

Hence, we can restate our equilibrium condition as the condition that the *real money supply*—that is, the money stock in terms of goods, not dollars—be equal to the *real money demand*, which depends on real income, Y, and the interest rate, i.

The notion of a "real" demand for money may feel a bit abstract, so an example will help. Instead of thinking of your demand for money in general, think of your demand for coins. Suppose you like to have coins in your pocket to buy two cups of coffee during the day. If a cup costs $1.20, you will want to keep about $2.40 in coins. This is your nominal demand for coins. Equivalently, you want to keep enough coins in your pocket to buy two cups of coffee. This is your demand for coins in terms of goods—here in terms of cups of coffee.

From now on, we shall refer to equation (5.3) as the LM *relation*. The advantage of writing things this way is that *real income*, Y, appears on the right side of the equation instead of *nominal income*, $\$Y$. And real income (equivalently real output) is the variable we focus on when looking at equilibrium in the goods market. To make the reading lighter, we will refer to the left and right sides of equation (5.3) simply as "money supply" and "money demand" rather than the more accurate but heavier "real money supply" and "real money demand." Similarly, we will refer to income rather than "real income."

Deriving the LM Curve

In deriving the IS curve, we took the two policy variables as government spending, G, and taxes, T. In deriving the LM curve—the curve corresponding to the LM relation—we have to decide how we characterize monetary policy: as the choice of M, the money stock, or the choice of i, the interest rate.

If we think of monetary policy as choosing the nominal money supply, M, and, by implication, given the price level that we shall take as fixed in the short run, choosing M/P, the real money stock, equation (5.3) tells us that real money demand, the right-hand side of the equation, must be equal to the *given* real money supply, the left-hand side of the equation. Thus, if, for example, real income increases, increasing money demand, the interest rate must increase so that money demand remains equal to the given money supply. In other words, for a given money supply, an increase in income automatically leads to an increase in the interest rate.

◀ Go back to Figure 4-4 in the previous chapter.

This is the traditional way of deriving the LM relation and the resulting LM curve. As we discussed in Chapter 4, however, the assumption that the central bank chooses the money stock and then just lets the interest rate adjust is at odds with reality today. Although, in the past, central banks thought of the money supply as the monetary policy variable, they now focus directly on the interest rate. They choose an interest rate, call it $\bar{i}$, and adjust the money supply so as to achieve it. Thus, in the rest of the book, we shall think of the central bank as choosing the interest rate (and doing what it needs to do with the money supply to achieve this interest rate). This will make for an extremely simple LM curve, namely, a horizontal line (in Figure 5-4) at the value of the interest rate, $\bar{i}$, chosen by the central bank.

◀ LM curve is a bit of a misnomer as, under our assumption, the LM relation is a simple horizontal line. But the use of the term *curve* is traditional, and I shall follow tradition.

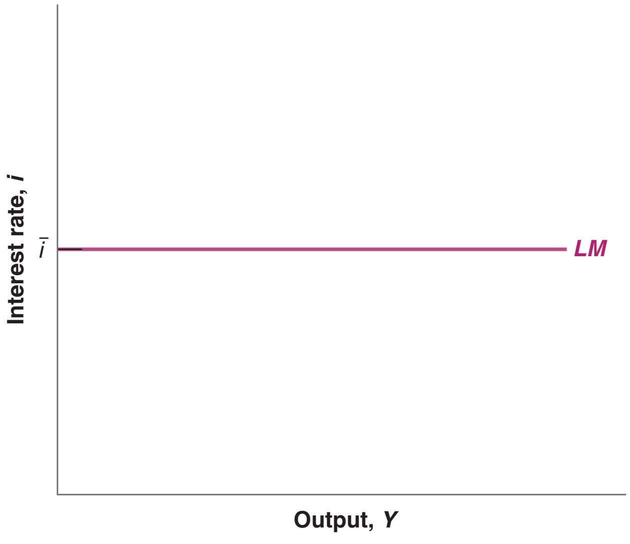


Figure 5-4

The LM Curve

The central bank chooses the interest rate (and adjusts the money supply so as to achieve it).

5-3 PUTTING THE IS AND LM RELATIONS TOGETHER

The IS relation follows from goods market equilibrium. The LM relation follows from financial market equilibrium. They must both hold.

$$\text{IS relation: } Y = C(Y - T) + I(Y, i) + G$$
$$\text{LM relation: } i = \bar{i}$$

Together they determine output. Figure 5-5 plots both the IS curve and the LM curve on one graph. Output—equivalently, production or income—is measured on the horizontal axis. The interest rate is measured on the vertical axis.

Any point on the downward-sloping IS curve corresponds to equilibrium in the goods market. *Any point* on the horizontal LM curve corresponds to equilibrium in financial markets. *Only at point A* are both equilibrium conditions satisfied. That means point *A*, with the associated level of output *Y* and interest rate $\bar{i}$, is the overall equilibrium—the point at which there is equilibrium in both the goods market and the financial markets.

You may ask: So what if the equilibrium is at point *A*, how does this fact translate into anything directly useful about the world? Don't despair: Figure 5-5 holds the answer to many questions in macroeconomics. The IS and LM relations that underlie the figure contain a lot of information about consumption, investment, and equilibrium conditions. Used properly, Figure 5-5 allows us to study what happens to output when the central bank decides to decrease the interest rate, or when the government decides to increase taxes, or when consumers become more pessimistic about the future, and so on.

In future chapters, you will see how we can extend it to think about the financial crisis, or about the role of expectations, or about the role of policy in an open economy.

Let's now see what the IS-LM model tells us, by looking separately at the effects of fiscal and monetary policy.

Fiscal Policy

Decrease in G-T ⇔ fiscal contraction ⇔ fiscal consolidation

Increase in G-T ⇔ fiscal expansion

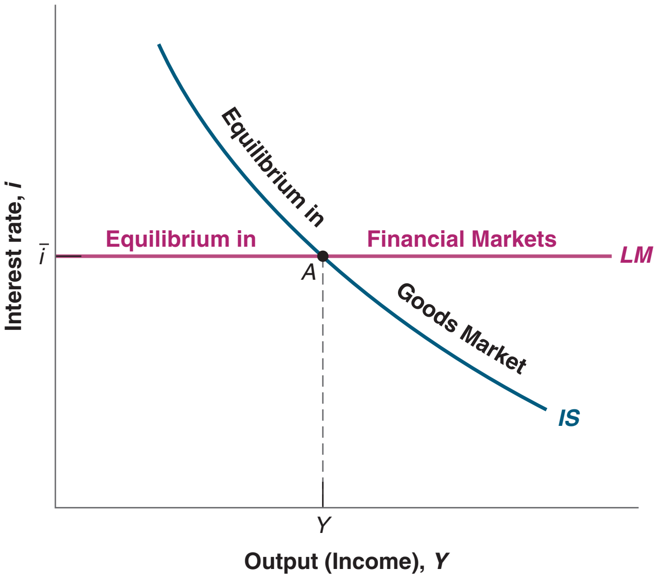
A reduction in the budget deficit, achieved either by increasing taxes, or by decreasing spending, or both, is called a **fiscal contraction** or a **fiscal consolidation**. Symmetrically, an increase in the budget deficit, achieved either by decreasing taxes, or by increasing spending, or both, is called a **fiscal expansion**.

Suppose the government decides to reduce the fiscal deficit, and to achieve this fiscal contraction through an increase in taxes. What will be the effects on output, on its composition, and on the interest rate?

Figure 5-5

The IS-LM Model

Equilibrium in the goods market implies that an increase in the interest rate leads to a decrease in output. This is represented by the IS curve. Equilibrium in financial markets is represented by the horizontal LM curve. Only at point *A*, which is on both curves, are both goods and financial markets in equilibrium.



When you answer this or any question about the effects of changes in policy (or, more generally, changes in exogenous variables), always go through the following three steps:

1. Ask how the change affects equilibrium in the goods market and how it affects equilibrium in the financial markets. Put another way: Does it shift the IS curve and/or the LM curve, and, if so, how?
2. Characterize the effects of these shifts on the intersection of the IS and LM curves. What does this do to equilibrium output and the equilibrium interest rate?
3. Describe the effects in words.

With time and experience, you will often be able to go directly to step 3. By then you will be ready to give an instant commentary on the economic events of the day. But until you get to that level of expertise, go step by step.

And when you feel really confident, put on a bow tie and go explain events on TV. (Why so many TV economists wear bow ties is a mystery.)

In this case, the three steps are easy. But going through them is good practice.

- Start with step 1. The first question is how the increase in taxes affects equilibrium in the goods market—that is, how it affects the relation between output and the interest rate captured in the IS curve. We derived the answer in Figure 5-3: At a given interest rate, the increase in taxes decreases output. The IS curve shifts to the left, from *IS* to *IS′* in Figure 5-6.

  Next, let's see if anything happens to the LM curve. By assumption, as we are looking at a change only in fiscal policy, the central bank does not change the interest rate. Thus, the LM curve, i.e., the horizontal line at $i = \bar{i}$, remains unchanged—it does not shift.
- Now consider step 2, the determination of the equilibrium.

  Before the increase in taxes, the equilibrium is given by point *A*, at the intersection of the *IS* and LM curves. After the increase in taxes and the shift to the left of the IS curve from *IS* to *IS′*, the new equilibrium is given by point *A′*. Output decreases from *Y* to *Y′*. By assumption, the interest rate does not change. Thus, as the IS curve *shifts*, the economy *moves along* the LM curve, from *A* to *A′*. The reason these words are italicized is that it is important always to distinguish between the *shift of* a curve (here the shift of the IS curve) and *movement along* a curve (here the movement along the LM curve). Many mistakes result from not distinguishing between the two.

The increase in taxes shifts the IS curve. The LM curve does not shift. The economy moves along the LM curve.

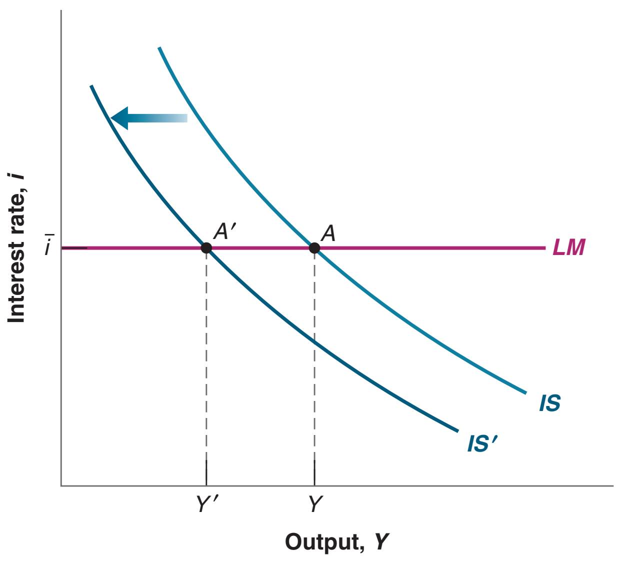


Figure 5-6

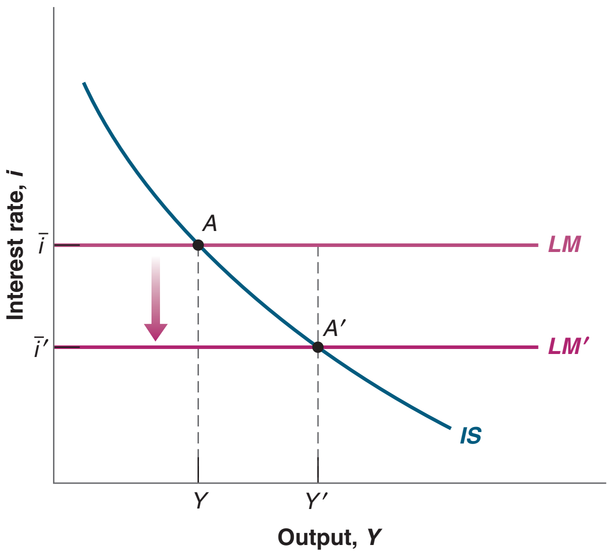
The Effects of an Increase in Taxes

An increase in taxes shifts the IS curve to the left. This leads to a decrease in the equilibrium level of output.

Figure 5-7

The Effects of a Decrease in the Interest Rate

A monetary expansion shifts the LM curve down and leads to higher output.



Note that we have just given a formal treatment of the informal discussion of the effects of an increase in public saving given in the Focus Box on "The Paradox of Saving" in Chapter 3.

- Step 3 is to tell the story in words:
  At a given interest rate, the increase in taxes leads to lower disposable income, which causes people to decrease their consumption. This decrease in demand leads, through a multiplier, to a decrease in output and income and, by implication, a decrease in investment.

Monetary Policy

Decrease in $i \Leftrightarrow$ increase in $M \Leftrightarrow$ monetary expansion.

Increase in $i \Leftrightarrow$ decrease in $M \Leftrightarrow$ monetary contraction (or monetary tightening).

Now we turn to monetary policy. Suppose the central bank decreases the interest rate. Recall that, to do so, it increases the money supply. Such a change in monetary policy is called a **monetary expansion**. (Conversely, an increase in the interest rate, which is achieved through a decrease in the money supply, is called a **monetary contraction** or **monetary tightening**.)

Make sure you understand why the IS curve does not shift.

- Again, step 1 is to see whether and how the IS and the LM curves shift.
  Let's look at the IS curve first. The change in the interest rate does not change the relation between output and the interest rate. It does not shift the IS curve.
  The change in the interest rate, however, leads (trivially) to a shift in the LM curve: it shifts down, from the horizontal line at $i = \bar{i}$ to the horizontal line $i = \bar{i}'$.
- Step 2 is to see how these shifts affect the equilibrium, represented in Figure 5-7. The economy moves down along the IS curve, and the equilibrium moves from point A to point A'. Output increases from Y to Y', and the interest rate decreases from i to i'.
- Step 3 is to say it in words: The lower interest rate leads to an increase in investment and, in turn, an increase in demand and output. Looking at the components of output: The increase in output and the decrease in the interest rate both lead to an increase in investment. The increase in income leads to an increase in disposable income and, in turn, consumption, so both consumption and investment increase.

5-4 USING A POLICY MIX

We have looked so far at fiscal policy and monetary policy in isolation. Our purpose was to show how each worked. In practice, the two are often used together. The combination of monetary and fiscal policies is known as the **monetary-fiscal policy mix**, or simply the *policy mix*.

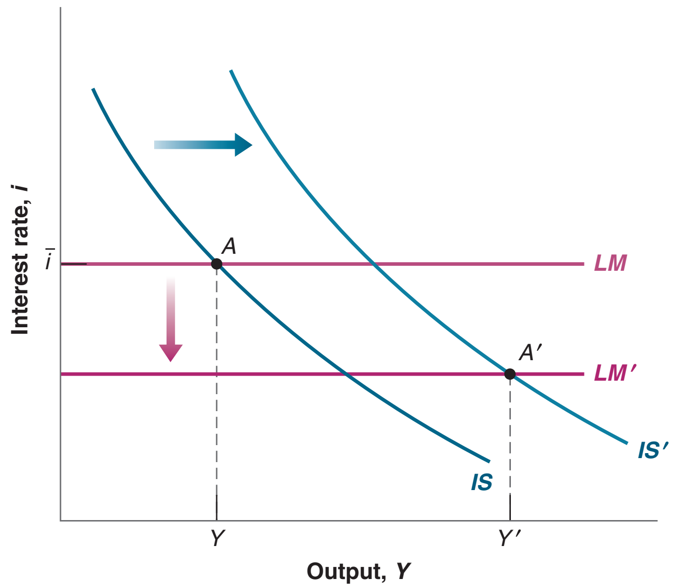


Figure 5-8

The Effects of a Combined Fiscal and Monetary Expansion

Fiscal expansion shifts the IS curve to the right. Monetary expansion shifts the LM curve down. Both lead to higher output.

Sometimes, the right mix is to use fiscal and monetary policy in the same direction. Suppose, for example, that the economy is in a recession and output is too low. Then, both fiscal and monetary policies can be used to increase output. This combination is represented in Figure 5-8. The initial equilibrium is given by the intersection of *IS* and *LM* at point *A*, with corresponding output *Y*. Expansionary fiscal policy, say through a decrease in taxes, shifts the IS curve to the right, from *IS* to *IS′*. Expansionary monetary policy shifts the LM curve down from *LM* to *LM′*. The new equilibrium is at *A′*, with corresponding output *Y′*. Thus, both fiscal and monetary policies contribute to the increase in output: Lower taxes and higher income result in higher consumption, which in turn leads to higher output and, together with a lower interest rate, higher investment.

Such a combination of fiscal and monetary policy is typically used to fight recessions. An example is given in the Focus Box "The US Recession of 2001."

We shall see other examples later in the book. Section 6-5 looks at the role of fiscal and monetary policy during the Great Financial Crisis.

You might ask: Why use both policies when either one on its own could achieve the desired increase in output? An increase in output could in principle be achieved just by using fiscal policy—say through a sufficiently large increase in government spending, or a sufficiently large decrease in taxes—or just by using monetary policy, through a sufficiently large decrease in the interest rate. The answer is that there are a number of reasons why policymakers may want to use a policy mix.

- A fiscal expansion means either an increase in government spending, or a decrease in taxes, or both. This means an increase in the budget deficit (or, if the budget was initially in surplus, a smaller surplus). As we shall see later, but you surely can guess why already, running a large deficit and increasing government debt could be dangerous later. In this case, it is better to rely, at least in part, on monetary policy.

More on this in Chapter 22.

- A monetary expansion means a decrease in the interest rate. If the interest rate is very low, then the room for using monetary policy may be limited. In this case, fiscal policy has to do more of the job. If the interest rate is already equal to zero because of *the zero lower bound*, fiscal policy has to do all the work. As we saw in Chapter 1, while US interest rates have become positive, they remain low. If demand were to decrease soon, the room for monetary policy to decrease interest rates would be limited, and fiscal policy would have to play the main role.

The US Recession of 2001

In 1992, the US economy embarked on a long expansion. For the rest of the decade, GDP growth was positive and high. In 2000, however, the expansion came to an end. From the third quarter of 2000 to the fourth quarter of 2001, GDP growth was either positive and close to zero, or negative. Based on data available at the time, it was thought that growth was negative through the first three quarters of 2001. Based on revised data, shown in Figure 1, which gives the growth rate for each quarter from 1999Q1 to 2002Q4 measured at an annual rate, it appears that growth was actually small but positive in the second quarter. (These data revisions happen often, so that what we see when we look back is not always what national income statisticians and policymakers perceived at the time.) The National Bureau of Economic Research (NBER), an academic organization that has traditionally dated US recessions and expansions, concluded that the US economy had indeed had a recession in 2001, starting in March 2001 and ending in December 2001; this period is represented by the shaded area in the figure.

What triggered the recession was a sharp decline in investment demand. Nonresidential investment—the demand for plant and equipment by firms—decreased by 4.5% in 2001. The cause was the end of what Alan Greenspan, the chairman of the Fed at the time, had dubbed a period of "irrational exuberance": During the second part of the 1990s, firms had been extremely optimistic about the future, and the rate of investment had been very high—the average yearly growth rate of investment from 1995 to 2000 exceeded 10%. In 2001, however, it became clear to firms that they had been overly optimistic and had invested too much. This led them to cut back on investment, leading to a decrease in demand and, through the multiplier, a decrease in GDP.

The recession could have been much worse. But its depth and the length were limited by a strong macroeconomic policy response.

Take monetary policy first. Starting in early 2001, the Fed, recognizing that the economy was slowing down, started decreasing the federal funds rate aggressively. (Figure 2 shows the behavior of the federal funds rate from 1991Q1 to 2002Q4.) It continued to do so throughout the year. The rate, which stood at 6.5% in January, was less than 2% at the end of the year.

Turn to fiscal policy. During the 2000 presidential campaign, then candidate George Bush ran on a platform of lower taxes. The argument was that the federal budget was in surplus, and so there was room to reduce tax rates while keeping the budget in balance. When President Bush took office in 2001 and it became clear that the economy was slowing down, he had an additional rationale to cut tax rates, namely to increase demand and fight the recession. Both the 2001 and 2002 budgets included substantial reductions in tax rates. On the spending side, the events of September 11, 2001 also led to an increase in spending, mostly on defense and homeland security.

Figure 3 shows federal government revenues and spending from 1999Q1 to 2002Q4, both expressed as ratios to GDP. Note the dramatic decrease in revenues starting in the third quarter of 2001. Even without decreases in tax rates, revenues would have gone down during the recession: Lower output and lower income mechanically imply lower tax revenues. But, because of the tax cuts, the decrease in revenues in 2001 and 2002 was much larger than can be explained by the recession. Note also the smaller but steady increase in spending starting around the same time. As a result, the budget surplus—the difference

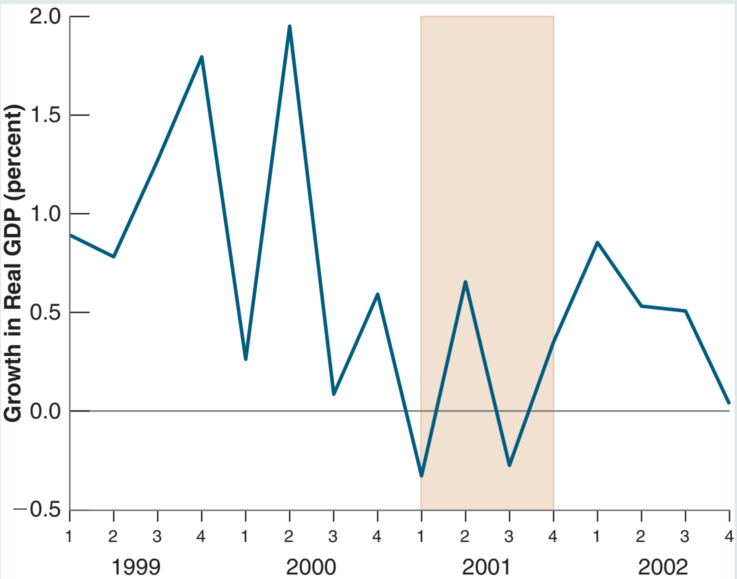


Figure 1

The US Growth Rate, 1999Q1 to 2002Q4

Source: Data from Calculated using Series GDPC1, Federal Reserve Economic Data (FRED) http://research.stlouisfed.org/fred2/

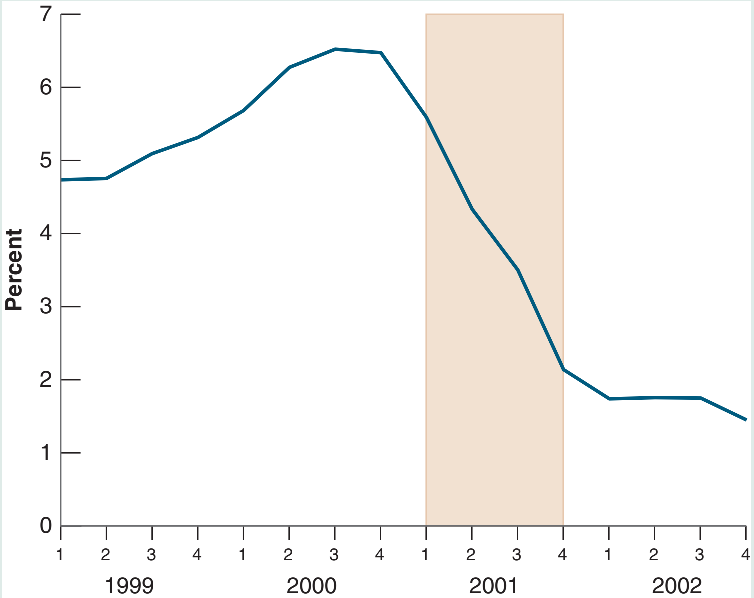


Figure 2

The Federal Funds Rate, 1999Q1 to 2002Q4

Source: Data from Calculated using Series GDP, FGRECPY, FGEXPND, Federal Reserve Economic Data (FRED) http://research.stlouisfed.org/fred2/

between revenues and spending—went from positive up until 2000, to negative in 2001 and much more so in 2002.

Let me end by taking up four questions you might be asking yourself at this point:

- Why weren't monetary and fiscal policy used to avoid rather than just limit the size of the recession? The reason is that changes in policy affect demand and output only over time (more on this in Section 5-5). Thus, by the time it became clear that the US economy was entering a recession, it was already too late to use policy to avoid it. But the policy did reduce both the depth and length of the recession.
- Weren't the events of September 11, 2001, also a cause of the recession? The answer, in short, is no, tragic as

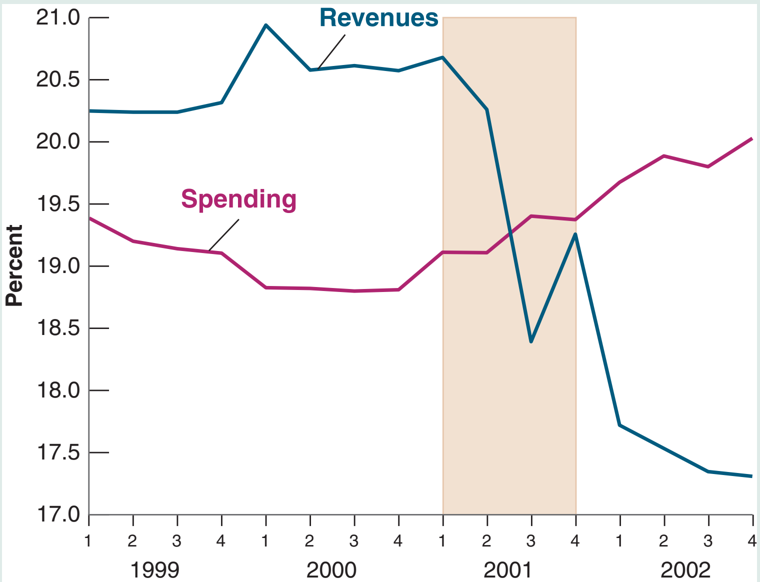


Figure 3

US Federal Government Revenues and Spending (as Ratios to GDP), 1999Q1 to 2002Q4

Source: Data from Calculated using Series GDP, FGRECPY, FGEXPND, Federal Reserve Economic Data (FRED) http://research.stlouisfed.org/fred2/

the event was. As we have seen, the recession started long before September 11, and ended soon after. Indeed, GDP growth was positive in the last quarter of 2001. One might have expected—and, indeed, most economists expected—the events of September 11 to have large adverse effects on output, leading, in particular, consumers and firms to delay spending decisions until the outlook was clearer. In fact, the drop in spending was short-lived and limited. Decreases in the federal funds rate after September 11—and large discounts by automobile producers in the last quarter of 2001—are believed to have been crucial in maintaining consumer confidence and consumer spending during that period.

- Was the monetary-fiscal mix used to fight the recession a textbook example of how policy should be conducted?

On this, economists differ. Most economists give high marks to the Fed for strongly decreasing interest rates as soon as the economy slowed down. But many economists are worried that the tax cuts introduced in 2001 and 2002 led to large budget deficits that lasted long after the recession was over. They argue that the tax cuts should have been temporary, helping the US economy get out of the recession but stopping thereafter.

- Why were monetary and fiscal policy unable to avoid the recession of 2009? The answer, in short, is twofold. The shocks were much larger, and much harder to react to. And the room for policy responses was more limited. We shall return to these two aspects in Chapter 6.

- Fiscal and monetary policies have different effects on the composition of output. A decrease in income taxes, for example, tends to increase consumption relative to investment. A decrease in the interest rate affects investment more than consumption. Thus, depending on the initial composition of output, policymakers may want to rely more on fiscal or more on monetary policy.
- Finally, neither fiscal policy nor monetary policy works perfectly. A decrease in taxes may fail to increase consumption. A decrease in the interest rate may fail to increase investment. Thus, in case one policy does not work as well as hoped for, it is better to use both.

Sometimes, the right policy mix is instead to use the two policies in opposite directions, for example, combining a fiscal consolidation with a monetary expansion. Suppose, for example, that the government is running a large budget deficit and would like to reduce it, but does not want to trigger a recession. In Figure 5-9, the initial equilibrium is given by the intersection of the IS and LM curves at point A, with associated output Y. Output is thought to be at the right level, but the budget deficit, T-G, is too large.

If the government reduces the deficit, say by increasing T or by decreasing G (or both), the IS curve will shift to the left, from IS to IS'. The equilibrium will be at point A', with level of output Y'. At a given interest rate, higher taxes or lower spending

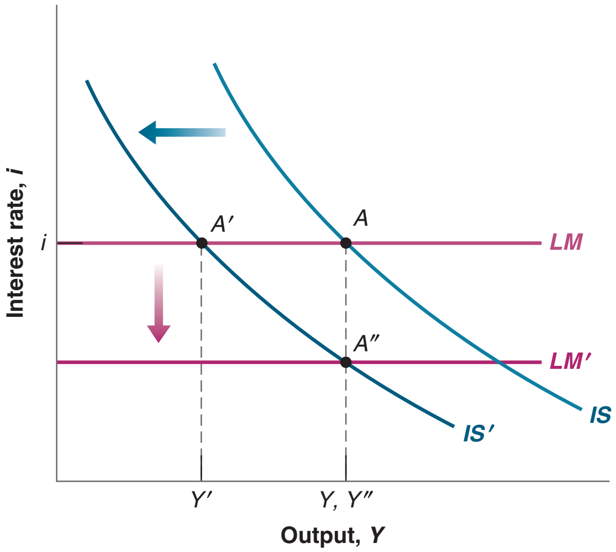


Figure 5-9

The Effects of a Combined Fiscal Consolidation and a Monetary Expansion

The fiscal consolidation shifts the IS curve to the left. A monetary expansion shifts the LM curve down. Together, they leave output unchanged while the budget deficit is reduced.

Deficit Reduction: Good or Bad for Investment?

FOCUS

You may have heard this argument in some form: "Private saving goes either toward financing the budget deficit or financing investment. It does not take a genius to conclude that reducing the budget deficit leaves more saving available for investment, so investment increases."

This argument sounds convincing. But, as we have seen earlier in the book, it must be wrong. If, for example, deficit reduction is not accompanied by a decrease in the interest rate, then we know that output decreases (see Figure 5-7) and, by implication, so does investment since it depends on output. So what is going on in this case?

Go back to Chapter 3, equation (3.10). There we learned that we can also think of the goods-market equilibrium condition as

$$\text{Investment} = \text{Private saving} + \text{Public saving}$$
$$I = S + (T - G)$$

In equilibrium, investment is equal to private saving plus public saving. If public saving is positive, the government is said to be running a budget surplus; if public saving is negative, the government is said to be running a budget deficit. So it is true that *given private saving*, if the government reduces its deficit—either by increasing taxes or reducing government spending so that *T-G* goes up—investment must go up: Given *S*, *T-G* going up implies that *I* goes up.

The crucial part of this statement, however, is "given private saving." The point is that a fiscal contraction affects private saving as well: The contraction leads to lower output and therefore lower income. As consumption goes down by less than income, private saving also goes down. It actually goes down by more than the reduction in the budget deficit, leading to a decrease in investment. In terms of the equation: *S* decreases by more than *T-G* increases, and so *I* decreases. (You may want to do the algebra and convince yourself that saving actually goes down by *more* than the increase in *T-G*.)

Does this mean that deficit reduction always decreases investment? The answer is clearly *no*. We saw this in Figure 5-9. If, when the deficit is reduced, the central bank also decreases the interest rate so as to keep output constant, then investment necessarily goes up. Although output is unchanged, the lower interest rate leads to higher investment.

The moral of this box is clear: Whether deficit reduction leads to an increase in investment is far from automatic. It may or it may not, depending on the response of monetary policy.

will decrease demand, and through the multiplier, decrease output. Thus, the reduction in the deficit will lead to a recession.

The recession can be avoided, however, if monetary policy is also used. If the central bank reduces the interest rate to $\bar{i}'$, the equilibrium is given by point A'' with corresponding output $Y'' = Y$. The combination of both policies thus allows for the reduction in the deficit, but without a recession.

What happens to consumption and investment in this case? What happens to consumption depends on how the deficit is reduced. If the reduction takes the form of a decrease in government spending rather than an increase in taxes, income is unchanged, disposable income is unchanged, and so consumption is unchanged. If the reduction takes the form of an increase in income taxes, then disposable income is lower, and so is consumption. What happens to investment is unambiguous: Unchanged output and a lower interest rate imply higher investment. The relation between deficit reduction and investment is discussed further in the Focus Box "Deficit Reduction: Good or Bad for Investment?"

Such a policy mix was used in the early 1990s. When Bill Clinton was elected president in 1992, one of his priorities was to reduce the budget deficit using a combination of cuts in spending and increases in taxes. He was worried, however, that, by itself, such a fiscal contraction would lead to a decrease in demand and trigger another recession. The right strategy was to combine a fiscal contraction (to get rid of the deficit) with a monetary expansion (to make sure that demand and output remained high). This was the strategy adopted and carried out by Clinton (who was in charge of fiscal policy) and Alan Greenspan (who was in charge of monetary policy). The result of this strategy—and a bit of economic luck—was a steady reduction of the budget deficit (which turned into a budget surplus at the end of the 1990s) and a steady increase in output throughout the rest of the decade.

A similar discussion is taking place today in the euro area but with a twist. Worried about high public debt levels, governments would like to reduce fiscal deficits to decrease debt levels over time, a policy known as **fiscal austerity**. The problem, however, is that interest rates are already very low, so there is little room for monetary policy to offset the adverse effects of a fiscal contraction on output. This is leading to an intense debate between those who think that fiscal austerity is needed, even if it has adverse effects on output, and those who think that fiscal consolidation should wait until monetary policy can be used to offset its adverse effect on output.

5-5 HOW DOES THE IS-LM MODEL FIT THE FACTS?

We have so far ignored dynamics. For example, when looking at the effects of an increase in taxes in Figure 5-6—or the effects of a monetary expansion in Figure 5-7—we made it look as if the economy moved instantaneously from A to A', as if output went instantaneously from Y to Y'. This is clearly not realistic: The adjustment of output takes time. To capture this time dimension, we need to reintroduce dynamics.

Introducing dynamics formally would be difficult. But, as we did in Chapter 3, we can describe the basic mechanisms in words. Some of the mechanisms will be familiar from Chapter 3, some are new:

- Consumers are likely to take some time to adjust their consumption following a change in disposable income.
- Firms are likely to take some time to adjust investment spending following a change in their sales.
- Firms are likely to take some time to adjust investment spending following a change in the interest rate.
- Firms are likely to take some time to adjust production following a change in their sales.

So, in response to an increase in taxes, it takes some time for consumption spending to respond to the decrease in disposable income, some more time for production to decrease in response to the decrease in consumption spending, yet more time for investment to decrease in response to lower sales, and so on.

In response to the decrease in the interest rate, it takes some time for investment spending to respond, some more time for production to increase in response to the increase in demand, yet more time for consumption and investment to increase in response to the induced change in output, and so on.

Describing precisely the adjustment process implied by all these sources of dynamics is complicated, but the basic implication is straightforward: Time is needed for output to adjust to changes in fiscal and monetary policy. How much time? This question can be answered only by looking at the data and using econometrics. Figure 5-10 shows the results of such an econometric study, which uses data from the United States from 1960 to 1990.

We discussed the federal funds market and the federal funds rate in Chapter 4, Section 4–3.

The study looks at the effects of a decision by the Fed to increase the federal funds rate by 1%. It traces the typical effects of such an increase on a number of macroeconomic variables.

Each panel in Figure 5-10 represents the effects of the change in the interest rate on a given variable: retail sales, output, employment, unemployment, and prices. Each panel plots three lines: the solid line in the center of a band gives the best estimate of the

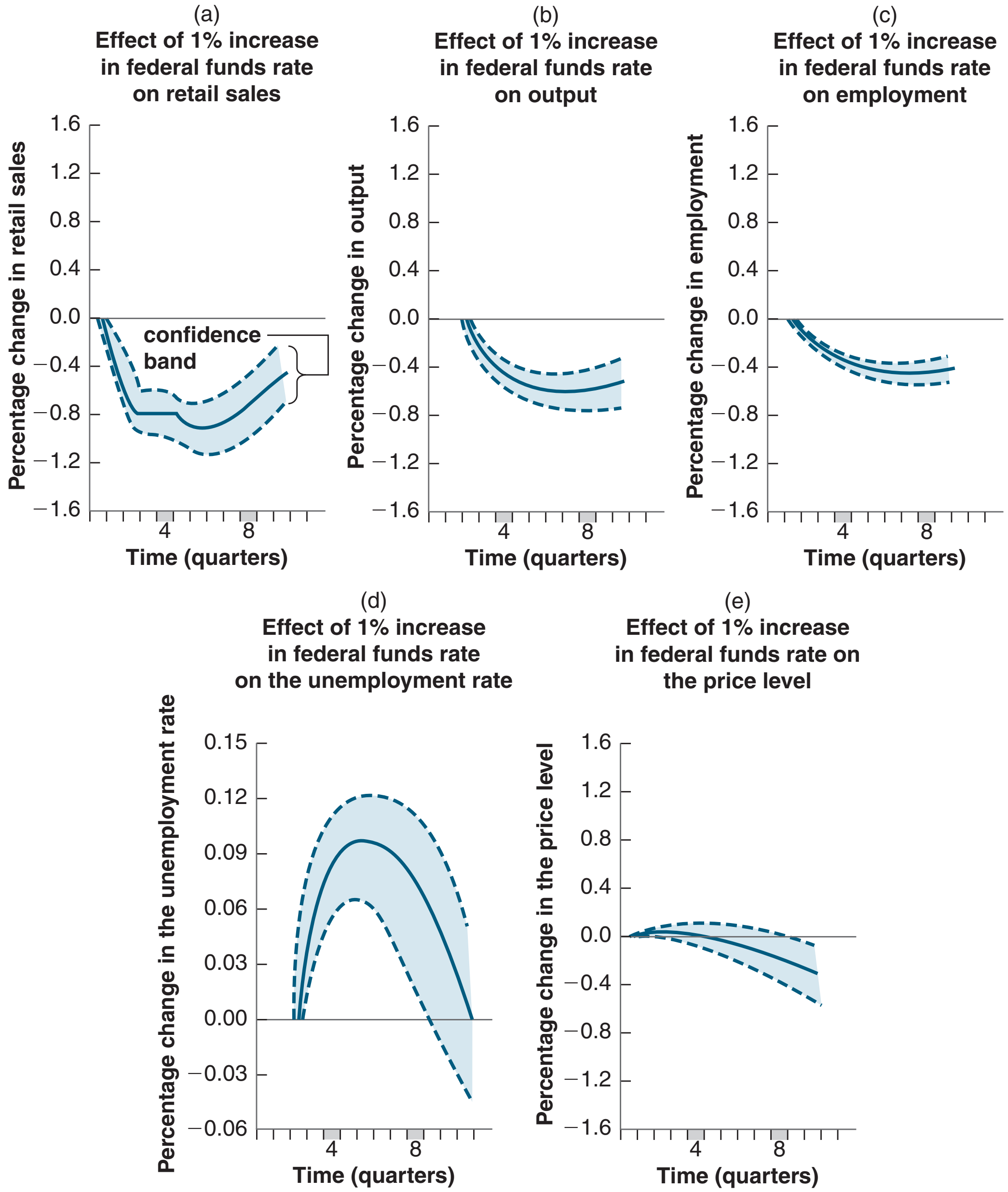


Figure 5-10

The Empirical Effects of an Increase in the Federal Funds Rate

In the short run, an increase in the federal funds rate leads to a decrease in sales, output, and employment and an increase in unemployment, but it has little effect on the price level.

Source: Lawrence Christiano, Martin Eichenbaum, and Charles Evans, "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds," *Review of Economics and Statistics*, 1996, 78 (February): pp. 16–34.

effect of the interest rate change on the variable, and the two dashed lines and tinted space between them represent a **confidence band**, within which the true value of the effect lies with 60% probability.

There is no such thing in econometrics as learning the exact value of a coefficient or the exact effect of one variable on another. Rather, econometrics provides a best estimate (here, the solid line) and a measure of confidence in the estimate (the confidence band).

- Figure 5-10(a) shows the effects of an increase in the federal funds rate of 1% on retail sales over time (12 quarters). The percentage change in retail sales is plotted on the vertical axis; time, measured in quarters, is on the horizontal axis.
  Focusing on the best estimate—the solid line—we see that the increase in the federal funds rate of 1% leads to a decline in retail sales. The largest decrease in retail sales, −0.9%, occurs after five quarters.
- Figure 5-10(b) shows that lower sales lead to lower output. In response to the decrease in sales, firms cut production, but by less than the decrease in sales. This is

This explains why monetary policy could not have prevented the 2001 recession (see the Focus Box "The US Recession of 2001"). When the Fed started decreasing the federal funds rate at the start of 2001, it was already too late for these cuts to have much effect that year.

because firms accumulate inventories for some time, so the adjustment of production is smoother and slower than the adjustment of sales. The largest decrease, −0.7%, is reached after eight quarters.

In other words, monetary policy works, but with long lags. It takes nearly two years for monetary policy to have its full effect on output.

- Figure 5-10(c) shows that lower output leads to lower employment: As firms cut production, they also cut employment. As with output, the decline in employment is slow and steady, reaching −0.5% after eight quarters. The decline in employment leads to an increase in the unemployment rate, shown in Figure 5-10(d).
- Figure 5-10(e) looks at the behavior of the price level. Remember, one of the *assumptions* of the IS-LM model is that the price level is given, and so it does not change in response to changes in demand. Figure 5-10(e) shows that this assumption is not a bad approximation of reality in the short run. The price level is nearly unchanged for the first six quarters or so. Only after the first six quarters does it decline. This gives us a strong hint as to why the IS-LM model becomes less reliable as we look at the medium run: In the medium run, we can no longer assume that the price level is given, and movements in the price level become important.

Figure 5-10 provides two important lessons. First, it gives a sense of the dynamic adjustment of output and other variables to monetary policy.

Second, and more fundamentally, it shows that what we observe in the economy is consistent with the implications of the IS-LM model. This does not *prove* that the IS-LM model is the right model. It may be that what we observe in the economy is the result of a completely different mechanism, and the fact that the IS-LM model fits well is a coincidence. But this seems unlikely. The IS-LM model looks like a solid basis on which to build when looking at movements in activity in the short run. Later on, we shall extend the model to look at the role of expectations (Chapters 14 to 16) and the implications of openness in goods and financial markets (Chapters 17 to 20). But we must first understand what determines output in the medium run. This is the topic of Chapters 7 to 9.

SUMMARY

- The IS-LM model characterizes the implications of equilibrium in both the goods and financial markets.
- The IS relation and the IS curve show the combinations of the interest rate and the level of output that are consistent with equilibrium in the goods market. An increase in the interest rate leads to a decline in output. Consequently, the IS curve is downward sloping.
- The LM relation and the LM curve show the combinations of the interest rate and the level of output consistent with equilibrium in financial markets. Under the assumption that the central bank chooses the interest rate, the LM curve is a horizontal line at the interest rate chosen by the central bank.
- A fiscal expansion shifts the IS curve to the right, leading to an increase in output. A fiscal contraction shifts the IS curve to the left, leading to a decrease in output.
- A monetary expansion shifts the LM curve down, leading to a decrease in the interest rate and an increase in output. A monetary contraction shifts the LM curve up, leading to an increase in the interest rate and a decrease in output.
- The combination of monetary and fiscal policies is known as the monetary-fiscal policy mix, or simply the policy mix. Sometimes monetary and fiscal policy are used in the same direction, sometimes in opposite directions. Together, a fiscal contraction and a monetary expansion can, for example, achieve a decrease in the budget deficit while avoiding a decrease in output.
- The IS-LM model appears to describe well the behavior of the economy in the short run. In particular, the effects of monetary policy are similar to those implied by the IS-LM model once dynamics are introduced in the model. An increase in the interest rate due to a monetary contraction leads to a decrease in output, with the maximum effect taking place after about eight quarters.

KEY TERMS

IS curve, 91
LM curve, 93
fiscal contraction, 94
fiscal consolidation, 94
fiscal expansion, 94
monetary expansion, 96
monetary contraction, 96
monetary tightening, 96
monetary-fiscal policy mix, 96
fiscal austerity, 102
confidence band, 103

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The main determinants of investment are the level of sales and the interest rate.
b. If all the exogenous variables in the IS relation are constant, then a higher level of output can be achieved only by lowering the interest rate.
c. The IS curve is downward sloping because goods market equilibrium implies that an increase in taxes leads to a lower level of output.
d. If government spending and taxes increase by the same amount, the IS curve does not shift.
e. The LM curve is horizontal at the central bank's policy choice of the interest rate.
f. The real money supply is constant along the LM curve.
g. If the nominal money supply is \$400 billion and the price level rises from an index value of 100 to an index value of 103, the real money supply rises.
h. If the nominal money supply rises from \$400 billion to \$420 billion and the price level rises from an index value of 100 to 102, the real money supply rises.
i. An increase in government spending leads to a decrease in investment in the IS-LM model.

2. Consider first the goods market model with constant investment that we saw in Chapter 3. Consumption is given by

$$C = c_0 + c_1(Y - T)$$

and I, G, and T are given.

a. Solve for equilibrium output. What is the value of the multiplier for a change in autonomous spending?
   Now let investment depend on both sales and the interest rate:
   $$I = b_0 + b_1 Y - b_2 i$$
b. Solve for equilibrium output using the methods learned in Chapter 3. At a given interest rate, why is the effect of a change in autonomous spending bigger than what it was in part a? (Assume $c_1 + b_1 < 1$.)
c. Suppose the central bank chooses an interest rate of $\bar{i}$. Solve for equilibrium output at that interest rate.
d. Draw the equilibrium of this economy using an IS-LM diagram.

3. The response of the economy to fiscal policy

a. Use an IS-LM diagram to show the effects on output of a decrease in government spending. Can you tell what happens to investment? Why?
   Now consider the following IS-LM model:
   $$C = c_0 + c_1(Y - T)$$
   $$I = b_0 + b_1 Y - b_2 i$$
   $$Z = C + I + G$$
   $$i = \bar{i}$$
b. Solve for equilibrium output when the interest rate is $\bar{i}$. Assume $c_1 + b_1 < 1$. (*Hint*: You may want to rework through Problem 2 if you are having trouble with this step.)
c. Solve for equilibrium level of investment.
d. Let's go behind the scenes in the money market. Chapter 4 introduced equations that describe equilibrium in the money market. Let's write the equation characterizing the equilibrium as: $M/P = d_1 Y - d_2 i$. Solve for the equilibrium level of the real money supply when $i = \bar{i}$. How does the real money supply vary with government spending?

4. Consider the money market to better understand the horizontal LM curve in this chapter.

The LM relation (equation 5.3) is $\frac{M}{P} = Y\,L(i)$

a. What is on the left-hand side of equation (5.3)?
b. What is on the right-hand side of equation (5.3)?
c. Go back to Figure 4-2 in the previous chapter. How is the function $L(i)$ represented in that figure?
d. Modify Figure 4-2 to represent equation (5.3) in two ways. How does the horizontal axis have to be relabeled? What is the variable that now shifts the money demand function? Draw a modified Figure 4-2 with the appropriate labels.
e. Use your modified Figure 4-2 to show that (1) as output rises, to keep the interest rate constant, the central bank must increase the real money supply; (2) as output falls, to keep the interest rate constant, the central bank must decrease the real money supply.

5. Consider the following numerical example of the IS-LM model:

$$C = 200 + 0.25Y_D$$
$$I = 150 + 0.25Y - 1000i$$
$$G = 250$$
$$T = 200$$
$$\bar{i} = .05$$

a. Derive the IS relation. (*Hint:* You want an equation with Y on the left side and everything else on the right.)
b. The central bank sets an interest rate of 5%. How is that decision represented in the equations?
c. What is the level of real money supply when the interest rate is 5%? Use the expression:

$$(M/P) = 2Y - 8000i$$

d. Solve for the equilibrium values of C and I, and verify the value you obtained for Y by adding C, I, and G.
e. Now suppose that the central bank cuts the interest rate to 3%. How does this change the LM curve? Solve for Y, I, and C, and describe in words the effects of an expansionary monetary policy. What is the new equilibrium value of M/P supply?
f. Return to the initial situation in which the interest rate set by the central bank is 5%. Now suppose that government spending increases to $G = 400$. Summarize the effects of an expansionary fiscal policy on Y, I, and C. What is the effect of the expansionary fiscal policy on the real money supply?
g. Starting from an interest rate equal to 5% and government spending equal to 250 units, increase government spending to 400 units while fixing the real money supply at 1600 units. [*Hint*: The money market must be in equilibrium so $1600 = 2Y - 8000i$ (part c) and the goods market must be in equilibrium so $Y = C + I + G$ at the same values of Y and i.] Compare the effect of the increase in government spending on Y, I and C to the same increase in G in part g and explain the difference.

DIG DEEPER

6. Investment and the interest rate

The chapter argues that investment depends negatively on the interest rate because an increase in the cost of borrowing discourages investment. However, firms often finance their investment projects using their own funds.

If a firm is considering using its own funds (rather than borrowing) to finance investment projects, will higher interest rates discourage the firm from undertaking these projects? Explain. (*Hint*: Think of yourself as the owner of a firm that has earned profits and imagine that you are going to use the profits either to finance new investment projects or to buy bonds. Will your decision to invest in new projects in your firm be affected by the interest rate?)

7. The Bush-Greenspan policy mix

In 2001, the Fed pursued an expansionary monetary policy and reduced interest rates. At the same time, President George W. Bush pushed through legislation that lowered income taxes.

a. Illustrate the effect of such a policy mix on output.
b. How does this policy mix differ from the Clinton-Greenspan mix?
c. What happened to output in 2001? How do you reconcile the fact that both fiscal and monetary policies were expansionary with the fact that growth was so low in 2002? (*Hint*: What else happened?)

8. What mix of monetary and fiscal policy is needed to meet the following objectives?

a. Increase Y while keeping $\bar{i}$ constant. Would investment (I) change?
b. Decrease a fiscal deficit while keeping Y constant. Why must $\bar{i}$ also change?

9. The (less paradoxical) paradox of saving

A chapter problem at the end of Chapter 3 considered the effect of a drop in consumer confidence on private saving and investment, when investment depended on output but not on the interest rate. Here, we consider the same experiment in the context of the IS-LM framework, in which investment depends on the interest rate and output but the central bank moves interest rates to keep output constant.

a. Suppose consumer confidence falls, so households save a higher proportion of their income. In an IS-LM diagram where the central bank moves interest rates to keep output constant, show the effect of the fall in consumer confidence on the equilibrium in the economy.
b. How will the fall in consumer confidence affect consumption, investment, and private saving? Will the attempt to save more necessarily lead to more saving? Will this attempt necessarily lead to less saving?

10. Fiscal policy and investment. Read Focus Box "Deficit Reduction: Good or Bad for Investment?"

In each case below, there is a fiscal consolidation. Remember that equilibrium condition in good markets can also be written

$$I = S + (T - G)$$

a. How does this fiscal consolidation increase public saving? Calculate the change in public saving and the change in private saving? What must happen to the target interest rate for this policy to describe the changes in parts a, b, and c?

Year	Y	C	I	G	T	S
Pre-policy	1000	500	200	300	200	
Post-policy	950	400	300	250	250	

b.

Year	Y	C	I	G	T	S
Pre-policy	1000	500	200	300	200	
Post-policy	900	450	250	200	200	

c.

Year	Y	C	I	G	T	S
Pre-policy	1000	500	200	300	200	
Post-policy	975	480	195	300	300	

EXPLORE FURTHER

11. *The Clinton-Greenspan policy mix*

As described in this chapter, during the Clinton administration the policy mix changed toward more contractionary fiscal policy and more expansionary monetary policy. This question explores the implications of this policy mix, in both theory and fact.

a. What must the Federal Reserve do to ensure that, if G falls and T rises, that combination of policies has no effect on output? Show the effects of these policies in an IS-LM diagram. What happens to the interest rate? What happens to investment?

b. Go to the website of the *Economic Report of the President* www.govinfo.gov/app/collection/erp/2019 and look at Table B-46 in the statistical appendix. What happened to federal receipts (tax revenues), federal outlays, and the budget deficit as a percentage of GDP over the period 1992 to 2000? (Note that federal outlays include transfer payments, which would be excluded from the variable G, as we define it in our IS-LM model. Ignore the difference.)

c. The Federal Reserve Board of Governors posts the recent history of the federal funds rate at www.federalreserve.gov/releases/h15/data.htm. You can look at the rate on a daily, weekly, monthly, or annual interval. Look at the years between 1992 and 2000. When did monetary policy become more expansionary?

d. Go to Table B-2 of the *Economic Report of the President* and collect data on real GDP and real gross domestic investment for the period 1992 to 2000. Calculate investment as a percentage of GDP for each year. What happened to investment over the period?

e. Finally, go to Table B-31 and retrieve data on real GDP per capita (in chained 2005 dollars) for the period. Calculate the growth rate for each year. What was the average annual growth rate over the period 1992 to 2000? In Chapter 10 you will learn that the average annual growth rate of US real GDP per capita was 2.6% between 1950 and 2004. How did growth between 1992 and 2000 compare to the post–World War II average?

12. *Consumption, investment, and the recession of 2001*

This question asks you to examine the movements of investment and consumption before, during, and after the recession of 2001. It also asks you to consider the response of investment and consumption to the events of September 11, 2001.

Go to the website of the Bureau of Economic Analysis (www.bea.gov) and find the NIPA tables, in particular the quarterly versions of Table 1.1.1, which shows the percentage change in real GDP and its components, and Table 1.1.2, which shows the contribution of the components of GDP to the overall percentage change in GDP. Table 1.1.2 weighs the percentage change of the components by their size. Investment is more variable than consumption, but consumption is much bigger than investment, so smaller percentage changes in consumption can have the same impact on GDP as much larger percentage changes in investment. Note that the quarterly percentage changes are annualized (i.e., expressed as annual rates). Retrieve the quarterly data on real GDP, consumption, gross private domestic investment, and nonresidential fixed investment for the years 1999 to 2002 from Tables 1.1.1 and 1.1.2.

a. Identify the quarters of negative growth in 2000 and 2001.

b. Track consumption and investment in 2000 and 2001. From Table 1.1.1, which variable had the bigger percentage change around this time? Compare nonresidential fixed investment with overall investment. Which variable had the bigger percentage change?

c. From Table 1.1.2, get the contribution to GDP growth of consumption and investment for 1999 to 2001. Calculate the average of the quarterly contributions for each variable for each year. Now calculate the change in the contribution of each variable for 2000 and 2001 (i.e., subtract the average contribution of consumption in 1999 from the average contribution of consumption in 2000, subtract the average contribution of consumption in 2000 from the average contribution of consumption in 2001, and do the same for investment for both years). Which variable had the largest decrease in its contribution to growth? What do you think was the proximate cause of the recession of 2001? (Was it a fall in investment demand or a fall in consumption demand?)

d. Now look at what happened to consumption and investment after the events of September 11 in the third and fourth quarters of 2001 and in the first two quarters of 2002. Does the drop in investment at the end of 2001 make sense to you? How long did this drop in investment last? What happened to consumption about this time? How do you explain, in particular, the change in consumption in the fourth quarter of 2001? Did the events of September 11, 2001, cause the recession of 2001? Use the discussion in the chapter and your own intuition as guides in answering these questions.

FURTHER READING

- A description of the US economy, from the period of "irrational exuberance" to the 2001 recession, and the role of fiscal and monetary policy is given by Paul Krugman in *The Great Unraveling* (New York: W. W. Norton, 2003). (Warning: Krugman did not like the Bush administration or its policies!)

This page is intentionally left blank

6

Financial Markets II: The Extended IS-LM Model

Until now, we assumed that there were only two financial assets—money and bonds—and just one interest rate—the rate on bonds—determined by monetary policy. As you know, the financial system is vastly more complex than that. There are many interest rates and many financial institutions. And the financial system plays a major role in the economy: In the United States, financial activity accounts for 7% of GDP, a large number.

Before the 2008 crisis, the importance of the financial system was downplayed in macroeconomics. All interest rates were often assumed to move with the rate determined by monetary policy, so one could just focus on the rate determined by monetary policy and assume that other rates would move with it. The crisis made painfully clear that this assumption was too simplistic and that the financial system can be subject to crises with major macroeconomic implications. This chapter looks more closely at the role of the financial system and its macroeconomic implications, and then gives an account of what happened during the Great Financial Crisis.

This chapter cannot replace a text in finance. But it will tell you enough to understand why the financial system is central to macroeconomics.

Section 6-1 introduces the distinction between nominal and real interest rates.

Section 6-2 introduces the notion of risk and how it affects the interest rates charged to different borrowers.

Section 6-3 looks at the role of financial intermediaries.

Section 6-4 extends the IS-LM model to integrate what we have just learned.

Section 6-5 then uses this extended model to describe the recent financial crisis and its macroeconomic implications.

If you remember one basic message from this chapter, it should be: The financial system matters, and financial crises can have large macroeconomic effects.

6-1 NOMINAL VERSUS REAL INTEREST RATES

At the time of this writing, the one-year T-bill rate is even lower than it was in 2006, but comparing 1981 with 2006 is the best way to convey my point.

In January 1980, the one-year US T-bill rate—the interest rate on one-year government bonds—was 10.9%. In January 2006, the rate was only 4.2%. It was clearly much cheaper to borrow in 2006 than it was in 1981.

Or was it? In January 1980, expected inflation was around 9.5%. In January 2006, expected inflation was around 2.5%. This would seem relevant. The interest rate tells us how many dollars we shall have to pay in the future in exchange for having one more dollar today. But when we borrow, what we really want to know is how many goods we will have to give up in the future in exchange for the goods we get today. Likewise, when we lend, we want to know how many goods—not how many dollars—we will get in the future for the goods we give up today. The presence of inflation makes this distinction important. What is the point of receiving high interest payments in the future if inflation between now and then is so high that, with the return we shall receive then, we shall be unable to buy more goods?

This is where the distinction between nominal interest rates and real interest rates comes in.

The **nominal interest rate** is the interest rate in terms of dollars.

The **real interest rate** is the interest rate in terms of a basket of goods.

- Interest rates expressed in terms of dollars (or, more generally, in units of the national currency) are called **nominal interest rates**. The interest rates printed in the financial pages of newspapers are typically nominal interest rates. For example, when we say that the one-year T-bill rate is 4.2%, we mean that for every dollar the government borrows by issuing one-year T-bills, it promises to pay 1.042 dollars a year from now. More generally, if the nominal interest rate for year t is i_t, borrowing 1 dollar this year requires you to pay $1 + i_t$ dollars next year. (I shall use interchangeably "this year" for "today," and "next year" for "one year from today.")
- Interest rates expressed in terms of a basket of goods are called **real interest rates**. If we denote the real interest rate for year t by r_t, then, by definition, borrowing the equivalent of one basket of goods this year requires you to pay the equivalent of $1 + r_t$ baskets of goods next year.

What is the relation between nominal and real interest rates? How do we go from nominal interest rates—which we do observe—to real interest rates—which we typically do not observe? The intuitive answer: We must adjust the nominal interest rate to take into account expected inflation.

Let's go through the step-by-step derivation.

Assume there is only one good in the economy, bread (we shall add jam and other goods later). Denote the one-year nominal interest rate, in terms of dollars, by i_t. If you borrow 1 dollar this year, you will have to repay $(1 + i_t)$ dollars next year. But what you really want to know is: If you borrow enough to eat one more pound of bread this year, how much will you have to repay, in terms of pounds of bread, next year?

Figure 6-1 helps us derive the answer. The top part repeats the definition of the one-year real interest rate. The bottom part shows how we can derive the one-year real interest rate from information about the one-year nominal interest rate and the price of bread.

- Start with the downward pointing arrow in the lower left of Figure 6-1. Suppose you want to eat one more pound of bread this year. If the price of a pound of bread this year is P_t dollars, to eat one more pound of bread you must borrow P_t dollars.

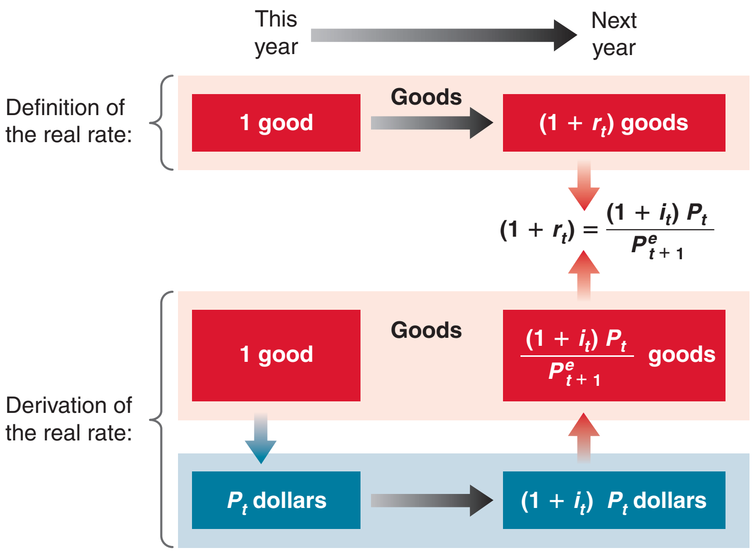


Figure 6-1

Definition and Derivation of the Real Interest Rate

- If i_t is the one-year nominal interest rate—the interest rate in terms of dollars—and you borrow P_t dollars, you will have to repay $(1 + i_t)P_t$ dollars next year. This is represented by the arrow from left to right at the bottom of Figure 6-1.
- What you care about, however, is not dollars but pounds of bread. Thus, the last step involves converting dollars back to pounds of bread next year. Let P^e_{t+1} be the price of bread you expect to pay next year. (The superscript e indicates that this is an expectation; you do not know yet what the price of bread will be next year.) How much you expect to repay next year, in terms of pounds of bread, is therefore equal to $(1 + i_t)P_t$ (the number of dollars you have to repay next year) divided by P^e_{t+1} (the price of bread in terms of dollars expected for next year), so $(1 + i_t)P_t/P^e_{t+1}$. This is represented by the arrow pointing up in the lower right of Figure 6-1.

If you have to repay \$10 next year and you expect the price of bread to be \$2 a loaf, you expect to have to repay the equivalent of $10/2 = 5$ loaves of bread next year. This is why we divide the dollar amount $(1 + i_t)P_t$ by the expected price of bread next year, P^e_{t+1}.

Putting together what you see in the top and bottom parts of Figure 6-1, it follows that the one-year real interest rate, r_t, is given by:

$$1 + r_t = (1 + i_t)\frac{P_t}{P^e_{t+1}} \tag{6.1}$$

This relation looks intimidating. Two simple manipulations make it look friendlier:

- Denote expected inflation between t and $t+1$ by π^e_{t+1}. Given that there is only one good—bread—the expected rate of inflation equals the expected change in the dollar price of bread between this year and next year, divided by the dollar price of bread this year:

$$\pi^e_{t+1} \equiv \frac{(P^e_{t+1} - P_t)}{P_t} \tag{6.2}$$

Using equation (6.2), rewrite P_t/P^e_{t+1} in equation (6.1) as $1/(1 + \pi^e_{t+1})$. Replace in equation (6.1) to get

$$(1 + r_t) = \frac{1 + i_t}{1 + \pi^e_{t+1}} \tag{6.3}$$

One plus the real interest rate equals the ratio of one plus the nominal interest rate, divided by one plus the expected rate of inflation.

Add 1 to both sides in equation (6.2):

$$1 + \pi^e_{t+1} = 1 + \frac{(P^e_{t+1} - P_t)}{P_t}$$

Reorganize:

$$1 + \pi^e_{t+1} = \frac{P^e_{t+1}}{P_t}$$

Take the inverse on both sides:

$$\frac{1}{1 + \pi^e_{t+1}} = \frac{P_t}{P^e_{t+1}}$$

Replace in equation (6.1) to get equation (6.3)

See Proposition 6 in Appendix 2 at the end of the book. Suppose $i = 10\%$ and $\pi^e = 5\%$. The exact relation in equation (6.3) gives $r_t = 4.8\%$. The approximation given by equation (6.4) gives 5%—close enough. The approximation can be quite bad, however, when i and π^e are high. If, for example, $i = 100\%$ and $\pi^e = 80\%$, the exact relation gives $r = 11\%$, but the approximation gives $r = 20\%$—a big difference.

- Equation (6.3) gives us the exact relation of the real interest rate to the nominal interest rate and expected inflation. However, when both the nominal interest rate and expected inflation are not too large—say, less than 10% per year—a close approximation (denoted by ≈) to this equation is given by

$$r_t \approx i_t - \pi^e_{t+1} \tag{6.4}$$

Make sure you remember equation (6.4). It says that the real interest rate is (approximately) equal to the nominal interest rate minus expected inflation. (In the rest of the book, I shall often treat the relation in equation (6.4) as if it were an equality. Remember, however, that it is only an approximation.)

Note some of the implications of equation (6.4):

- When expected inflation equals zero, the nominal and the real interest rates are equal.
- Because expected inflation is typically positive, the real interest rate is typically lower than the nominal interest rate.
- For a given nominal interest rate, the higher the expected rate of inflation, the lower the real interest rate.

The case where expected inflation happens to be equal to the nominal interest rate is worth looking at more closely. Suppose the nominal interest rate and expected inflation both equal 10%, and you are the borrower. For every dollar you borrow this year, you will have to repay 1.10 dollars next year. This looks expensive. But dollars will be worth 10% less in terms of bread next year. So, if you borrow the equivalent of one pound of bread, you will have to repay the equivalent of one pound of bread next year. The real cost of borrowing—the real interest rate—is equal to zero. Now suppose you are the lender: For every dollar you lend this year, you will receive 1.10 dollars next year. This looks attractive, but dollars next year will be worth 10% less in terms of bread. If you lend the equivalent of one pound of bread this year, you will get the equivalent of one pound of bread next year: Despite the 10% nominal interest rate, the real interest rate is equal to zero.

We have assumed so far that there is only one good—bread—but what we have done generalizes easily to many goods. All we need to do is substitute the price level—the price of a basket of goods—for the price of bread in equation (6.1) or (6.3). If we use the consumer price index (CPI) to measure the price level, the real interest rate tells us how much consumption we must give up next year to consume more today.

Nominal and Real Interest Rates in the United States since 1978

Let us return to the question at the start of this section. We can now restate it as follows: Was the real interest rate lower in 2006 than it was in 1981? More generally, what has happened to the real interest rate in the United States over the past four decades?

The answer is shown in Figure 6-2, which plots both nominal and real interest rates since 1978. For each year, the nominal interest rate is the one-year T-bill rate at the beginning of the year. To construct the real interest rate, we need a measure of expected inflation—more precisely, the rate of inflation expected as of the beginning of each year. We use, for each year, the forecast of inflation, using the GDP deflator, for that year published at the end of the previous year by the OECD. For example, the forecast of inflation used to construct the real interest rate for 2006 is 2.5%, the 2006 inflation forecast published by the OECD in November 2005.

Note that the real interest rate $(i - \pi^e)$ is based on expected inflation. If actual inflation turns out to be different from expected inflation, the realized real interest rate $(i - \pi)$ will be different from the real interest rate. For this reason, the real interest rate is

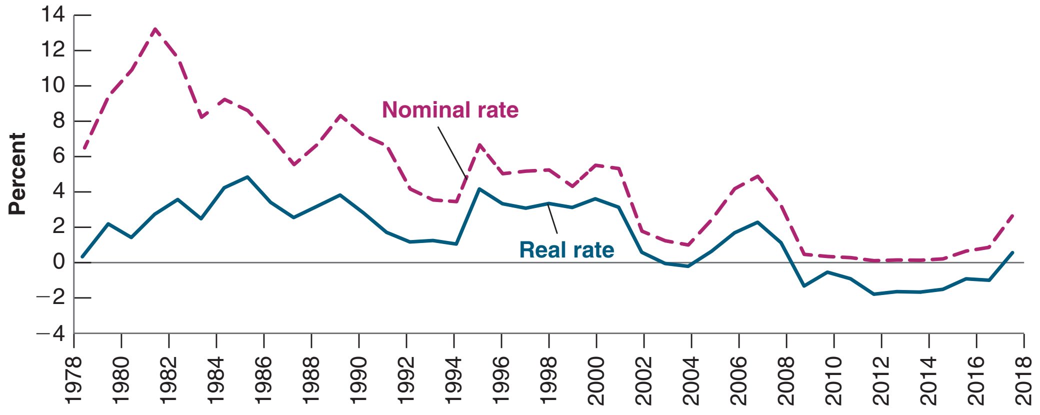


Figure 6-2

Nominal and Real One-Year T-Bill Rates in the United States, since 1978

The nominal rate has declined considerably since the early 1980s, but because expected inflation has declined as well, the real rate has declined much less than the nominal rate.

Source: FRED: Nominal interest rate is the one-year Treasury bill in December of the previous year: Series TB1YR (Series TB6MS in December 2001, 2002, 2003, and 2004). Expected inflation is the 12-month forecast of inflation, using the GDP deflator, from the November OECD Economic Outlook from the previous year.

sometimes called the *ex-ante real interest rate* (*ex-ante* means "before the fact"; here, before inflation is known). The realized real interest rate is called the *ex-post real interest rate* (*ex-post* means "after the fact"; here, after inflation is known).

Figure 6-2 shows the importance of adjusting for inflation. Although the nominal interest rate was much lower in 2006 than it was in 1981, the real interest rate was actually higher in 2006 (about 1.7%) than it was in 1981 (about 1.4%). Put another way, despite the large decline in nominal interest rates, borrowing was actually more expensive in 2006 than it was 1981. This is because inflation (and with it, expected inflation) has steadily declined since the early 1980s.

Nominal and Real Interest Rates: The Zero Lower Bound and Deflation

Which interest rate should enter the IS relation? Clearly, in thinking about consumption or investment decisions, what matters to people or to firms is the real interest rate, the rate in terms of goods. This has a straightforward implication for monetary policy. Although the central bank chooses the nominal rate (as we saw in Chapter 4), it cares about the real interest rate because this is the rate that affects spending decisions. To set the real interest rate it wants, it thus has to take into account expected inflation.

If the central bank wants to set the real interest rate equal to r, it must choose the nominal rate, i, so that, given expected inflation, π^e, the real interest rate, $r = i - \pi^e$, is at the desired level. For example, if it wants the real interest rate to be 4% and expected inflation is 2%, it will set the nominal interest rate, i, at 6%.

But, as we discussed in Chapter 4 in the context of the liquidity trap, the zero lower bound means that the nominal interest rate cannot be negative; otherwise, people would not want to hold bonds. This implies that the real interest rate cannot be lower than the negative of the rate of inflation. If expected inflation is 2%, for example, then the lowest the real interest rate can be is $0\% - 2\% = -2\%$. So long as expected inflation is positive, this allows for negative real interest rates. But if expected inflation turns negative, if people anticipate deflation, then the lower bound on the real rate becomes positive and can turn out to be high. If, for example, expected deflation is 2%, the real rate cannot be less than 2%. This may not be low enough to increase the demand for goods by much, and the economy may remain in recession. As we shall see in Section 6-5, the zero lower bound turned out to be a serious concern during the 2008 crisis.

6-2 RISK AND RISK PREMIA

Until now, we assumed there was only one type of bond. But bonds differ in several ways. They differ in terms of maturity—i.e., the length of time over which they promise payments. For example, 1-year government bonds promise one payment a year hence; 10-year government bonds promise a stream of payments over 10 years. They also differ in terms of risk. Some bonds are nearly riskless; the probability that the borrower will not repay is negligible. Some bonds instead are risky, with a nonnegligible probability that the borrower will not be able or willing to repay. In this chapter, we shall focus on risk, leaving aside the issue of maturity.

We shall return to a discussion of maturity, and the relation between interest rates on bonds of different maturities, once I have introduced a more formal treatment of expectations, in Chapter 14.

Neither you nor I can borrow at the federal funds rate set by the Fed. Nor can we borrow at the same rate as the US government. There is a good reason for this. Whoever might be lending to us knows that there is a chance that we may not be able to repay. The same is true for firms that issue bonds. Some firms present little risk and others more. To compensate for the risk, bond holders require a **risk premium**.

What determines this risk premium?

- The first factor is the probability of default itself. The higher this probability, the higher the interest rate investors will ask for. More formally, let i be the nominal interest rate on a riskless bond, and $i + x$ be the nominal interest rate on a risky bond, which has probability, p, of defaulting. Call x the risk premium. Suppose investors are risk neutral, so they want to have the same expected return on a riskless and a risky bond. For the expected return on both, the following relation must hold:

$$(1 + i) = (1 - p)(1 + i + x) + p(0)$$

  The left-hand side gives the return on the riskless bond. The right-hand side gives the expected return on the risky bond. With probability $(1 - p)$, there is no default and the bond will pay $(1 + i + x)$. With probability p, there is default and the bond will pay nothing. Reorganizing gives:

$$x = (1 + i)\, p/(1 - p)$$

  So, for example, if the interest rate on a riskless bond is 4% and the probability of default on the risky bond is 2%, then the risk premium required to give the same expected rate of return as on the riskless bond is 2.1%.

For small values of i and p, a good approximation to this formula is simply $x = p$.

- The second factor is the degree of **risk aversion** of the bond holders. Even if the expected return on the risky bond is the same as on a riskless bond, the risk itself will make them reluctant to hold the risky bond. Thus, they will ask for an even higher premium to compensate for the risk, so x will be higher than the value we just derived. How much higher will depend on bond holders' degree of risk aversion. And if they become more risk averse, the risk premium will go up even if the probability of default has not changed.

To show why this matters, Figure 6-3 plots the interest rates on three types of bonds since 2000: US government bonds, which are considered nearly riskless; and corporate bonds rated as safe (AAA) and less safe (BBB) by ratings agencies.

Different rating agencies use different rating systems. The rating scale used here is that of Standard and Poor's and ranges from AAA (nearly riskless) and BBB to C (bonds with a high probability of default).

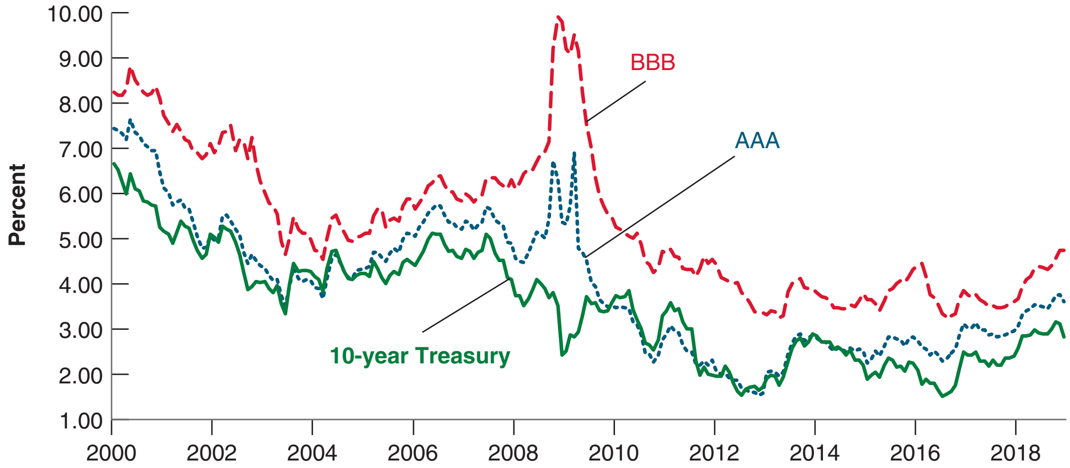


Figure 6-3

Yields on 10-Year US Government Treasury, AAA, and BBB Corporate Bonds, since 2000

In September 2008, the financial crisis led to a sharp increase in the rates at which firms could borrow.

Source: FRED: Series DGS10; For AAA and BBB corporate bonds, Bank of America Merrill Lynch Series BAMLC0A4CBBB, BAMLC0A1CAAAEY

Note three things about Figure 6-3. First, the rate on even the most highly rated (AAA) corporate bonds is higher than the rate on US government bonds, by a premium of about 1% on average. The US government can borrow at cheaper rates than US corporations. Second, the rate on lower-rated (BBB) corporate bonds is higher than the rate on the most highly rated bonds by a premium often exceeding 2%. Third, note what happened during the financial crisis in 2008 and 2009. Although the rate on government bonds decreased, reflecting the Fed's decision to decrease the policy rate, the interest rate on lower-rated bonds increased sharply, reaching 10% at the height of the crisis. Put another way, despite the fact that the Fed was lowering the policy rate to zero, the rate at which lower-rated firms could borrow became much higher, making it extremely unattractive for these firms to invest. In terms of the IS-LM model, this shows why we cannot assume that the policy rate enters the IS relation. The rate at which many borrowers can borrow may be much higher than the policy rate.

To summarize: In the last two sections, I have introduced the concepts of real versus nominal rates and the concept of a risk premium. In Section 6-4, we shall extend the IS-LM model to take both concepts into account. Before we do, let's turn to the role of financial intermediaries.

6-3 THE ROLE OF FINANCIAL INTERMEDIARIES

Until now, we have looked at **direct finance**, that is, borrowing directly by the ultimate borrowers from the ultimate lenders. In fact, much of the borrowing and lending takes place through financial intermediaries, which are financial institutions that receive

Figure 6-4

Bank Assets, Capital, and Liabilities

Bank Balance Sheet

Assets		Liabilities and capital	
Assets	100	Liabilities	80
		Capital	20

Because it grew in the "shadow" of banks, the nonbank part of the financial system is called **shadow banking**. But it is now large and no longer in the shadows.

funds from some investors and then lend these funds to others. Among these institutions are banks, but also, increasingly, "nonbanks," such as mortgage companies, money market funds, and hedge funds.

Financial intermediaries perform an important function. They develop expertise about specific borrowers and can tailor lending to their specific needs. In normal times, they function smoothly. They borrow and lend, charging a slightly higher interest rate than the rate at which they borrow so as to make a profit. Once in a while, however, they run into trouble, and this is what happened in the Great Financial Crisis. To understand why, let's first focus on banks and start, in Figure 6-4, with a much simplified bank balance sheet (the discussion applies to nonbanks as well and we shall return to them later).

One wishes that the balance sheets of banks were that simple and transparent. Had that been the case, the crisis would have been much more limited.

Consider a bank that has assets of 100, liabilities of 80, and capital of 20. You can think of the owners of the bank as having directly invested 20 of their own funds, borrowed another 80 from other investors, and bought various assets for 100. The liabilities may be checkable deposits, interest-paying deposits, or borrowing from investors and other banks. The assets may be reserves (central bank money), loans to consumers, loans to firms, loans to other banks, mortgages, government bonds, or other forms of securities.

In drawing a bank balance sheet in Chapter 4, we ignored capital (and focused instead on the distinction between reserves and other assets). Ignoring capital was unimportant there. But it is important here. Let's see why.

The Choice of Leverage

Start with two definitions. The **capital ratio** of a bank is defined as the ratio of its capital to its assets, so, for the bank in Figure 6-4, $20/100 = 20\%$. The **leverage ratio** of a bank is defined as the ratio of assets to capital, so as the inverse of the capital ratio, in this case $100/20 = 5$. It is traditional to think in terms of leverage and to focus on the leverage ratio. I shall follow tradition. But given the simple relation between the two, the discussion could equivalently be in terms of the capital ratio.

In thinking about what leverage ratio it should choose, the bank has to balance two factors. A higher leverage ratio implies a higher expected profit rate. But it also implies a higher risk of bankruptcy. Let's look at each factor in turn.

- Suppose the expected rate of return on assets is 5% and the expected rate of return on liabilities is 4%. Then, the expected profit of the bank is equal to $(100 \times 5\% - 80 \times 4\%) = 1.8$. Given that the owners of the bank have put in 20 of their own funds, the expected profit per unit of capital is equal to $1.8/20 = 9\%$. Now suppose the owners of the bank decided instead to put in only 10 of their own funds and borrowed 90. The capital ratio of the bank would then be equal to $10/100 = 10\%$, its leverage would be 10, and its expected profit would be equal to $(100 \times 5\% - 90 \times 4\%) = 1.4$. Its expected profit per unit of capital would be $1.4/10 = 14\%$, so substantially higher. By increasing its leverage, and decreasing its own funds, the bank would increase its expected profit per unit of capital.

What would be the expected profit per unit of capital if the bank chose to have zero leverage? If the bank chose to have full leverage (no capital)? (The second question is a trick question.)

- Why shouldn't the bank choose a high leverage ratio? Because higher leverage also implies a higher risk that the value of the assets becomes less than the value of liabilities, in turn implying a higher risk of **insolvency**. For the bank in Figure 6-4, its assets can decrease in value to 80 without the bank becoming insolvent and going bankrupt. But if it were to choose a leverage ratio of 10, any decrease in the value of the assets below 90 would lead the bank to become insolvent. The risk of bankruptcy would be much higher.

A bank is solvent if the value of its assets exceeds the value of its liabilities. It is insolvent otherwise.

Thus, the bank must choose a leverage ratio that takes into account both factors. Too low a leverage ratio means less profit. Too high a leverage ratio means too high a risk of bankruptcy.

Leverage and Lending

Suppose a bank has chosen its preferred leverage ratio, and suppose that the value of its assets declines. For example, the assets of the bank in Figure 6-4 decrease in value from 100 to 90, say as a result of bad loans. The capital of the bank is now down to $90 - 80 = 10$. Its leverage ratio increases from 5 to 9. The bank is still solvent, but it is clearly more at risk than it was before. What will it want to do? It may want to increase capital, for example, by asking other investors to provide funds. But it is also likely to want to decrease the size of its balance sheet. For example, if it can call back some loans for an amount of 40 and thus reduce its assets to $90 - 40 = 50$, and then use the 40 to decrease its liabilities to $80 - 40 = 40$, its capital ratio will be $10/50 = 20\%$, back to its original value. But although the capital ratio of the bank is back to its desired level, the effect is a sharp decrease in the bank's lending.

Let's go one step further. Suppose that, starting from the balance sheet in Figure 6-4, the decline in the value of the assets is large, say down from 100 to 70. The bank will become insolvent and go bankrupt. The borrowers that depended on the bank may have a hard time finding another lender.

Why is this relevant to us? Because whether banks remain solvent but cut lending or become insolvent, the decrease in lending that this triggers may well have major adverse macroeconomic effects. Again, let's defer a discussion of macroeconomic implications to the next section. And before we get there, let's explore things further.

Liquidity

We looked at the case where bank assets declined in value and saw that this led banks to reduce lending. Now consider a case in which investors are unsure of the value of the assets of the bank, and believe, right or wrong, that the value of the assets may have come down. Then, leverage can have disastrous effects. Let's see why.

- If investors have doubts about the value of the bank assets, the safe thing for them to do is to take their funds out of the bank. This creates serious problems for the bank, which needs to come up with the funds to repay the investors. The loans it has made cannot easily be called back. Typically, the borrowers no longer have the funds at the ready; they have used them to pay bills, buy a car, purchase a machine, and such. Selling the loans to another bank is likely to be difficult as well. Assessing the value of the loans is difficult for the other banks, which do not have the specific knowledge about the borrowers the original bank has.
- In general, the harder it is for others to assess the value of a bank's assets, the more likely the bank is to be unable to sell them or to have to do it at **fire sale prices**, prices far below the true value of the loans. Such sales make matters worse for the bank. As the value of the assets decreases, the bank may well become insolvent and go bankrupt. In turn, as investors realize this may happen, they are more likely to want to get their funds out, forcing more fire sales, and making the problem worse. Note that this can happen even if the initial doubts of investors were totally unfounded, even if the value of the bank assets had not decreased in the first place. The decision by

FOCUS

Bank Runs

Take a healthy bank, that is, a bank with a portfolio of good loans. Suppose rumors start that the bank is not doing well and some loans will not be repaid. Believing that the bank may fail, people with deposits at the bank will want to close their accounts and withdraw cash. If enough people do so, the bank will run out of funds. Given that the loans cannot easily be called back, the bank will not be able to satisfy the demand for cash, and it will have to close.

Conclusion: Fear that a bank will close can actually cause it to close—even if all its loans were good in the first place. The financial history of the United States up to the 1930s is full of such bank runs. One bank fails for the right reason (because it has made bad loans). This causes depositors at other banks to panic and withdraw money from their banks, forcing them to close. You have probably seen *It's a Wonderful Life*, a classic movie with James Stewart that runs on TV every year around Christmas. After another bank in Stewart's town fails, depositors at the savings and loan he manages get scared and want to withdraw their money, too. Stewart successfully persuades them this is not a good idea. *It's a Wonderful Life* has a happy ending. But in real life, most bank runs didn't end well. (For another famous movie bank run, and how it can start, watch *Mary Poppins*.)

What can be done to avoid bank runs?

One potential solution is called **narrow banking**. Narrow banking would restrict banks to holding liquid and safe government bonds, like T-bills. Loans would have to be made by financial intermediaries other than banks. This would likely eliminate bank runs. Some recent changes in US regulation have gone in that direction, restricting banks that rely on deposits from engaging in some financial operations, but they stop far short of imposing narrow banking. One worry with narrow banking is that, although it might indeed eliminate runs on banks, the problem might migrate to shadow banking and create runs there.

In practice, the problem has been tackled in two ways: First, by trying to limit bank runs in the first place; second, if bank runs happen nevertheless, by having the central bank provide funds to banks so that they do not have to engage in fire sales.

To limit bank runs, governments in most advanced countries have put in place a system of deposit insurance. The United States, for example, introduced **federal deposit insurance** in 1934. The US government now insures each checkable deposit account up to a ceiling, which, since 2008, is $250,000. As a result, there is no reason for depositors to run and withdraw their money.

Deposit insurance leads, however, to problems of its own. Depositors, who do not have to worry about their deposits, no longer look at the activities of the banks in which they have their accounts. Banks may then misbehave, by making loans they wouldn't have made in the absence of deposit insurance. They may take too much risk, take too much leverage.

And as the 2008–2009 crisis unfortunately showed, deposit insurance is no longer enough. First, banks rely on sources of funds other than deposits, often borrowing overnight from other financial institutions and investors. These other funds are not insured, and during the crisis there was in effect a run on many banks—not from the traditional depositors but from wholesale funders. Second, financial institutions other than banks can be subject to the same problem, with investors wanting their funds back quickly and with assets difficult to dispose of or sell quickly.

So, to the extent that runs cannot be fully prevented, central banks have put in place programs to provide funds to banks in case they face a run. In such circumstances, the central bank will agree to lend to a bank against the value of the bank's assets so that the bank does not have to sell its assets and fire sales can be avoided. Access to such a provision was traditionally reserved for banks. But the recent crisis showed that other financial institutions may be subject to runs and may also need access.

Just like deposit insurance, the central bank's **liquidity provision** (as it is called) is not a perfect solution. In practice, central banks may face a difficult choice. Assessing which financial institutions beyond banks can have access to the liquidity provision is delicate. Assessing the value of the assets, and thus deciding how much can be lent to a financial institution, can also be difficult. The central bank would not want to provide funds to an institution that is actually insolvent; but, in the middle of a financial crisis, the difference between insolvency and illiquidity may be difficult to establish.

To watch the bank run in *It's a Wonderful Life*, go to *www.youtube.com/watch?v=lbwjS9iJ2Sw*
To watch the bank run in *Mary Poppins*, go to *www.youtube.com/watch?v=C6DGs3qjRwQ*

investors to ask for their funds, and the fire sales this triggers, can make the bank insolvent even if it was fully solvent to start.

- Note also that the problem is worse if investors can ask for their funds on short notice. This is clearly the case for checkable deposits at banks. Checkable deposits are also called **demand deposits**, precisely because people can ask for their funds on demand. The fact that banks' assets are largely composed of loans and their liabilities are largely composed of demand deposits makes them particularly exposed to the risk of runs, and the history of the financial system is full of examples of **bank runs**, when worries about banks' assets led to runs on banks, forcing them to close. Bank

runs were a major feature of the Great Depression, and as discussed in the Focus Box "Bank Runs," central banks have taken measures to limit them. As we shall see later in this chapter, however, this has not fully taken care of the problem, and a modern form of runs—not on banks but on other financial intermediaries—played a major role in the Great Financial Crisis.

We can summarize what we have just learned in terms of the **liquidity** of assets and liabilities. The lower the liquidity of the assets (i.e., the more difficult they are to sell), the higher the risk of fire sales and the risk that the bank becomes insolvent and goes bankrupt. The higher the liquidity of the liabilities (i.e., the easier it is for investors to get their funds at short notice), the higher the risk of fire sales as well, and the risk that the bank becomes insolvent and goes bankrupt. Again, the reason this is relevant for us is that such bankruptcies, if they occur, may well have major macroeconomic consequences. This is the topic of the next section.

6-4 EXTENDING THE IS-LM MODEL

The IS-LM model introduced in Chapter 5 had only one interest rate. It was determined by the central bank, and it entered spending decisions. It appeared both in the LM relation and the IS relation. The first three sections of this chapter showed that, although this was a useful first step, reality is substantially more complex, and we must extend our initial model.

First, we must distinguish between the nominal interest rate and the real interest rate. Second, we must distinguish the policy rate set by the central bank and the interest rates faced by borrowers. As we saw, these interest rates depend both on the risk associated with borrowers and on the state of health of financial intermediaries. The higher the risks, or the higher the leverage ratio of intermediaries, the higher the interest rate borrowers have to pay. We capture those two aspects by rewriting the IS-LM in the following way:

$$\textit{IS relation:} \quad Y = C(Y - T) + I(Y, i - \pi^e + x) + G$$
$$\textit{LM relation:} \quad i = \bar{i}$$

The way the central bank controls the nominal interest rate is by adjusting the money supply. (If you need a refresher, go back to Chapter 4.)

The LM relation remains the same. The central bank still controls the nominal interest rate. But there are two changes to the IS relation, the presence of expected inflation, π^e, and the risk premium, x.

- The expected inflation term reflects the fact that spending decisions depend, all other things equal, on the real interest rate, $r = i - \pi^e$, rather than on the nominal rate.
- The risk premium, x, captures, in a simple way, the factors we discussed previously: probability of default and level of risk aversion. It may be high because lenders perceive a higher risk that borrowers will not repay or because they are more risk averse. Or it may be high because financial intermediaries are reducing lending, because of worries about solvency or liquidity.

Two important distinctions: Real versus nominal interest rate, and policy rate versus borrowing rate.

The two equations make clear that the interest rate in the LM equation, i, is no longer the same as the interest rate in the IS relation, $r + x$. Let's call the rate in the LM equation the (nominal) **policy rate** (because it is determined by monetary policy), and that of the IS equation the (real) **borrowing rate** (because it is the rate at which consumers and firms can borrow).

One simplification: As we discussed in Section 6-2, although the central bank formally chooses the nominal interest rate, it can choose it in such a way as to achieve the real interest rate it wants (this ignores the issue of the zero lower bound to which we shall return). Thus, we can think of the central banks as choosing the real policy rate directly and rewrite the two equations as:

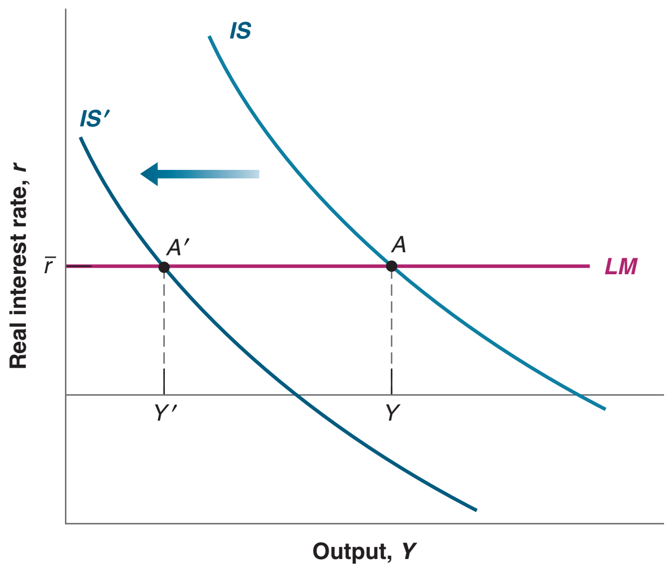
$$\textit{IS relation:} \quad Y = C(Y - T) + I(Y, r + x) + G \tag{6.5}$$
$$\textit{LM relation:} \quad r = \bar{r} \tag{6.6}$$

The central bank chooses the real policy rate, r. But the real interest rate relevant for spending decisions is the borrowing rate, $r + x$, which depends not only on the policy rate but also on the risk premium.

Figure 6-5

Financial Shocks and Output

An increase in the risk premium x leads to a shift of the IS curve to the left and a decrease in equilibrium output.



The two equations are represented in Figure 6-5. The policy rate—the real interest rate implicitly chosen by the central bank—is measured on the vertical axis and output on the horizontal axis. The IS curve is drawn for given values of G, T, and x. All other things equal, an increase in the real policy rate decreases spending and in turn output: The IS curve is downward sloping. The LM is a horizontal line at the policy rate. Equilibrium is given by point A, with associated level of output Y.

Financial Shocks and Policies

Suppose that, for some reason, x increases. There are many potential scenarios here. This may be, for example, because investors have become more risk averse and require a higher risk premium, or because one financial institution has gone bankrupt and investors have become worried about the health of other banks, starting a run, forcing these other banks to reduce lending. In Figure 6-5, the IS curve shifts to the left. At the same policy rate r, the borrowing rate, $r + x$, increases, leading to a decrease in demand and output. The new equilibrium is at point A′. Problems in the financial system lead to a recession and a financial crisis becomes a macroeconomic crisis.

For simplicity, we have looked at an exogenous increase in x. But x may well depend on output. A decrease in output, say a recession, increases the probability that some borrowers will be unable to repay: workers who become unemployed may not be able to repay loans, firms that lose sales may go bankrupt. The increase in risk leads to a further increase in the risk premium, and thus to a further increase in the borrowing rate, which can further decrease output.

What can policy do? As explained in Chapter 5, fiscal policy, be it an increase in G or a decrease in T, can shift the IS curve to the right and increase output. But a large increase in spending or a cut in taxes may imply a large increase in the budget deficit, and the government may be reluctant to cause this.

Given that the cause of the low output is that the interest rate for borrowers is too high, monetary policy appears to be a better tool. Indeed, a sufficient decrease in the policy rate, as drawn in Figure 6-6, can in principle be enough to take the economy to point A″ and keep output to its initial level. In effect, in the face of the increase in x, the central bank must decrease r so as to keep $r + x$, the rate relevant to spending decisions, unchanged.

Note that the policy rate necessary to increase demand sufficiently and return output to its previous level may well be negative, as shown in Figure 6-6. Suppose that, for example, in the initial equilibrium, r was equal to 2% and x was equal to 1%. Suppose that x increases by 4%, from 1% to 5%. To maintain the same value of $r + x$, the central bank must decrease the policy rate from 2% to $2\% - 4\% = -2\%$. This raises the issue of the constraint arising from the zero lower bound on the nominal interest rate.

Given the zero lower bound on the nominal rate, the lowest real rate the central bank can achieve is given by $r = i - \pi^e = 0 - \pi^e = -\pi^e$. In words, the lowest real policy rate the central bank can achieve is the negative of inflation. If inflation is high enough,

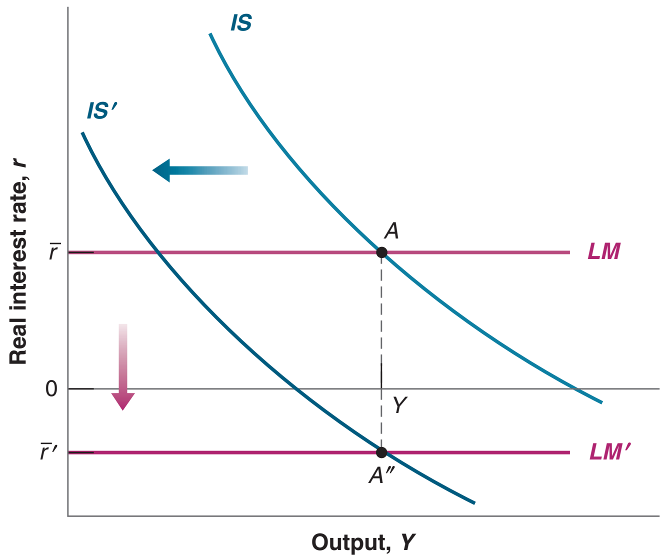


Figure 6-6

Financial Shocks, Monetary Policy, and Output

If sufficiently large, a decrease in the policy rate can in principle offset the increase in the risk premium. The zero lower bound may, however, put a limit on the decrease in the real policy rate.

say 5%, then a zero nominal rate implies a real rate of −5%, which is likely to be low enough to offset the increase in x. But if inflation is low or even negative, then the lowest real rate the central bank can achieve may not be enough to offset the increase in x. It may not be enough to return the economy to its equilibrium. As we shall see, two characteristics of the recent crisis were indeed a large increase in x and low actual and expected inflation, limiting how much central banks could use monetary policy to offset the increase in x.

We now have the elements we need to understand what triggered the financial crisis in 2008 and how it morphed into a major macroeconomic crisis. This is the topic of the next and last section of this chapter.

6-5 FROM A HOUSING PROBLEM TO A FINANCIAL CRISIS

When housing prices started declining in the United States in 2006, most economists forecast that this would lead to a decrease in demand and a slowdown in growth. Few economists anticipated that it would lead to a major macroeconomic crisis. What most had not anticipated was the effect of the decline in housing prices on the financial system, and in turn, on the economy.

Housing Prices and Subprime Mortgages

Figure 6-7 shows the evolution of an index of US housing prices since 2000. The index is known as the Case-Shiller index, named for the two economists who constructed it. The index is normalized to equal 100 in January 2000. You can see the large increase in prices in the early 2000s, followed by a large decrease. From a value of 100 in 2000, the index increased to 226 in mid-2006. It then started to decline. By the end of 2008, at the start of the financial crisis, the index was down to 162. It reached a low of 150 in early 2012 and started recovering thereafter. At the time of this writing, it is 227, close to its mid-2006 peak.

Look up *Case-Shiller* on the internet to find the index and see its recent evolution. You can also see what has happened to prices in the city where you live. ▶

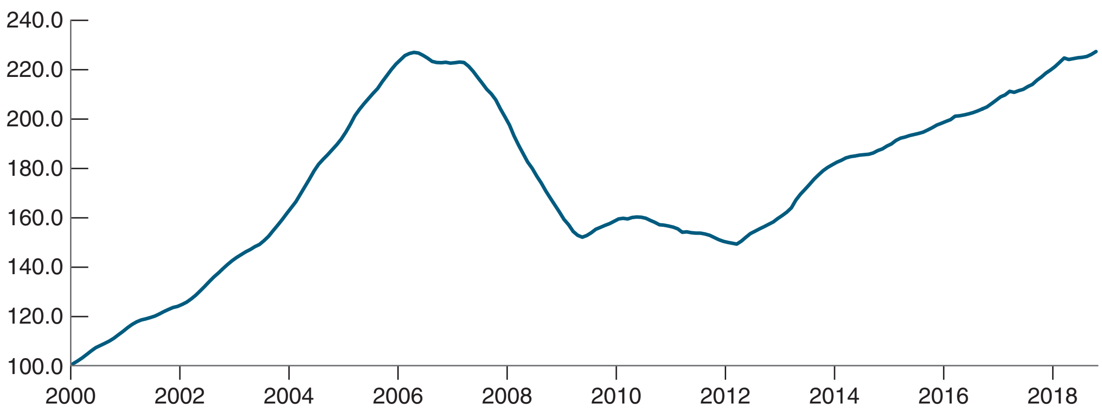


Figure 6-7

US Housing Prices since 2000

The increase in housing prices from 2000 to 2006 was followed by a sharp decline.

Source: FRED: Case-Shiller Home Price Indices, 10-city home price index, Series SPCS10RSA

Was the sharp price increase from 2000 to 2006 justified? In retrospect, and given the ensuing collapse, surely not. But at the time, when prices were increasing, economists were not so sure. Some increase in prices was clearly justified:

- The 2000s were a period of unusually low interest rates. Mortgage rates were low, increasing the demand for housing and thus pushing up the price.
- Other factors were also at work. **Mortgage lenders** became increasingly willing to make loans to more risky borrowers. These mortgages, known as **subprime mortgages** (**subprimes**), had existed since the mid-1990s but became more prevalent in the 2000s. By 2006, about 20% of all US mortgages were subprimes. Was this necessarily bad? Again, at the time, it was seen by most economists as a positive development. It allowed more poor people to buy homes, and under the assumption that housing prices would continue to increase, so the value of the mortgage would decrease over time relative to the price of the house, it looked safe for both lenders and borrowers. Judging from the past, the assumption that housing prices would not decrease also seemed reasonable. As you can see from Figure 6-7, housing prices had not decreased even during the 2000–2001 recession.

Even if people did not finance the purchase of a house by taking a mortgage, low interest rates would lead to an increase in the price of houses. More on this when we discuss present discounted values in Chapter 14.

Some economists were worried even as housing prices were going up. Robert Shiller, one of the two economists behind the Case-Shiller index, was among them, warning that the price increase was a bubble that would most likely crash. He received the Nobel Prize in 2013 for his work on asset prices.

In retrospect, these developments were much less benign than most economists thought. First, housing prices *could* go down, as became evident from 2006 on. When this happened, many borrowers found that the mortgage they owed now exceeded the value of their house (when this happens, the mortgage is said to be **underwater**). Second, it became clear that, in many cases, the mortgages were in fact much riskier than either the lender pretended or the borrower understood. In many cases, borrowers had taken mortgages with low initial interest rates, known as "teaser rates," and thus low initial interest payments, probably not fully realizing that their payments would increase sharply later on. Even if house prices had not declined, many of these borrowers would have been unable to meet their mortgage payments.

Some of these loans became known as NINJA loans (for "no income, no job, no assets").

Thus, as many borrowers defaulted, lenders found themselves faced with large losses. In mid-2008, losses on mortgages were estimated to be around $300 billion. This is a large number, but, relative to the size of the US economy, it is only about 2% of US GDP. One might have thought that the US financial system could absorb the shock and that the adverse effect on output would be limited. This was not to be. Although the trigger of the crisis was the decline in housing prices, its effects were enormously amplified. Even economists who had anticipated the housing price decline did not realize how strong the amplification mechanisms would be. To understand those, we return to the role of financial intermediaries.

The Role of Financial Intermediaries

In the previous section, we saw that high leverage, illiquidity of assets, and liquidity of liabilities all increase the risk of trouble in the financial system. All three elements were present in 2008, creating a perfect storm.

Leverage

Banks were highly leveraged. Why? For a number of reasons: First, they probably underestimated the risk they were taking: Times were good, and in good times, banks, just like people, tend to underestimate the risk of bad times. Second, the compensation and bonus system gave incentives to managers to go for high expected returns without fully taking the risk of bankruptcy into account. Third, although financial regulation required banks to keep their capital ratio above some minimum, banks found new ways to avoid the regulation by creating **structured investment vehicles (SIVs)**.

On the liability side, SIVs borrowed from investors, typically in the form of short-term debt. On the asset side, SIVs held various forms of securities. To reassure investors that they would get repaid, SIVs typically had a guarantee from the bank that had created them that, if needed, the bank would provide funds to the SIV. Although the first SIV was set up by Citigroup in 1988, SIVs rapidly grew in the 2000s. You may ask why banks did not simply do all these things on their own balance sheet rather than create a separate vehicle. The main reason was to be able to increase leverage. If the banks had done these operations themselves, the operations would have appeared on their balance sheet and been subject to regulatory capital requirements, forcing them to hold enough capital to limit the risk of bankruptcy. Doing these operations through an SIV did not require banks to put down capital. For that reason, by setting up an SIV, banks could increase leverage and increase expected profits, and they did.

When housing prices started declining and many mortgages turned out to be bad, the securities held by SIVs dropped in value. Questions arose about the solvency of the SIVs, and given the banks' guarantee to provide funds to the SIVs if needed, questions arose about the solvency of the banks themselves. Then, two other factors, securitization and wholesale funding, came into play.

Securitization and Illiquidity of Assets

An important financial development of the 1990s and 2000s was the growth of **securitization**. Traditionally, the financial intermediaries that made loans or issued mortgages kept them on their own balance sheet. This had obvious drawbacks. A local bank with local loans and mortgages on its books was much exposed to the local economic situation. When, for example, oil prices came down sharply in the mid-1980s and Texas was in recession, many local banks went bankrupt. With a more diversified portfolio of mortgages, say from many parts of the country, these banks might have avoided bankruptcy.

This is the idea behind securitization. Securitization is the creation of securities based on a bundle of assets (e.g., loans or mortgages). For instance, a **mortgage-backed security (MBS)** is a title to the returns from a bundle of mortgages, with the number of underlying mortgages often in the tens of thousands. The advantage is that many investors who would not want to hold individual mortgages will be willing to buy and hold these securities. This increase in the supply of funds from investors is, in turn, likely to decrease the cost of borrowing.

Securitization can go further. For example, instead of issuing identical claims to the returns on the underlying bundle of assets, one can issue different types of securities. For example, one can issue **senior securities**, which have first claims on the returns from the bundle, and **junior securities**, which come after and pay only if anything remains after the senior securities have been paid. Senior securities appeal to investors who want little risk; junior securities appeal to investors who are willing to take more risk. Such securities, known as **collateralized debt obligations (CDOs)**, were first issued in the late 1980s but grew in importance in the 1990s and 2000s. Securitization went even further with the creation of CDOs combining previously created CDOs, or CDO^2.

One of the obstacles to understanding the financial system is the alphabet soup of acronyms: SIVs, MBS, CDOs, etc. ►

Securitization would seem like a good idea, a way of diversifying risk and getting a larger group of investors involved in lending to households or firms. And, indeed, it is. But it also came with two large costs, which became clear during the crisis. The first was that if a bank sold a mortgage as part of a securitization bundle and thus did not keep it on its balance sheet, it had fewer incentives to make sure that the borrower could repay. The second was a risk that **rating agencies**, firms that assess the risk of various securities, had largely missed: When underlying mortgages went bad, assessing their value in an MBS, or, even more so, in a CDO, was extremely hard to do. These assets came to be known as **toxic assets**. This led investors to assume the worst and be reluctant either to hold them or to continue lending to institutions such as SIVs that held them. Many of the assets held by banks, SIVs, and other financial intermediaries were illiquid. They were extremely hard to assess and thus hard to sell except at fire sale prices.

Wholesale Funding and Liquidity of Liabilities

Yet another development of the 1990s and 2000s was the emergence of sources of finance other than checkable deposits by banks. Increasingly, they relied on borrowing from other banks or other investors, in the form of short-term debt, to finance the purchase of their assets, a process known as **wholesale funding**. SIVs, the financial entities set up by banks, were entirely funded through such wholesale funding.

Wholesale funding again would seem like a good idea, giving banks more flexibility in the amount of funds they could use to make loans or buy assets. But it had a cost, which became clear during the crisis. Although holders of checkable deposits were protected by deposit insurance and did not have to worry about the value of their deposits, this was not the case for other investors. Thus, when those investors worried about the value of the assets held by the banks or SIVs, they asked for their funds back. Banks and SIVs had liquid liabilities, much more liquid than their assets.

The combination of high leverage, illiquid assets, and liquid liabilities led to a major financial crisis. As housing prices declined and some mortgages went bad, high leverage implied a sharp decline in the capital of banks and SIVs. This forced them to sell some of their assets. Because these assets were often hard to value, the banks had to sell them at fire sale prices. This, in turn, decreased the value of similar assets remaining on their balance sheet, or on the balance sheet of other financial intermediaries, leading to a further decline in capital ratios and forcing further sales of assets and further declines in prices. The complexity of the securities held by banks and SIVs made it difficult to assess their solvency. Investors therefore became reluctant to continue to lend to them, and wholesale funding came to a stop, which forced further asset sales and price declines. Even the banks became reluctant to lend to each other. On September 15, 2008, Lehman Brothers, a major bank with more than $600 billion in assets, declared bankruptcy, leading financial participants to conclude that many, if not most, other banks and financial institutions were at risk. By late September, the financial system became paralyzed. Banks basically stopped lending to each other or to anyone else. Quickly, what had been largely a financial crisis turned into a macroeconomic crisis.

Macroeconomic Implications

The immediate effects of the financial crisis on the macroeconomy were twofold: first, a large increase in the interest rates at which people and firms could borrow, if they could borrow at all; second, a dramatic decrease in confidence.

We saw the effect on various interest rates in Figure 6-3. In late 2008, interest rates on highly rated (AAA) bonds increased to 7% and on lower-rated (BBB) bonds to 10%. Suddenly, borrowing became extremely expensive for most firms. And for the many firms too small to issue bonds and thus dependent on bank credit, it became nearly impossible to borrow at all.

The events of September 2008 also triggered wide anxiety among consumers and firms. Thoughts of another Great Depression and, more generally, confusion and fear about what was happening in the financial system led to a large drop in confidence. The evolution of consumer and business confidence indexes is shown in Figure 6-8; both are normalized to equal 100 in January 2007. Note how consumer confidence, which had started declining in mid-2007, took a sharp turn in the fall of 2008 and reached a low of 22 in early 2009, far below historical lows. The result of lower confidence and lower housing and stock prices was a sharp decrease in consumption.

◄ See the Focus Box "The Lehman Bankruptcy, Fears of Another Great Depression, and Shifts in the Consumption Function" in Chapter 3.

Policy Responses

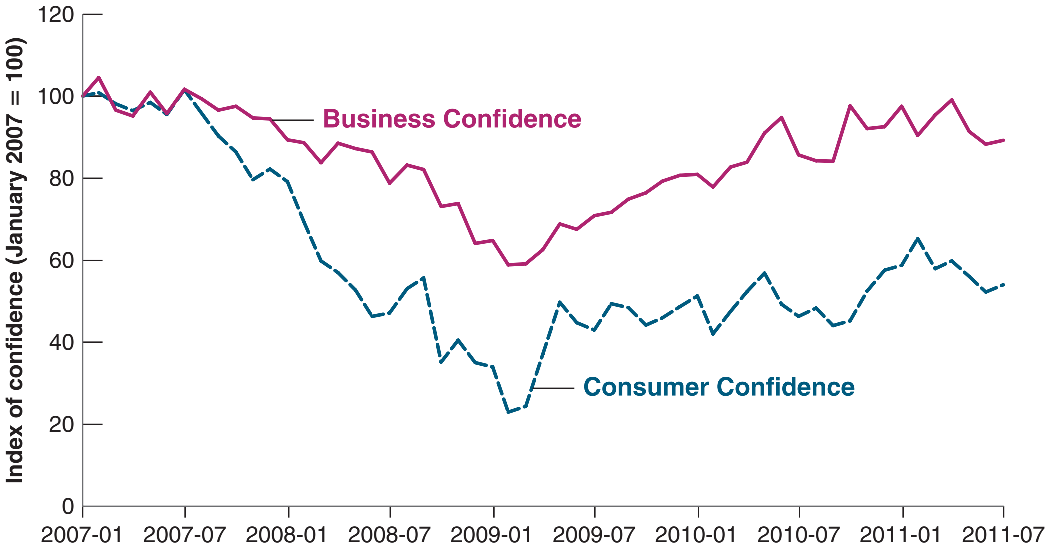
The high cost of borrowing, lower stock prices, and lower confidence all combined to decrease the demand for goods. In terms of the IS-LM model, there was a sharp adverse shift of the IS curve, just as we drew in Figure 6-5. In the face of this large decrease in demand, policymakers did not remain passive.

Figure 6-8

US Consumer and Business Confidence, 2007–2011

The financial crisis led to a sharp drop in confidence, which bottomed in early 2009.

Source: Bloomberg L.P.



Financial Policies

The most urgent measures were aimed at strengthening the financial system:

- To prevent a run by depositors, federal deposit insurance was increased from $100,000 to $250,000 per account. Recall, however, that much of banks' funding came not from deposits but from the issuance of short-term debt to investors. To allow the banks to continue to fund themselves through wholesale funding, the federal government offered a program guaranteeing new debt issues by banks.
- The Federal Reserve provided widespread liquidity to the financial system. We have seen that, if investors wanted to take their funds back, the banks had to sell some of their assets, often at fire sale prices. In many cases, this would have meant bankruptcy. To avoid this, the Fed put in place a number of **liquidity facilities** to make it easier for banks and other financial intermediaries to borrow from the Fed. The Fed also increased the set of assets that financial institutions could use as **collateral** when borrowing from the Fed (collateral refers to the asset a borrower pledges when borrowing from a lender; if the borrower defaults, the asset goes to the lender). Together, these facilities allowed banks and financial intermediaries to pay back investors without having to sell their assets. They also decreased the incentives of investors to ask for their funds by decreasing the risk that banks and financial intermediaries would go bankrupt.
- In addition, the government introduced the **Troubled Asset Relief Program (TARP)**, aimed at cleaning up banks. The initial goal of the $700 billion program, introduced in October 2008, was to remove the complex assets from bank balance sheets, thus decreasing uncertainty, reassuring investors, and making it easier to assess the health of each bank. The Treasury, however, faced the same problems as private investors. If the complex assets were going to be exchanged for, say, Treasury bills, at what price should the exchange be done? Within a few weeks, it became clear that the task of assessing the value of each of these assets was extremely hard and would take a long time, and the initial goal was abandoned. The new goal became to increase the capital of banks. This was done by the government's acquiring shares and thus providing funds to most of the largest US banks. By increasing their capital ratio, and thus decreasing their leverage, the banks could avoid bankruptcy and, over time, return to normal. As of the end of September 2009, total spending under the TARP was $360 billion, of which $200 billion was spent through the purchase of shares in banks.

At the time of writing, all banks have bought back their shares and reimbursed the government. Indeed, in the final estimation, TARP actually has made a small profit. ▶

Fiscal and monetary policies were used aggressively as well.

Monetary Policy

In the summer of 2007, the Fed had begun to worry about a slowdown in growth and started decreasing the policy rate, slowly at first, faster later as evidence of the crisis mounted. The evolution of the federal funds rate from 2000 on was shown in Figure 1-4 in Chapter 1. By December 2008, the rate was down to zero. But monetary policy was constrained by the zero lower bound and the policy rate could not be decreased further. The Fed then turned to what has become known as **unconventional monetary policy**, buying other assets so as to directly affect the rate faced by borrowers. We shall explore the various dimensions of unconventional monetary policy at more length in Chapter 23. Suffice it to say that, although these measures were useful, the efficacy of monetary policy was nevertheless severely constrained by the zero lower bound.

Recall that the interest rate faced by borrowers is given by $r + x$. You can think of conventional monetary policy as the choice of r, and unconventional monetary policy as measures to reduce x.

Fiscal Policy

When the size of the adverse shock became clear, the US government turned to fiscal policy. When the Obama administration assumed office in 2009, its first priority was to design a fiscal program that would increase demand and reduce the size of the recession. Such a fiscal program, called the **American Recovery and Reinvestment Act**, was passed in February 2009. It called for $780 billion in new measures, in the form of both tax reductions and spending increases, over 2009 and 2010. The US budget deficit increased from 1.7% of GDP in 2007 to a high of 9.0% in 2010. The increase was largely the mechanical effect of the crisis because the decrease in output led automatically to a decrease in tax revenues and an increase in transfer programs such as unemployment benefits. But it was also the result of the specific measures in the fiscal program aimed at increasing either private or public spending. Some economists argued that the increase in spending and the cuts in taxes should be even larger, given the seriousness of the situation. Others, however, worried that the deficit was becoming too large, that it might lead to an explosion of public debt, and that it had to be reduced. Beginning in 2011, the deficit was indeed reduced, down to 2.4% in 2015.

Since 2015, the deficit has increased, reaching 4% in 2018. This increase reflects in large part tax cuts implemented by the Trump administration. More on this in Chapter 22.

We can summarize our discussion by going back to the IS-LM model we developed in the previous section. This is done in Figure 6-9. The financial crisis led to a large shift of the IS curve to the left, from *IS* to *IS*′. In the absence of changes in policy, the equilibrium

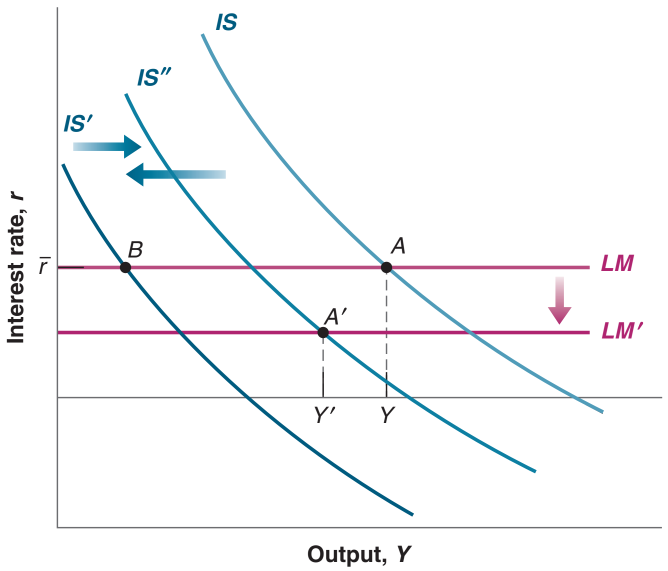


Figure 6-9

The Financial Crisis and the Use of Financial, Fiscal, and Monetary Policies

The financial crisis led to a shift of the IS to the left. Financial and fiscal policies led to some shift of the IS back to the right. Monetary policy led to a shift of the LM curve down. Policies were not enough, however, to avoid a major recession.

It is difficult to know what would have happened in the absence of those policies. It is reasonable to think, but impossible to prove, that the decrease in output would have been much larger, leading to a repeat of the Great Depression. ▶

would have moved from point A to point B. Financial and fiscal policies offset some of the shift, so that, instead of shifting to IS', the economy shifted to IS''. And monetary policy led to a shift of the LM down, from LM to LM', so the resulting equilibrium was at point A'. At that point, the zero lower bound on the nominal policy rate implied that the real policy rate could not be decreased further. The result was a decrease in output from Y to Y'. The initial shock was so large that the combination of financial, fiscal, and monetary measures was just not enough to avoid a large decrease in output. US GDP fell by 3.5% in 2009 and recovered only slowly thereafter.

SUMMARY

- The nominal interest rate tells you how many dollars you need to repay in the future in exchange for one dollar today.
- The real interest rate tells you how many goods you need to repay in the future in exchange for one good today.
- The real interest rate is approximately equal to the nominal rate minus expected inflation.
- The zero lower bound on the nominal interest rate implies that the real interest rate cannot be lower than expected inflation.
- The interest rate on a bond depends both on the risk that the bond issuer will default and on the bond holder's degree of risk aversion. A higher probability or a higher degree of risk aversion leads to a higher interest rate.
- Financial intermediaries receive funds from investors and then lend these funds to others. In choosing their leverage ratio, financial intermediaries trade off expected profit against the risk of insolvency.
- Because of leverage, the financial system is exposed to both in solvency and illiquidity risks. Both may lead financial intermediaries to decrease lending.
- The higher the leverage ratio, or the more illiquid the assets, or the more liquid the liabilities, the higher the risk of a bank run or a run on financial intermediaries.
- The IS-LM model must be extended to take into account the difference between the nominal and the real interest rate, and the difference between the policy rate chosen by the central bank and the interest rate at which firms and people can borrow.
- A shock to the financial system leads to an increase in the interest rate at which people and firms can borrow for a given policy rate. It leads to a decrease in output.
- The financial crisis of the late 2000s was triggered by a decrease in housing prices. It was amplified by the financial system.
- At the time of the crisis, financial intermediaries were highly leveraged. Because of securitization, their assets were hard to assess and thus illiquid. Because of wholesale funding, their liabilities were liquid. Runs forced financial intermediaries to reduce lending, with strong adverse effects on output.
- Financial, fiscal, and monetary policies were used. They were not enough, however, to prevent a deep recession.

KEY TERMS

nominal interest rate, 110
real interest rate, 110
risk premium, 114
risk aversion, 114
direct finance, 115
shadow banking, 116
capital ratio, 116
leverage ratio, 116
insolvency, 116
fire sale prices, 117
demand deposits, 118
bank runs, 118
narrow banking, 118
federal deposit insurance, 118
liquidity provision, 118
liquidity, 119
policy rate, 120
borrowing rate, 120
mortgage lenders, 123
subprime mortgages, or subprimes, 123
underwater, 123
structured investment vehicles (SIVs), 123
securitization, 124
mortgage-backed security (MBS), 124
senior securities, 124
junior securities, 124
collateralized debt obligations (CDOs), 124
rating agencies, 124
toxic assets, 124
wholesale funding, 125
liquidity facilities, 126
collateral, 126
Troubled Asset Relief Program (TARP), 126
unconventional monetary policy, 127
American Recovery and
Reinvestment Act, 127

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The nominal interest rate is measured in terms of goods; the real interest rate is measured in terms of money.
b. As long as expected inflation remains roughly constant, the movements in the real interest rate are roughly equal to the movements in the nominal interest rate.
c. The nominal policy interest rate was at the zero lower bound in the United States in 2019.
d. When expected inflation increases, the real rate of interest falls.
e. All bonds have equal risk of default and thus pay equal rates of interest.
f. The nominal policy interest rate is set by the central bank.
g. An increase in a bank's leverage ratio tends to increase both the expected profit of the bank and the risk of the bank going bankrupt.
h. The real borrowing rate and the real policy rate always move in the same direction.
i. It can be difficult to value assets of banks and other financial intermediaries, particularly in a financial crisis.
j. When a bank has high leverage and low liquidity, it may have to sell assets at fire sale prices.
k. Banks and other financial intermediaries have assets that are less liquid than their liabilities.
l. House prices have risen constantly since the year 2000.
m. The fiscal stimulus program adopted by the United States in response to the financial crisis helped offset the decline in aggregate demand and reduce the size of the recession.
n. The fiscal stimulus program adopted by the United States included a large increase in the deficit measured as a percent of GDP.

2. Compute the real interest rate using the exact formula and the approximation formula for each set of assumptions listed in (a) through (c).

a. $i = 4\%; \quad \pi^e = 2\%$
b. $i = 15\%; \pi^e = 11\%$
c. $i = 54\%; \pi^e = 46\%$

3. Fill in the table below and answer the questions that relate to the data in the table.

Situation	Nominal policy interest rate	Expected inflation	Real policy interest rate	Risk premium	Nominal borrowing interest rate	Real borrowing interest rate
A	3	0		0		
B	4		2	1		
C	0	2		4		
D				2	6	3
E	0	–2				5

a. Which situations correspond to a liquidity trap as defined in Chapter 4?
b. Which situations correspond to the case where the nominal policy interest rate is at the zero lower bound?
c. Which situation has the highest risk premium? What two factors in bond markets lead to a positive risk premium?
d. Why is it so important when the nominal policy interest rate is at the zero lower bound to maintain a positive expected rate of inflation?

4. Modern bank runs

Consider a simple bank that has assets of 100, capital of 20, and checking deposits of 80. Recall from Chapter 4 that checking deposits are liabilities of a bank.

a. Set up the bank's balance sheet.
b. Now suppose that the perceived value of the bank's assets falls by 10. What is the new value of the bank's capital? What is the bank's leverage ratio?
c. Suppose the deposits are insured by the government. Despite the decline in the value of bank capital, is there any immediate reason for depositors to withdraw their funds from the bank? Would your answer change if the perceived value of the bank's assets fell by 15? 20? 25? Explain.

Now consider a different sort of bank, still with assets of 100 and capital of 20, but now with short-term credit of 80 instead of checkable deposits. Short-term credit must be repaid or rolled over (borrowed again) when it comes due.

d. Set up this bank's balance sheet.
e. Again suppose that the perceived value of the bank's assets falls. If lenders are nervous about the solvency of the bank, will they be willing to continue to provide short-term credit to the bank at low interest rates?
f. Assuming that the bank cannot raise additional capital, how can it raise the funds necessary to repay its debt coming due? If many banks are in this position at the same time (and if banks hold similar kinds of assets), what will likely happen to the value of their assets? How will this affect the willingness of lenders to provide short-term credit?

5. The IS-LM view of the world with more complex financial markets

Consider an economy described by Figure 6-6 in the text.

a. What are the units on the vertical axis of Figure 6-6?
b. If the nominal policy interest rate is 5% and the expected rate of inflation is 3%, what is the value for the vertical intercept of the LM curve?
c. Suppose the nominal policy interest rate is 5%. If expected inflation decreases from 3% to 2%, in order to keep the LM curve from shifting in Figure 6-6, what must the central bank do to the nominal policy rate of interest?
d. If the expected rate of inflation were to decrease from 3% to 2%, with the nominal policy rate unchanged, does the IS curve shift?
e. If the expected rate of inflation were to decrease from 3% to 2%, does the LM curve shift?
f. If the risk premium on risky bonds increases from 5% to 6%, does the LM curve shift?

g. If the risk premium on risky bonds increases from 5% to 6%, does the IS curve shift?
h. What are the fiscal policy options that prevent an increase in the risk premium on risky bonds from decreasing the level of output?
i. What are the monetary policy options that prevent an increase in the risk premium on risky bonds from decreasing the level of output?

DIG DEEPER

6. Nominal and real interest rates around the world

a. There are a few episodes of negative nominal interest rates around the world. Some may or may not be in play as you read this book. The Swiss nominal policy rate (the Swiss equivalent of the federal funds rate) is series IRSTCI01CHM156N from the FRED database maintained at the Federal Reserve Bank of St. Louis. The Swiss nominal policy rate was negative from 2014 to 2018. If so, why not hold cash instead of bonds?
b. The real rate of interest is frequently negative; see Figure 6-2. Under what circumstances can it be negative? If so, why not just hold cash instead of bonds?
c. What are the effects of a negative real interest rate on borrowing and lending?
d. Find a recent issue of *The Economist* and look at the table in the back, titled "Economic and financial indicators." Use the three-month money market rate as a proxy for the nominal policy interest rate, and the most recent three-month rate of change in consumer prices as a measure of the expected rate of inflation (both are expressed in annual terms). Which countries have the lowest nominal interest rates? Do any countries have a negative nominal policy rate? Which countries have the lowest real interest rates? Are some of these real interest rates negative?

7. The Troubled Asset Relief Program (TARP)

Consider a bank that has assets of 100, capital of 20, and short-term credit of 80. Among the bank's assets are securitized assets whose value depends on the price of houses. These assets have a value of 50.

a. Set up the bank's balance sheet.

Suppose that as a result of a housing price decline, the value of the bank's securitized assets falls by an uncertain amount, so that these assets are now worth somewhere between 25 and 45. Call the securitized assets "troubled assets." The value of the other assets remains at 50. As a result of the uncertainty about the value of the bank's assets, lenders are reluctant to provide any short-term credit to the bank.

b. Given the uncertainty about the value of the bank's assets, what is the range in the value of the bank's capital?

As a response to this problem, the government considers purchasing the troubled assets, with the intention of reselling them again when the markets stabilize. (This is the original version of the TARP.)

c. If the government pays 25 for the troubled assets, what will be the value of the bank's capital? How much would the government have to pay for the troubled assets to ensure that the bank's capital does not have a negative value? If the government pays 45 for the troubled assets, but the true value turns out to be much lower, who bears the cost of this mistaken valuation? Explain.

Suppose that instead of buying the troubled assets, the government provides capital to the bank by buying ownership shares, with the intention of reselling the shares when the markets stabilize. (This is what the TARP became.) The government exchanges Treasury bonds (which become assets for the bank) for ownership shares.

d. Suppose the government exchanges 25 Treasury bonds for ownership shares. Assuming the worst-case scenario (so that the troubled assets are worth only 25), set up the new balance sheet of the bank. (Remember that it now has three assets: 50 of untroubled assets, 25 of troubled assets, and 25 of Treasury bonds.) What is the total value of the bank's capital? Will the bank be insolvent?
e. Given your answers and the material in the text, why might recapitalization be a better policy than buying the troubled assets?

8. Calculating the risk premium on bonds

$$(1 + i) = (1 - p)(1 + i + x) + p(0)$$

p is the probability that the bond does not pay at all (the bond issuer is bankrupt) and has a zero return.

i is the nominal policy interest rate.

x is the risk premium.

a. If the probability of bankruptcy is zero, what is the rate of interest on the risky bond?
b. Calculate the probability of bankruptcy when the nominal interest rate for a risky borrower is 8% and the nominal policy rate of interest is 3%.
c. Calculate the nominal interest rate for a borrower when the probability of bankruptcy is 1% and the nominal policy rate of interest is 4%.
d. Calculate the nominal interest rate for a borrower when the probability of bankruptcy is 5% and the nominal policy rate of interest is 4%.
e. The formula assumes that payment upon default is zero. In fact, it is often positive. How would you change the formula in this case?

9. Unconventional monetary policy: financial policy and quantitative easing

We have written the IS-LM model in the following terms:

IS relation: $$Y = C(Y - T) + I(Y, r + x) + G \quad (6.5)$$

LM relation: $$r = \bar{r} \quad (6.6)$$

Interpret the real policy rate as the federal funds rate adjusted for expected inflation. Assume that the rate at which firms can borrow is much higher than the federal funds rate, equivalently that the premium, x, in the IS equation is high.

a. Suppose that the government takes action to improve the solvency of the financial system. If the government's action is successful and banks become more willing to lend—both to one another and to nonfinancial firms—what is likely to happen to the premium? What will happen to the IS-LM diagram based on Figure 6-6? Can we consider financial policy as a kind of macroeconomic policy?

b. Faced with a zero nominal interest rate, suppose the Fed decides to purchase securities directly to facilitate the flow of credit in the financial markets. This policy is called quantitative easing. If quantitative easing is successful, so that it becomes easier for financial and nonfinancial firms to obtain credit, what is likely to happen to the premium? What effect will this have on the IS-LM diagram? If quantitative easing has some effect, is it true that the Fed has no policy options to stimulate the economy when the federal funds rate is zero?

c. We will see later in the course that one argument for quantitative easing is that it increases expected inflation. Suppose it does. How does that affect the LM curve in Figure 6-6?

EXPLORE FURTHER

10. The spread between riskless and risky bonds

The text used Figure 6-3 to describe fluctuations in the spreads between riskless rate on 10-year US Treasury bonds and 10-year AAA and BBB corporate bonds. This figure can be updated by going to the Federal Reserve Bank of St. Louis FRED database. The 10-year Treasury bond yield is variable DGS10; Moody's 10-year seasoned AAA bond is series DAAA; and the Bank of America BBB bond yield is series BAMLC0A4CBBBEY.

a. Find the values of these three yields for the day closest to the day you are looking at this question. Which is the highest yield and which is the lowest yield? What is the spread between the BBB and AAA yield? What is the spread between the BBB and Treasury yield?

b. Now go back one calendar year and find the same yields and calculate the three spreads. Fill in the table below:

Date	BBB	AAA	Treasury	BBB-AAA	AAA-Treasury	BBB-Treasury
Today						
One year ago						

c. Do you see any evidence of a change in the risk premium over the past year or has it been relatively stable? Explain.

11. Inflation-indexed bonds

Some bonds issued by the US Treasury make payments indexed to inflation, compensating investors for inflation. Therefore, the current interest rates on these bonds are real interest rates—interest rates in terms of goods. These interest rates can be used, together with nominal interest rates, to provide a measure of expected inflation. Let's see how.

Go to the website of the Federal Reserve Board and get the most recent statistical release listing interest rates (www.federalreserve.gov/releases/h15/Current). Find the current nominal interest rate on Treasury securities with a five-year maturity. Now find the current interest rate on "inflation-indexed" Treasury securities with a five-year maturity. What do you think participants in financial markets think the average inflation rate will be over the next five years?

FURTHER READINGS

- A more in-depth description of what happened, but along the same lines as this chapter, is given in "What Happened: Financial Factors in the Great Recession," by Mark Gertler and Simon Gilchrist, *Journal of Economic Perspectives*, 2018, 32(3), pp. 3–30.
- There are many good books on the crisis, among them Michael Lewis's *The Big Short* (2010, W. W. Norton) and Gillian Tett's *Fool's Gold* (2009, Free Press). Both books show how the financial system became increasingly risky until it finally collapsed. Both read like detective novels, with a lot of action and fascinating characters. *The Big Short* was made into a movie in 2015.
- *In Fed We Trust* (2009, Crown Business), written by David Wessel, a former economics editor of the *Wall Street Journal*, describes how the Fed reacted to the crisis. It also makes for fascinating reading. Read also the insider version, *The Courage to Act: A Memoir of a Crisis and Its Aftermath* (2015, W. W. Norton), by Ben Bernanke, who was chairman of the Fed throughout the crisis.

This page is intentionally left blank

THE CORE

The Medium Run

In the medium run, the economy returns to a level of output associated with the natural rate of unemployment.

Chapter 7

Chapter 7 looks at equilibrium in the labor market. It characterizes the natural rate of unemployment, which is the unemployment rate to which the economy tends to return in the medium run.

Chapter 8

Chapter 8 looks at the relation between inflation and unemployment, a relation known as the Phillips curve. In the short run, unemployment typically deviates from its natural rate. The behavior of inflation depends on the deviation of unemployment from its natural rate. In the medium run, unemployment returns to the natural rate.

Chapter 9

Chapter 9 presents a model of the short run and the medium run. The model puts together the IS-LM model and the Phillips curve and thus is called the IS-LM-PC model. It describes the dynamics of output and unemployment, in both the short and the medium run.

This page is intentionally left blank

7 The Labor Market

Think about what happens when firms respond to an increase in demand by increasing production. Higher production leads to higher employment. Higher employment leads to lower unemployment. Lower unemployment leads to higher wages. Higher wages increase production costs, leading firms to increase prices. Higher prices lead workers to ask for higher wages. Higher wages lead to further increases in prices, and so on.

So far, we have simply ignored this sequence of events. By assuming a constant price level in the IS-LM model, we in effect assumed that firms were able and willing to supply any amount of output at a given price level. As long as our focus was on the *short run*, this assumption was fine. But, as our attention now turns to the *medium run*, we must now abandon this assumption, explore how prices and wages adjust over time, and how this affects output. This will be our task in this and the next two chapters.

◀ Recall the behavior of the price level in Figure 5-10.

At the center of the sequence of events described in the first paragraph is *the labor market*, which is the market in which wages are determined. This chapter focuses on the labor market. It has six sections:

Section 7-1 provides an overview of the labor market.

Section 7-2 focuses on unemployment, how it moves over time, and how its movements affect individual workers.

Sections 7-3 and 7-4 look at wage and price determination.

Section 7-5 then looks at equilibrium in the labor market. It characterizes the *natural rate of unemployment*, which is the rate of unemployment to which the economy tends to return in the medium run.

Section 7-6 gives a map of where we will be going next.

If you remember one basic message from this chapter, it should be: The natural rate of unemployment is the rate at which the wage demands of workers are consistent with the price decisions of firms. ▶▶▶

7-1 A TOUR OF THE LABOR MARKET

The total US population in 2018 was 327.2 million (Figure 7-1). Excluding those who were either younger than working age (under 16), in the armed forces, or behind bars, the number of people potentially available for civilian employment, the **noninstitutional civilian population**, was 257.7 million.

The civilian **labor force**, which is the sum of those either working or looking for work, was only 162.0 million. The other 95.7 million people were **out of the labor force**, neither working in the marketplace nor looking for work. The **participation rate**, which is defined as the ratio of the labor force to the noninstitutional civilian population, was therefore 162.0/257.7, or 62%. The participation rate has steadily increased over time, reflecting mostly the increasing participation rate of women. In 1950, one woman out of three was in the labor force; now the number is close to two out of three.

Work in the home, such as cooking or raising children, is not classified as work in the official statistics. This is a reflection of the difficulty of measuring these activities, not a value judgment about what constitutes work and what does not.

Of those in the labor force, 155.7 million were employed and 6.3 million were unemployed—looking for work. The **unemployment rate**, which is defined as the ratio of the unemployed to the labor force, was thus 3.9%.

The Large Flows of Workers

To get a sense of what a given unemployment rate implies for individual workers, consider the following analogy.

Take an airport full of passengers. It may be crowded because many planes are coming and going, and many passengers are quickly moving in and out of the airport. Or it may be because bad weather is delaying flights and passengers are stranded, waiting for the weather to improve. The number of passengers in the airport will be high in both cases, but their plights are quite different. Passengers in the second scenario are likely to be much less happy.

In the same way, a given unemployment rate may reflect two different realities. It may reflect an active labor market, with many **separations** and many **hires**, and thus many workers entering and exiting unemployment, or it may reflect a sclerotic labor market, with few separations, few hires, and a stagnant unemployment pool.

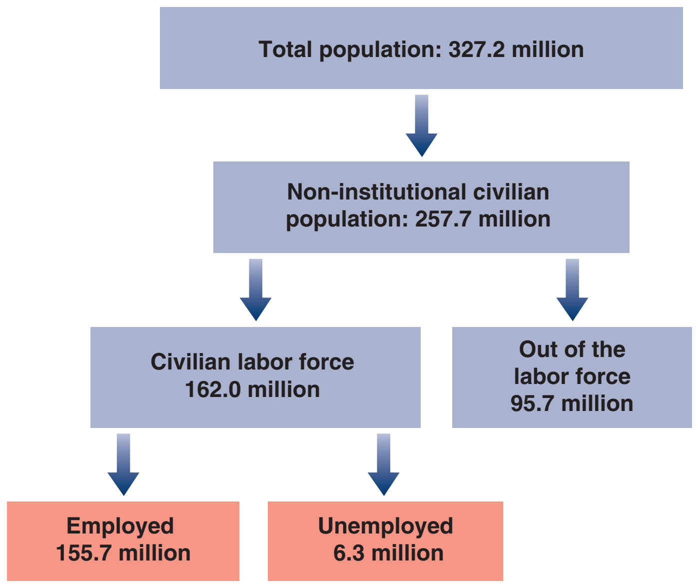
Sclerosis, a medical term, means "hardening of the arteries." By analogy, it is used in economics to describe markets that function poorly and have few transactions.

Finding out which reality hides behind the aggregate unemployment rate requires data on the movements of workers. In the United States, the data are available from

Figure 7-1

Population, Labor Force, Employment, and Unemployment in the United States (in Millions), 2018

Source: Current Population Survey www.bls.gov/cps/.



a monthly survey called the **Current Population Survey (CPS)**. Average monthly flows, computed from the CPS from 1994 to 2018, are reported in Figure 7-2. (For more on the ins and outs of the CPS, see the Focus Box "The Current Population Survey.")

Figure 7-2 has three striking features.

- The flows of workers in and out of employment are large.

  On average, there are 8.5 million separations each month in the United States (out of an employment pool of 132.0 million), 3.0 million people change jobs (shown by the circular arrow at the top), 3.7 million move from employment out of the labor force, and 1.8 million move from employment to unemployment.

  Why are there so many separations each month? About three-fourths of all separations are **quits**, workers leaving their jobs for what they perceive as a better alternative. The remaining one-fourth are **layoffs**. Layoffs come mostly from changes in employment levels across firms. The slowly changing aggregate employment numbers hide a reality of continual job destruction and job creation across firms. At any given time, some firms are suffering decreases in demand and decreasing their employment, while other firms are enjoying increases in demand and increasing employment.

- The flows in and out of unemployment are large relative to the number of unemployed. The average monthly flow out of unemployment each month is 3.8 million: 2.0 million people get a job, and 1.9 million stop searching for a job and drop out of the labor force. Put another way, the proportion of unemployed people leaving unemployment equals 3.8/8.8 or about 44% each month. Put yet another way, the average **duration of unemployment**, which is the average length of time people spend unemployed, is between two and three months.

  This fact has an important implication. You should not think of unemployment in the United States as a stagnant pool of workers waiting indefinitely for jobs. For most (but obviously not all) of the unemployed, being unemployed is more a quick transition than a long wait between jobs. One needs, however, to make two remarks at this point. First, the United States is unusual in this respect. In many European countries, the average duration is much longer than in the United States. Second, as we shall see, even in the United States, when unemployment is high, as was the

The numbers for employment, unemployment, and those out of the labor force in Figure 7-1 refer to 2018. The numbers for the same variables in Figure 7-2 refer to averages from 1994 to 2018. This is why they are different.

Put in another, and perhaps more dramatic, way: On average, every day in the United States, about 60,000 workers become unemployed.

The average duration of unemployment equals the inverse of the proportion of unemployed people leaving unemployment each month. To see why, consider an example. Suppose the number of unemployed is constant and equal to 100, and each unemployed person remains unemployed for two months. So, at any given time, there are 50 people who have been unemployed for one month and 50 who have been unemployed for two months. Each month, the 50 unemployed who have been unemployed for two months leave unemployment. In this example, the proportion of unemployed leaving unemployment each month is 50/100, or 50%. The duration of unemployment is two months, which is the inverse of 1/50%.

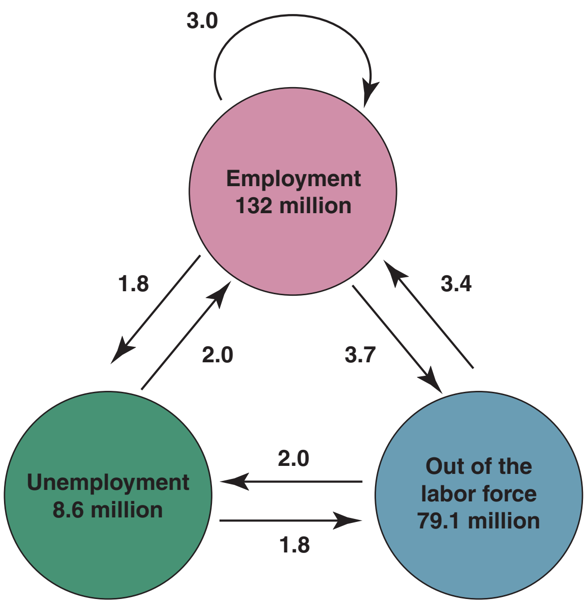


Figure 7-2

Average Monthly Flows between Employment, Unemployment, and Nonparticipation in the United States, 1994 to 2018 (Millions)

(1) The flows of workers in and out of employment are large. (2) The flows in and out of unemployment are large relative to the number of unemployed. (3) There are also large flows in and out of the labor force, much of it directly to and from employment.

Source: Calculated from the series constructed by Fleischman and Fallick, www.federalreserve.gov/econresdata/researchdata/feds200434.xls.

FOCUS

The Current Population Survey

The Current Population Survey (CPS) is the main source of statistics on the labor force, employment, participation, and earnings in the United States.

When the CPS began in 1940, it was based on interviews of 8,000 households. The sample has grown considerably, and now about 60,000 households are interviewed every month. The households are chosen so that the sample is representative of the US population. Each household stays in the sample for four months, leaves the sample for the following eight months, then comes back for another four months before leaving the sample permanently.

The survey is now based on computer-assisted interviews. Interviews are done either in person, in which case interviewers use laptop computers, or by phone. Some questions are asked in every survey. Other questions are specific to a particular survey and are used to find out about particular aspects of the labor market.

The Labor Department uses the data to compute and publish numbers on employment, unemployment, and participation by age, gender, education, and industry. Economists use these data, which are available in large computer files, in two ways.

The first is to get snapshots of how things are at various points in time, to answer such questions as: What is the distribution of wages for Hispanic American workers with only primary education, and how does it compare with the same distribution 10 or 20 years ago?

The second way, of which Figure 7-2 is an example, relies on the fact that the survey follows people through time. By looking at the same people in two consecutive months, economists can find out, for example, how many of those who were unemployed last month are employed this month. This number gives them an estimate of the probability that somebody who was unemployed last month found a job this month.

For more on the CPS, go to the CPS homepage (www.bls.gov/cps/home.htm).

case during the crisis, the average duration of unemployment becomes much longer. Being unemployed becomes much more painful.

- The flows in and out of the labor force are also surprisingly large. Each month, 5.5 million workers drop out of the labor force (3.7 plus 1.8), and a roughly equal slightly larger number, 5.4 million, join the labor force (3.4 plus 2.0). You might have expected these two flows to be composed, on one side, of those completing school and entering the labor force for the first time, and on the other side, of workers entering retirement. But each of these two groups actually represents a small fraction of the total flows. Each month only about 450,000 new people enter the labor force, and about 350,000 retire. But the actual flows in and out of the labor force are 10.9 million, about 14 times larger.

What this fact implies is that many of those classified as "out of the labor force" are in fact willing to work and move back and forth between participation and nonparticipation. Indeed, among those classified as out of the labor force, a large proportion report that although they are not looking, they "want a job." What they really mean by this statement is unclear, but the evidence is that many do take jobs when offered them.

This fact has another important implication. The sharp focus on the unemployment rate by economists, policymakers, and news media is partly misdirected. Some of the people classified as out of the labor force are much like the unemployed. They are in effect **discouraged workers**. And although they are not actively looking for a job, they will take it if they find one.

Conversely, some of the unemployed may be unwilling to accept any job offered to them and should probably not be counted as unemployed because they are not really looking for a job.

This is why economists sometimes focus on the **employment rate**, which is the ratio of employment to the population available for work, rather than on the unemployment rate. The higher unemployment, or the higher the number of people out of the labor force, the lower the employment rate.

In 2018, employment was 155.7 million and the population available for work was 257.7 million. The employment rate was 60.4%. The employment rate is sometimes called the *employment-to-population ratio*. (Make sure you understand why the sum of the employment rate and the unemployment rate is not equal to 1. *Hint*: look at the denominators.)

I shall follow tradition in this text and focus on the unemployment rate as an indicator of the state of the labor market, but you should keep in mind that the unemployment rate is not the best estimate of the number of people available for work.

7-2 MOVEMENTS IN UNEMPLOYMENT

Let's now look at movements in unemployment. Figure 7-3 shows the average annual value of the US unemployment rate since 1948. The shaded bars represent years when there was a recession.

Figure 7-3 has two important features.

- Year-to-year movements in the unemployment rate are closely associated with recessions and expansions. Look, for example, at the last four peaks in unemployment in Figure 7-3. The most recent peak, at 9.6% in 2010, was the result of the crisis. The previous two peaks, associated with the recessions of 2001 and 1990–1991, were much lower, around 7%. Only the recession of 1982, when the unemployment rate reached 9.7%, is comparable to the recent crisis. (Annual averages can mask larger values within the year. In the 1982 recession, although the average unemployment rate over the year was 9.7%, the rate actually reached 10.8% in November. Similarly, the monthly unemployment rate in the Great Financial Crisis peaked at 10.0% in October 2009.)
- There does not seem to be much of an underlying trend. Until the mid-1980s, it looked as if the US unemployment rate was on an upward trend, from an average of 4.5% in the 1950s to 4.7% in the 1960s, 6.2% in the 1970s, and 7.3% in the 1980s. From the 1980s on, however, the unemployment rate generally declined for more than two decades. By 2006, it was down to 4.6%. The unemployment rate increased sharply with the crisis, but has come down again. At the time of writing, it stands at 3.9%, the lowest rate since 1968.

Note also that the unemployment rate sometimes peaks in the year after the recession, not in the actual recession year. This occurred, for example, in the 2001 recession. The reason is that, although growth is positive, so the economy is technically no longer in recession, the additional output does not lead to enough new hires to reduce the unemployment rate.

How do these fluctuations in the overall unemployment rate affect individual workers? This is an important question because the answer determines both:

- The effect of movements in the aggregate unemployment rate on the welfare of individual workers, and
- The effect of the aggregate unemployment rate on wages.

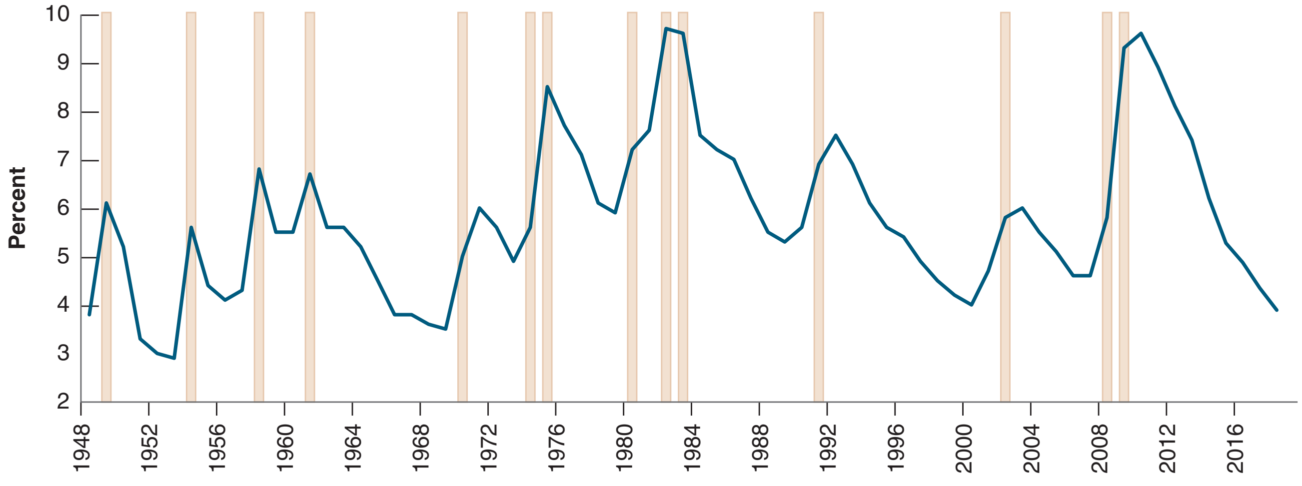


Figure 7-3

Movements in the US Unemployment Rate, 1948–2018

Since 1948, the average yearly US unemployment rate has fluctuated between 3% and 10%.

Source: Series UNRATE: Federal Reserve Economic Data (FRED) http://research.stlouisfed.org/fred2/.

Let's start by asking how firms can decrease their employment in response to a decrease in demand. They can hire fewer new workers, or they can lay off the workers they currently employ. Typically, firms prefer to slow or stop the hiring of new workers first, relying on quits and retirements to achieve a decrease in employment. But this may not be enough if the decrease in demand is large, so firms may then have to lay off workers.

Now think about the implications for both employed and unemployed workers:

- If the adjustment takes place through fewer hires, the chance that an unemployed worker will find a job diminishes. Fewer hires mean fewer job openings; higher unemployment means more job applicants. Fewer openings and more applicants combine to make it harder for the unemployed to find jobs.
- If the adjustment takes place instead through layoffs, then employed workers are at a higher risk of losing their job.

In general, as firms do both, higher unemployment is associated with both a lower chance of finding a job if one is unemployed and a higher chance of losing it if one is employed. Figures 7-4 and 7-5 show these two effects at work over the period 1996 to 2018.

Figure 7-4 plots two variables against time: the unemployment rate (measured on the left vertical axis) and the proportion of unemployed workers becoming employed each month (measured on the right vertical axis). This proportion is constructed by dividing the flow from unemployment to employment during each month by the number of unemployed at the start of the month. To show the relation between the two variables more clearly, the proportion of unemployed becoming employed is plotted on an inverted scale. Be sure to note that on the right vertical scale, the proportion is lowest at the top and highest at the bottom.

The relation between movements in the proportion of unemployed workers becoming employed and the unemployment rate is striking. Periods of higher unemployment are associated with much lower proportions of unemployed workers becoming employed. In 2010, for example, with unemployment close to 10%, only about 17% of the unemployed became employed within a month, as opposed to 28% in 2007, when unemployment was much lower.

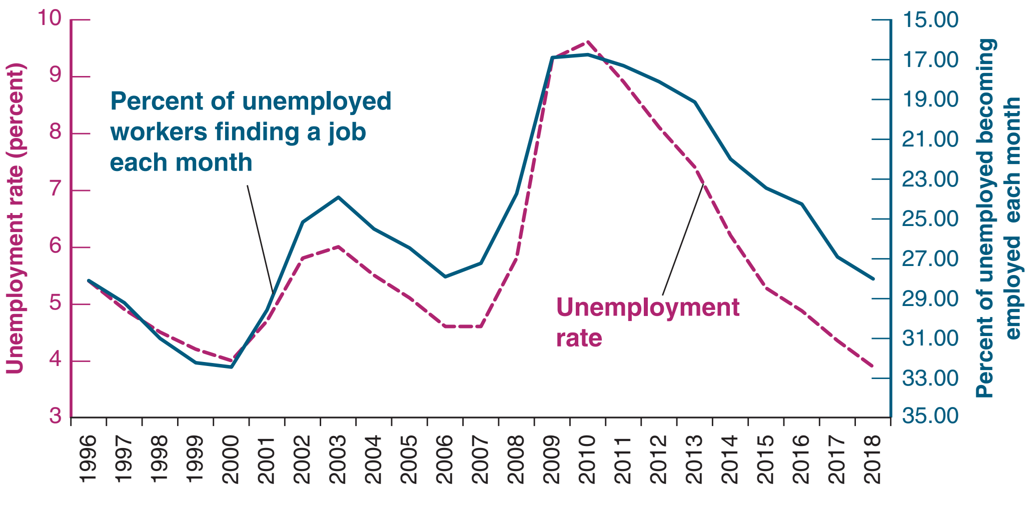
Similarly, Figure 7-5 plots two variables against time: the unemployment rate (measured on the left vertical axis) and the proportion of workers who become unemployed

Figure 7-4

The Unemployment Rate and the Proportion of Unemployed Becoming Employed within a Month, 1996–2018

When unemployment is higher, the proportion of unemployed becoming employed within one month is lower. Note that the scale on the right is an inverted scale.

Sources: FRED: Unemployment: UNRATE; Percent of unemployed workers becoming employed each month: Series constructed by Fleischman and Fallick, www.federalreserve.gov/econresdata/researchdata/.



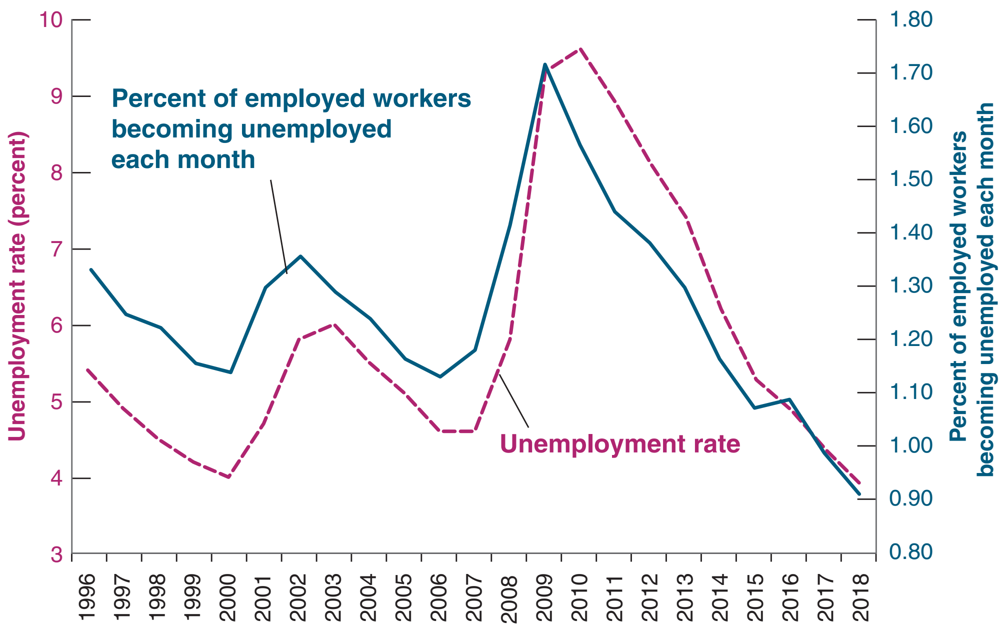


Figure 7-5

The Unemployment Rate and the Proportion of Workers Becoming Unemployed, 1996–2018

When unemployment is higher, a higher proportion of workers become unemployed.

Sources: FRED. Unemployment rate UNRATE: Federal Reserve Economic Data; Proportion of employed workers becoming unemployed each month: Fleischman and Fallick, www.federalreserve.gov/econresdata/research data/feds200434.xls.

each month (measured on the right vertical axis). This proportion is constructed by dividing the flow from employment to unemployment during each month by the number of employed in the month. The relation between the proportion of workers who become unemployed and the unemployment rate plotted is quite strong. Higher unemployment implies a higher probability for employed workers to become unemployed. The probability nearly doubles between times of low unemployment and times of high unemployment.

We could look instead at the separation rate, the sum of the flows to unemployment and out of the labor force. It would yield similar conclusions.

Let's summarize:

When unemployment is high, workers are worse off in two ways:

- Unemployed workers face a lower probability of becoming employed; equivalently, they can expect to remain unemployed for a longer time.
- Employed workers face a higher probability of becoming unemployed.

7-3 WAGE DETERMINATION

Having looked at unemployment, let's turn to wage determination and to the relation between wages and unemployment.

Wages are set in many ways. Sometimes they are set by **collective bargaining**, that is, bargaining between firms and unions. In the United States, collective bargaining plays a limited role, especially outside the manufacturing sector. Today, barely more than 10% of US workers have their wages set by collective bargaining agreements. For the rest, wages are either set by employers or by bargaining between the employer and individual employees. The higher the skills needed to do the job, the more likely there is to be bargaining. Wages offered for entry-level jobs at McDonald's are on a take-it-or-leave-it basis. New college graduates, on the other hand, can typically negotiate a few aspects of their contracts. CEOs and baseball stars can negotiate a lot more.

Collective bargaining is bargaining between a union (or a set of unions) and a firm (or a set of firms).

There are also large differences across countries. Collective bargaining plays an important role in Japan and in most European countries. Negotiations may take place at the firm level, at the industry level, or at the national level. Sometimes contract agreements apply only to firms that have signed the agreement. Sometimes they are automatically extended to all firms and all workers in the sector or the economy.

Given these differences across workers and across countries, can we hope to formulate anything like a general theory of wage determination? Yes. Although institutional differences influence wage determination, there are common forces at work in all countries. Two sets of facts stand out:

- Workers are typically paid a wage that exceeds their **reservation wage**, which is the wage that would make them indifferent between working or being unemployed. In other words, most workers are paid a high enough wage that they prefer being employed to being unemployed.
- Wages typically depend on labor market conditions. The lower the unemployment rate, the higher the wages. (I shall state this more precisely in the next section.)

To think about these facts, economists have focused on two broad lines of explanation. The first is that even in the absence of collective bargaining, most workers have some bargaining power, which they can and do use to obtain wages above their reservation wages. The second is that firms themselves may, for a number of reasons, want to pay wages higher than the reservation wage. Let's look at each explanation in turn.

Bargaining

How much **bargaining power** workers have depends on two factors. The first is how costly it would be for the firm to find other workers were they to leave the firm. The second is how hard it would be for them to find another job were they to leave the firm. The costlier it is for the firm to replace them, and the easier it is for them to find another job, the more bargaining power they have. This has two implications:

- How much bargaining power a worker has depends first on the nature of the job. Replacing a worker at McDonald's is not costly. The required skills can be taught quickly, and typically a large number of willing applicants have already filled out job application forms. In this situation, the worker is unlikely to have much bargaining power. If he or she asks for a higher wage, the firm can lay him or her off and find a replacement at minimum cost. In contrast, a highly skilled worker who knows in detail how the firm operates may be difficult and costly to replace. This gives him or her more bargaining power. If he or she asks for a higher wage, the firm may decide that it is best to provide it.
- How much bargaining power a worker has also depends on labor market conditions. When the unemployment rate is low, it is more difficult for firms to find acceptable replacement workers. At the same time, it is easier for workers to find other jobs. Under these conditions, workers are in a stronger bargaining position and may be able to obtain a higher wage. Conversely, when the unemployment rate is high, finding good replacement workers is easier for firms, whereas finding another job is harder for workers. Being in a weak bargaining position, workers may have no choice but to accept a lower wage.

Peter Diamond, Dale Mortensen, and Christopher Pissarides received the 2010 Nobel Prize in economics precisely for working out the characteristics of a labor market with large flows and wage bargaining. ▶

Efficiency Wages

Regardless of workers' bargaining power, firms may want to pay more than the reservation wage. They may want their workers to be productive, and a higher wage can help them achieve that goal. If, for example, it takes a while for workers to learn how to do a job correctly, firms will want their workers to stay for some time. But if workers are paid only their reservation wage, they will be indifferent between staying or leaving. In this

From Henry Ford to Jeff Bezos

FOCUS

In 1914, Henry Ford—the builder of the most popular car in the world at the time, the Model T—made a stunning announcement. His company would pay all qualified employees a minimum of $5.00 a day for an eight-hour day. This was a large salary increase for most employees, who had been earning an average of $2.30 for a nine-hour day. From the point of view of the Ford Company, this increase in pay was far from negligible; it represented about half of the company's profits at the time.

What Ford's motivations were is not entirely clear. Ford himself gave too many reasons for us to know which ones he actually believed. The reason was not that the company had a hard time finding workers at the previous wage. But the company clearly had a hard time retaining workers. There was a high turnover rate, as well as high dissatisfaction among workers.

Whatever the reasons behind Ford's decision, as Table 1 shows, the results of the wage increase were astounding.

The annual turnover rate (the ratio of separations to employment) plunged from a high of 370% in 1913 to a low of 16% in 1915. (An annual turnover rate of 370% means that on average 31% of the company's workers left each month, so that over the course of a year the ratio of separations to employment was $31\% \times 12 = 370\%$.) The layoff rate collapsed from 62% to nearly 0%. The average rate of absenteeism (not shown in the table), which ran at close to 10% in 1913, was down to 2.5% one year later. There is little question that higher wages were the main source of these changes.

Table 1 Annual Turnover and Layoff Rates (%) At Ford, 1913–1915

	1913	1914	1915
Turnover rate (%)	370	54	16
Layoff rate (%)	62	7	0.1

Did productivity at the Ford plant increase enough to offset the cost of increased wages? The answer to this question is less clear. Productivity was much higher in 1914 than in 1913. Estimates of the productivity increases range from 30% to 50%. Despite higher wages, profits were also higher in 1914 than in 1913. But how much of this increase in profits was the result of changes in workers' behavior and how much was because of the increasing success of Model T cars is harder to establish.

Although the effects support efficiency wage theories, it may be that the increase in wages to $5 a day was excessive, at least from the point of view of profit maximization. But Henry Ford probably had other objectives as well, from keeping the unions out—which he did—to generating publicity for himself and the company—which he also surely did.

Interestingly, we may be witnessing another similar experiment. In October 2018, Jeff Bezos announced that Amazon would increase the minimum wage paid to its 250,000 regular workers and 100,000 seasonal workers to $15 an hour (compared to a federal minimum wage of $7.50). One reason is certainly the need to attract and keep workers given the very low unemployment rate. Another is to counter some of the negative publicity Amazon has attracted. Another reason may be, just as it was for Ford, the desire to improve morale and productivity. It is too early to know what the effects will be.[1]

case, many of them will quit, and the turnover rate will be high. Paying a wage above the reservation wage makes it more attractive for workers to stay.

Behind this example lies a more general proposition. Most firms want their workers to feel good about their jobs. Feeling good promotes good work, which leads to higher productivity. Paying a high wage is one instrument the firm can use to achieve these goals. (See the Focus Box "From Henry Ford to Jeff Bezos.") Economists call the theories that link the productivity or the efficiency of workers to the wage they are paid **efficiency wage theories**.

Before September 11, 2001, the approach to airport security was to hire workers at low wages and accept the resulting high turnover. Now that airport security has become a higher priority, the approach has been to make the jobs more attractive and increase pay, so as to get more motivated and more competent workers and reduce turnover. The average salary of a TSA officer is $40,000, and turnover at the Transport Security Administration (TSA) is now roughly equal to the service industry average.

Like theories based on bargaining, efficiency wage theories suggest that wages depend on both the nature of the job and labor market conditions.

- Firms, such as high-tech firms, that see employee morale and commitment as essential to the quality of their work will pay more than firms in sectors where workers' activities are more routine.
- Labor market conditions affect wages. A low unemployment rate makes it more attractive for employed workers to quit: When unemployment is low, it is easy to find

[1]Sources: *Daniel Raff and Lawrence Summers, "Did Henry Ford Pay Efficiency Wages?"* Journal of Labor Economics, *1987, 5(4), Part 2: pp. S57–S87. Louise Matsakis, "Why Amazon Really Raised Its Minimum Wage to $15," Wired, October 2, 2018, www.wired.com/story/why-amazon-really-raised-minimum-wage.*

another job. A firm that wants to avoid an increase in quits when unemployment is low has to increase wages to induce workers to stay with the firm. When this happens, lower unemployment again leads to higher wages. Conversely, higher unemployment leads to lower wages.

Wages, Prices, and Unemployment

We can capture our discussion of wage determination by using the following equation:

$$W = P^e \underset{(-,+)}{F(u, z)} \tag{7.1}$$

The aggregate nominal wage W depends on three factors:

- The expected price level, P^e
- The unemployment rate, u
- A catch-all variable, z, that stands for all other variables that may affect the outcome of wage setting.

Let's look at each factor.

The Expected Price Level

First, ignore the difference between the expected and the actual price level and ask: Why does the price level affect nominal wages? The answer: Because both workers and firms care about *real* wages, not nominal wages.

- Workers do not care about how many dollars they receive but about how many goods they can buy with those dollars. In other words, they care about the nominal wages (W) they receive relative to the price of the goods they buy (P). They care about W/P.
- In the same way, firms do not care about the nominal wages they pay but about the nominal wages (W) they pay relative to the price of the goods they sell (P). So they also care about W/P.

Think of it another way. If workers expect the price level—the price of the goods they buy—to double, they will ask for a doubling of their nominal wage. If firms expect the price level—the price of the goods they sell—to double, they will be willing to double the nominal wage. So, if both workers and firms expect the price level to double, they will agree to double the nominal wage, keeping the real wage constant. This is captured in equation (7.1): A doubling in the expected price level leads to a doubling of the nominal wage chosen when wages are set.

An increase in the expected price level leads to an increase in the nominal wage in the same proportion. ►

Why do wages depend on the *expected price level*, P^e, rather than the *actual price level*, P? Because wages are set in nominal (dollar) terms, and when they are set, the relevant price level is not yet known.

For example, in some union contracts in the United States, nominal wages are set in advance for three years. Unions and firms have to decide what nominal wages will be over the following three years based on what they expect the price level to be over those three years. Even when wages are set by firms, or by bargaining between the firm and each worker, nominal wages are typically set for a year. If the price level goes up unexpectedly during the year, nominal wages are typically not readjusted. (How workers and firms form expectations of the price level will occupy us for much of the next two chapters; we will leave this issue aside for the moment.)

The Unemployment Rate

Also affecting the wage in equation (7.1) is the unemployment rate, u. The minus sign under u indicates that an increase in the unemployment rate *decreases* wages.

The fact that wages depend on the unemployment rate was one of the main conclusions of our previous discussion. If we think of wages as being determined by bargaining, then higher unemployment weakens workers' bargaining power, forcing them to accept lower wages. If we think of wages as being determined by efficiency wage considerations, then higher unemployment allows firms to pay lower wages and still keep workers willing to work.

An increase in unemployment leads to a decrease in the nominal wage.

The Other Factors

The third variable in equation (7.1), z, is a catch-all variable that stands for all the factors that affect wages given the expected price level and the unemployment rate. By convention, we will define z so that an increase in z implies an increase in the wage (thus, the positive sign under z in the equation). Our previous discussion suggests a long list of potential factors here.

By the definition of z, an increase in z leads to an increase in the nominal wage.

Take, for example, **unemployment insurance**, the payment of unemployment benefits to workers who lose their jobs. There are good reasons why society should provide some insurance to workers who lose their job and have a hard time finding another. But there is little question that, by making the prospects of unemployment less distressing, more generous unemployment benefits do increase wages at a given unemployment rate. If unemployment insurance did not exist, some workers would have little to live on and would be willing to accept low wages to avoid remaining unemployed. But unemployment insurance does exist, and it allows unemployed workers to hold out for higher wages. In this case, we can think of z as representing the level of unemployment benefits. At a given unemployment rate, higher unemployment benefits increase the wage.

It is easy to think of other factors. An increase in the minimum wage may also increase wages just above the minimum wage, leading to an increase in the average wage, W, at a given unemployment rate. Or take an increase in **employment protection**, which makes it more expensive for firms to lay off workers. Such a change is likely to increase the bargaining power of workers covered by this protection (laying them off and hiring other workers is now costlier for firms), increasing the wage for a given unemployment rate.

We will explore some of these factors as we go along.

7-4 PRICE DETERMINATION

Having looked at wage determination, let's now turn to price determination.

The prices set by firms depend on the costs they face. These costs depend, in turn, on the nature of the **production function**, which is the relation between the inputs used in production and the quantity of output produced, and on the prices of these inputs.

For the moment, we will assume that firms produce goods using labor as the only factor of production. We will write the production function as follows:

$$Y = AN$$

where Y is output, A is labor productivity, and N is employment. This way of writing the production function implies that **labor productivity**, which is defined as output per worker, is constant.

Using a term from microeconomics, this assumption implies *constant returns to labor in production*. If firms double the number of workers they employ, they double the amount of output they produce.

It should be clear that this is a strong simplification. In reality, firms use other factors of production in addition to labor. They use capital—machines and factories. They use raw materials—oil, for example. Moreover, there is technological progress, so that labor productivity (A) is not constant but steadily increases over time. We shall introduce these complications later. We shall introduce raw materials in Chapter 9 when we discuss changes in the price of oil. We shall focus on the role of capital and technological progress when we turn to the determination of output in the *long run* in Chapter 10 through Chapter 13. For the moment, though, this simple relation between output and employment will serve our purposes.

Given the assumption that labor productivity, A, is constant, we can make one further simplification. We can choose the units of output so that one worker produces one unit of output—in other words, so that $A = 1$. (This way we do not have to carry the letter A around, and this will simplify notation.) With this assumption, the production function becomes:

$$Y = N \tag{7.2}$$

The production function $Y = N$ implies that the cost of producing one more unit of output is the cost of employing one more worker, at wage W. Using the terminology introduced in your microeconomics course: The marginal cost of production—the cost of producing one more unit of output—is equal to W.

If there was perfect competition in the goods market, the price of a unit of output would be equal to marginal cost: P would be equal to W. But many goods markets are not competitive, and firms charge a price higher than their marginal cost. A simple way of capturing this fact is to assume that firms set their price according to

$$P = (1 + m)\,W \tag{7.3}$$

where m is the **markup** of the price over the cost. To the extent that goods markets are not competitive and firms have market power, m is positive, and the price, P, will exceed the cost, W, by a factor equal to $(1 + m)$.

7-5 THE NATURAL RATE OF UNEMPLOYMENT

Let's now look at the implications of wage and price determination for unemployment.

For the rest of this chapter, we shall do so under the assumption that nominal wages depend on the actual price level, P, rather than on the expected price level, P^e (why we make this assumption will become clear soon). Under this additional assumption, wage setting and price setting determine the equilibrium rate of unemployment, also called the **natural rate of unemployment**. Let's see how.

The rest of the chapter is based on the assumption that $P^e = P$. ►

The Wage-Setting Relation

Given the assumption that nominal wages depend on the actual price level (P) rather than on the expected price level (P^e), equation (7.1), which characterizes wage determination, becomes:

$$W = P\,F(u, z)$$

Dividing both sides by the price level,

$$\frac{W}{P} = \underset{(-,+)}{F(u, z)} \tag{7.4}$$

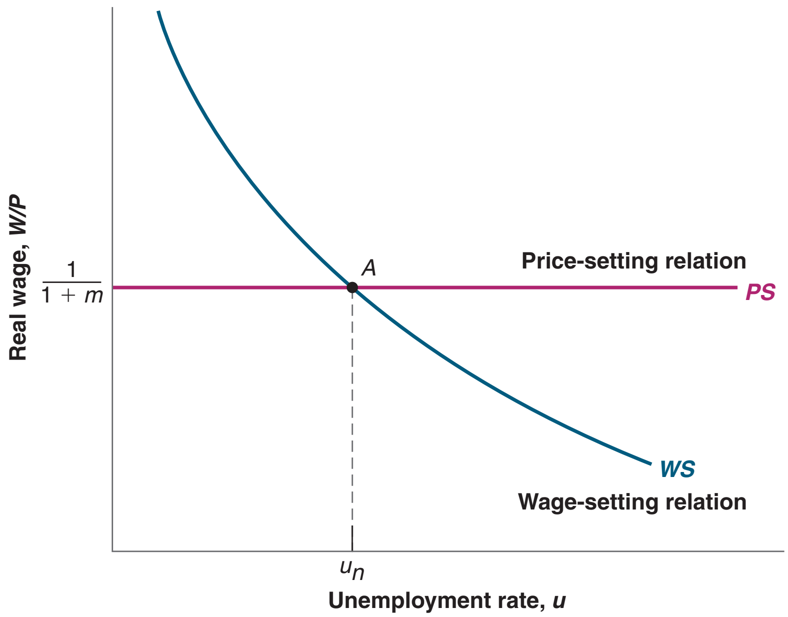


Figure 7-6

Wages, Prices, and the Natural Rate of Unemployment

The natural rate of unemployment is the unemployment rate such that the real wage chosen in wage setting is equal to the real wage implied by price setting.

Wage determination implies a negative relation between the real wage, W/P, and the unemployment rate, u: *The higher the unemployment rate, the lower the real wage chosen by wage setters*. The intuition is straightforward. The higher the unemployment rate, the weaker the workers' bargaining position, and the lower the real wage.

"Wage setters" are unions and firms if wages are set by collective bargaining; individual workers and firms if wages are set on a case-by-case basis; or firms if wages are set on a take-it-or-leave-it basis.

This relation between the real wage and the rate of unemployment—let's call it the **wage-setting relation**—is drawn in Figure 7-6. The real wage is measured on the vertical axis, the unemployment rate on the horizontal axis. The wage-setting relation is drawn as the downward-sloping curve WS. The higher the unemployment rate, the lower the real wage.

The Price-Setting Relation

Let's now look at the implications of price determination. If we divide both sides of the price determination equation, (7.3), by the nominal wage, we get

$$\frac{P}{W} = 1 + m \tag{7.5}$$

The ratio of the price level to the wage implied by the price-setting behavior of firms equals 1 plus the markup. Now invert both sides of this equation to get the implied real wage:

$$\frac{W}{P} = \frac{1}{1 + m} \tag{7.6}$$

Note what this equation says: *Price-setting decisions determine the real wage paid by firms.* An increase in the markup leads firms to increase their prices given the wage they have to pay; equivalently, it leads to a decrease in the real wage.

The step from equation (7.5) to equation (7.6) is algebraically straightforward. But how price setting actually determines the real wage paid by firms may not be intuitively obvious. Think of it this way: Suppose the firm you work for increases its markup and therefore increases the price of its product. Your real wage does not change much; you are still paid the same nominal wage, and the product produced by the firm is likely a small part of your consumption basket. Now suppose that not only the firm you work for but all the firms in the economy increase their markup. All the prices go up. Even if you are paid the same nominal wage, your real wage goes down. So, the higher the markup set by firms, the lower your (and everyone else's) real wage will be. This is what equation (7.6) says.

The **price-setting relation** in equation (7.6) is drawn as the horizontal line *PS* (for price setting) in Figure 7-6. The real wage implied by price setting is $1/(1 + m)$; it does not depend on the unemployment rate.

Equilibrium Real Wages and Unemployment

Equilibrium in the labor market requires that the real wage chosen in wage setting be equal to the real wage implied by price setting. (This way of stating equilibrium may sound strange if you learned to think in terms of labor supply and labor demand in your microeconomics course. The relation between wage setting and price setting, on the one hand, and labor supply and labor demand, on the other, is closer than it looks at first and is explored further in the appendix at the end of this chapter.) In Figure 7-6, equilibrium is therefore given by point *A*, and the equilibrium unemployment rate is given by u_n.

We can also characterize the equilibrium unemployment rate algebraically; eliminating W/P between equations (7.4) and (7.6) gives

$$F(u_n, z) = \frac{1}{1 + m} \tag{7.7}$$

The equilibrium unemployment rate, denoted u_n (*n* for natural), is such that the real wage chosen in wage setting—the left side of equation (7.7)—is equal to the real wage implied by price setting—the right side of equation (7.7).

The equilibrium unemployment rate u_n is called the *natural rate of unemployment* (which is why we have used the subscript *n* to denote it). The terminology has become standard, so we shall use it, but this is a really bad choice of words. The word *natural* suggests a constant of nature, one that is unaffected by institutions and policy. As its derivation makes clear, however, the natural rate of unemployment is anything but natural. The positions of the wage-setting and price-setting curves, and thus the equilibrium unemployment rate, depend on both *z* and *m*. Consider two examples:

Natural in *Webster's Dictionary* means "in a state provided by nature, without man-made changes."

- An increase in unemployment benefits. This can be represented by an increase in *z*. Because an increase in benefits makes the prospect of unemployment less painful, it increases the wage set by wage setters at a given unemployment rate. It shifts the wage-setting relation up, from *WS* to *WS'* in Figure 7-7. The economy moves along the *PS* line, from *A* to *A'*. The natural rate of unemployment increases from u_n to u'_n.
  In words: At a given unemployment rate, higher unemployment benefits lead to a higher real wage. A higher unemployment rate brings the real wage back to what firms are willing to pay.
- A less stringent enforcement of existing antitrust legislation. To the extent that this allows firms to collude more easily and increase their market power, it will lead to an increase in their markup *m*. This increase implies a decrease in the real wage paid by firms, and so it shifts the price-setting relation down, from *PS* to *PS'* in Figure 7-8. The economy moves along *WS*. The equilibrium moves from *A* to *A'*, and the natural rate of unemployment increases from u_n to u'_n.

  An increase in markups decreases the real wage and leads to an increase in the natural rate of unemployment. By letting firms increase their prices given the wage, less stringent enforcement of antitrust legislation leads to a decrease in the real wage. Higher unemployment is what makes workers accept this lower real wage, leading to an increase in the natural rate of unemployment.

An increase in unemployment benefits shifts the wage-setting curve up. The economy moves along the price-setting curve. Equilibrium unemployment increases. Does this mean that unemployment benefits are necessarily a bad idea? (*Hint*: No, but they have side effects.)

This has led some economists to call unemployment a "discipline device." It sounds harsh, but higher unemployment is indeed the economic mechanism through which wages get back in line with what firms are willing to pay.

An increase in the markup shifts the price-setting curve (line in this case). The economy moves along the wage-setting curve. Equilibrium unemployment increases.

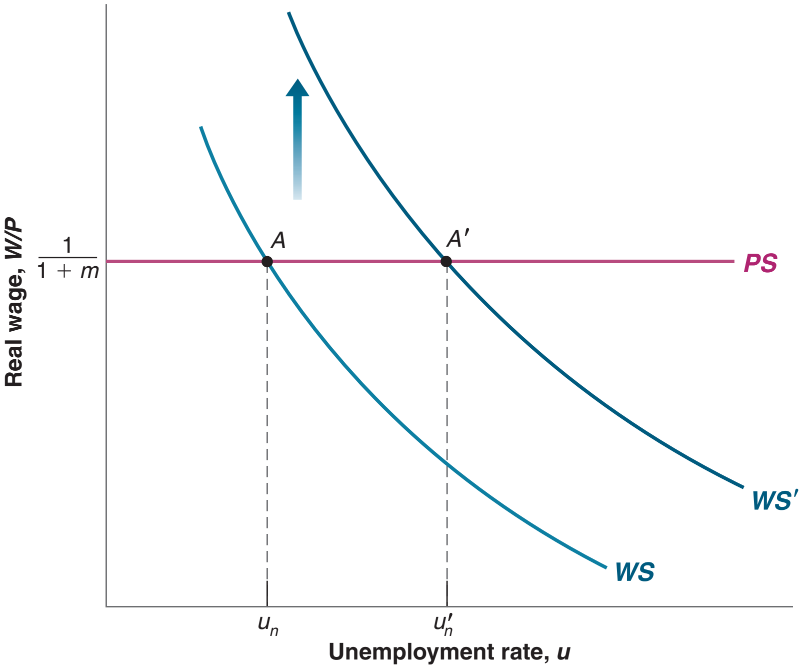


Figure 7-7

Unemployment Benefits and the Natural Rate of Unemployment

An increase in unemployment benefits leads to an increase in the natural rate of unemployment.

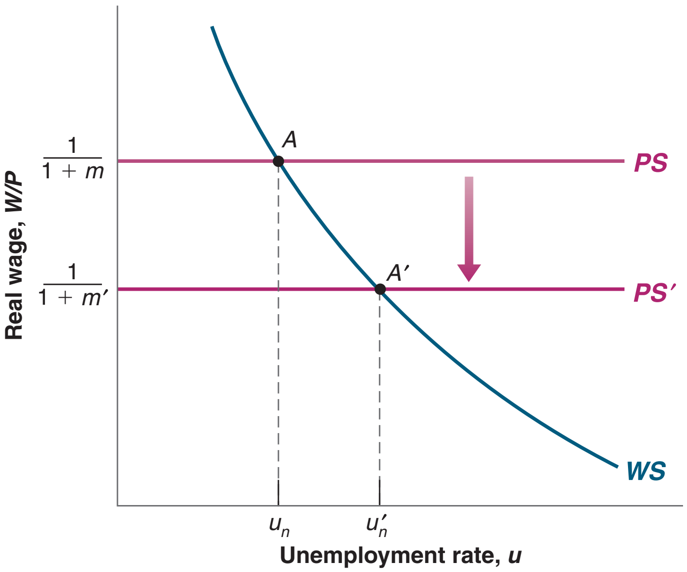


Figure 7-8

Markups and the Natural Rate of Unemployment

An increase in the markup leads to an increase in the natural rate of unemployment.

Factors like the generosity of unemployment benefits or antitrust legislation can hardly be thought of as the result of nature. Rather, they reflect various characteristics of the structure of the economy. For that reason, a better name for the equilibrium rate of unemployment would be the **structural rate of unemployment**, but so far the name has not caught on.

This name has been suggested by Edmund Phelps, of Columbia University. He was awarded the Nobel Prize in 2006. For more on some of his contributions, see Chapters 8 and 24.

7-6 WHERE WE GO FROM HERE

We have just seen how equilibrium in the labor market determines the equilibrium unemployment rate (called the *natural rate of unemployment*). Although we leave a precise derivation to Chapter 9, it is clear that, for a given labor force, the unemployment rate determines the level of employment, and that, given the production function, the level of employment determines the level of output. Thus, associated with the natural rate of unemployment is a *natural level of output*.

So, you may (and, indeed, you should) ask, What did we do in the previous four chapters? If equilibrium in the labor market determines the unemployment rate and, by implication, the level of output, why did we spend so much time looking at the goods and financial markets? What about our previous conclusions that the level of output was determined by factors such as monetary policy, fiscal policy, consumer confidence, and so on—all factors that do not enter equation (7.7) and therefore do not affect the natural level of output?

The key to the answer lies in the difference between the *short run* and the *medium run*:

- We have derived the natural rate of unemployment and, by implication, the associated level of output under two assumptions. First, we have assumed equilibrium in the labor market. Second, we have assumed that the price level was equal to the expected price level.
- However, there is no reason for the second assumption to be true in the *short run*. The price level may well turn out to be different from what was expected when nominal wages were set. Hence, in the short run, there is no reason for unemployment to be equal to the natural rate or for output to be equal to its natural level.

  As we shall see in Chapter 9, the factors that determine movements in output *in the short run* are indeed the factors we focused on in the preceding three chapters: monetary policy, fiscal policy, and so on. Your time (and mine) was not wasted.
- But expectations are unlikely to be systematically wrong (say, too high or too low) forever. That is why, in the medium run, output tends to return to its natural level. *In the medium run*, the factors that determine unemployment and output are the factors that appear in equation (7.7).

In the short run, the factors that determine movements in output are the factors we focused on in the preceding four chapters: monetary policy, fiscal policy, and so on.

In the medium run, output tends to return to the natural level. The factors that determine unemployment and, by implication, output are the factors we have focused on this chapter.

These, in short, are the answers to the questions asked in the first paragraph of this chapter. Developing these answers in detail will be our task in the next two chapters. Chapter 8 relaxes the assumption that the price level is equal to the expected price level and derives the relation between unemployment and inflation known as the Phillips curve. Chapter 9 puts all the pieces together.

SUMMARY

- The labor force consists of those who are working (employed) or looking for work (unemployed). The unemployment rate is equal to the ratio of the number of unemployed to the number in the labor force. The participation rate is equal to the ratio of the labor force to the working-age population.
- The US labor market is characterized by large flows between employment, unemployment, and "out of the labor force." On average, each month, about 44% of the unemployed move out of unemployment, either to take a job or to drop out of the labor force.
- Unemployment is high in recessions and low in expansions. During periods of high unemployment, the probability of losing a job increases and the probability of finding a job decreases.
- Wages are set unilaterally by firms or by bargaining between workers and firms. They depend negatively on the unemployment rate and positively on the expected price level. The reason why wages depend on the expected price level is that they are typically set in nominal terms for some period of time. During that time, even if the price level turns out to be different from what was expected, wages are typically not readjusted.
- The price set by firms depends on the wage and on the markup of prices over wages. A higher markup implies a higher price given the wage, and thus a lower real wage.
- Equilibrium in the labor market requires that the real wage chosen in wage setting be equal to the real wage implied by price setting. Under the additional assumption that the expected price level is equal to the actual price level, equilibrium in the labor market determines the unemployment rate. This unemployment rate is known as the *natural rate of unemployment*.

- In general, the actual price level may turn out to be different from the price level expected by wage setters. Therefore, the unemployment rate need not be equal to the natural rate.
- The coming chapters will show that, in the short run, unemployment and output are determined by the factors we focused on in the previous four chapters, but, in the medium run, unemployment tends to return to the natural rate and output tends to return to its natural level.

KEY TERMS

noninstitutional civilian population, 136
labor force, 136
out of the labor force, 136
participation rate, 136
unemployment rate, 136
separations, 136
hires, 136
Current Population Survey (CPS), 137
quits, 137
layoffs, 137
duration of unemployment, 137
discouraged workers, 138
employment rate, 138
collective bargaining, 141
reservation wage, 142
efficiency wage, 143
bargaining power, 142
efficiency wage theories, 143
unemployment insurance, 145
employment protection, 145
production function, 145
labor productivity, 145
markup, 146
natural rate of unemployment, 146
wage-setting relation, 147
price-setting relation, 148
structural rate of unemployment, 149

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. Since 1950, the participation rate in the United States has remained roughly constant at 60%.
b. Each month, the flows into and out of employment are very small compared to the size of the labor force.
c. Fewer than 10% of all unemployed workers exit the unemployment pool each year.
d. The unemployment rate tends to be high in recessions and low in expansions.
e. Most workers are typically paid their reservation wage.
f. Workers who do not belong to unions have no bargaining power.
g. It may be in the best interest of employers to pay wages higher than their workers' reservation wage.
h. The natural rate of unemployment is unaffected by policy changes.

2. Answer the following questions using the information provided in this chapter.

a. As a percentage of employed workers, what is the size of the flows into and out of employment (i.e., hires and separations) each month?
b. As a percentage of unemployed workers, what is the size of the flows from unemployment into employment each month?
c. As a percentage of the unemployed, what is the size of total flows out of unemployment each month? What is the average duration of unemployment?
d. As a percentage of the labor force, what is the size of the total flows into and out of the labor force each month?
e. In the text we say that there is an average of 450,000 new workers entering the labor force each month. What percentage of total flows into the labor force do new workers entering the labor force constitute?

3. The natural rate of unemployment

Suppose that the markup of goods prices over marginal cost is 5%, and that the wage-setting equation is

$$W = P(1 - u),$$

where u is the unemployment rate.

a. What is the real wage, as determined by the price-setting equation?
b. What is the natural rate of unemployment?
c. Suppose that the markup of prices over costs increases to 10%. What happens to the natural rate of unemployment? Explain the logic behind your answer.

DIG DEEPER

4. Reservation wages

In the mid-1980s, a famous supermodel once said that she would not get out of bed for less than $10,000 (presumably per day).

a. What is your own reservation wage?
b. Did your first job pay more than your reservation wage at the time?
c. Relative to your reservation wage at the time you accept each job, which job pays more: your first one or the one you expect to have in 10 years?

d. Explain your answers to parts a through c in terms of the efficiency wage theory.

e. Part of the policy response to the crisis was to extend the length of time workers could receive unemployment benefits. How would this affect reservation wages if this change were made permanent?

5. Bargaining power and wage determination

Even in the absence of collective bargaining, workers do have some bargaining power that allows them to receive wages higher than their reservation wage. Each worker's bargaining power depends both on the nature of the job and on the economywide labor market conditions. Let's consider each factor in turn.

a. Compare the job of a delivery person and a computer network administrator. In which of these jobs does a worker have more bargaining power? Why?

b. For any given job, how do labor market conditions affect a worker's bargaining power? Which labor market variable would you look at to assess labor-market conditions?

c. Suppose that for given labor market conditions (the variable you identified in part b), worker bargaining power throughout the economy increases. What effect would this have on the real wage in the medium run? in the short run? What determines the real wage in the model described in this chapter?

6. The existence of unemployment

a. Why does the wage-setting relation in Figure 1 have an upward slope? As N approaches L, what happens to the unemployment rate?

b. The price-setting relation is horizontal. How would an increase in the mark-up affect the position of the price-setting relation in Figure 1? How would an increase in the mark-up affect the natural rate of unemployment in Figure 1?

7. The informal labor market

You learned in Chapter 2 that informal work at home (e.g., preparing meals, taking care of children) is not counted as part of GDP. Such work also does not constitute employment in labor market statistics. With these observations in mind, consider two economies, each with 100 people, divided into 25 households each composed of four people. In each household, one person stays at home and prepares the food, two people work in the nonfood sector, and one person is unemployed. Assume that the workers outside food preparation produce the same actual and measured output in both economies.

In the first economy, EatIn, the 25 food preparation workers (one per household) cook for their families and do not work outside the home. All meals are prepared and eaten at home. The 25 food-preparation workers in this economy do not seek work in the formal labor market (and when asked, they say they are not looking for work). In the second economy, EatOut, the 25 food preparation workers are employed by restaurants. All meals are purchased in restaurants.

a. Calculate measured employment and unemployment and the measured labor force for each economy. Calculate the measured unemployment rate and participation rate for each economy. In which economy is measured GDP higher?

b. Suppose now that EatIn's economy changes. A few restaurants open, and the food preparation workers in 10 households take jobs in restaurants. The members of these 10 households now eat all of their meals at restaurants. The food preparation workers in the remaining 15 households continue to work at home and do not seek jobs in the formal sector. The members of these 15 households continue to eat all of their meals at home. Without calculating the numbers, what will happen to measured employment and unemployment and to the measured labor force, unemployment rate, and participation rate in EatIn? What will happen to measured GDP in EatIn?

c. Suppose that you want to include work at home in GDP and the employment statistics. How would you measure the value of work at home in GDP? How would you alter the definitions of *employment*, *unemployment*, and *out of the labor force*?

d. Given your new definitions in part c, would the labor market statistics differ for EatIn and EatOut? Assuming that the food produced by these economies has the same value, would measured GDP in these economies differ? Under your new definitions, would the experiment in part b have any effect on the labor market or GDP statistics for EatIn?

EXPLORE FURTHER

8. Unemployment durations and long-term unemployment

According to the data presented in this chapter, about 44% of unemployed workers leave unemployment each month.

a. Assume that the probability of leaving unemployment is the same for all unemployed persons, independent of how long they have been unemployed. What is the probability that an unemployed worker will still be unemployed after one month? two months? six months?

Now consider the composition of the unemployment pool. We will use a simple experiment to determine the proportion of the unemployed who have been unemployed six months or more. Suppose the number of unemployed workers is constant and equal to x. Each month, 47% of the unemployed find jobs, and an equivalent number of previously employed workers become unemployed.

b. Consider the group of x workers who are unemployed this month. After a month, what percentage of this group will still be unemployed? (*Hint*: If 47% of unemployed workers find jobs every month, what percentage of the original x unemployed workers did not find jobs in the first month?)

c. After a second month, what percentage of the original x unemployed workers has been unemployed for at least two months? (*Hint*: Given your answer to part b, what percentage of those unemployed for at least one month do not find jobs in the second month?) After the sixth month, what percentage of the original x unemployed workers has been unemployed for at least six months?

d. Using Table B-28 of the *Economic Report of the President* (this is the table number as of the 2019 report) you can compute the proportion of unemployed who have been unemployed six months or more (27 weeks or more) for each year between 2000 and 2019. How do the numbers between 2000 and 2008 (the precrisis years) compare with the answer you obtained in part c? Can you guess what may account for the difference between the actual numbers and the answer you obtained in this problem? (*Hint*: Suppose that the probability of exiting unemployment decreases the longer you are unemployed.)

e. For the unemployed who have been unemployed 6 months or more, what happens to them during the crisis years 2009 to 2011?

f. Is there any evidence of the crisis ending when you look at the percentage of the unemployed who have been unemployed 6 months or more?

g. Part of the policy response to the crisis was an extension of the length of time that an unemployed worker could receive unemployment benefits. How do you predict this change would affect the proportion of those unemployed more than six months? Did this occur?

9. *Go to the website maintained by the US Bureau of Labor Statistics (www.bls.gov). Find the latest Employment Situation Summary. Look under the link "National Employment."*

a. What are the latest monthly data on the size of the US civilian labor force, on the number of unemployed, and on the unemployment rate?

b. How many people are employed?

c. Compute the change in the number of unemployed from the first number in the table to the most recent month in the table. Do the same for the number of employed workers. Is the decline in unemployment equal to the increase in employment? Explain in words.

10. *The typical dynamics of unemployment over a recession*

The table below shows the behavior of annual real GDP growth during three recessions. These data are from Table B-4 of the Economic Report of the President.

Year	Real GDP Growth	Unemployment Rate
1981	2.5	
1982	–1.9	
1983	4.5	
1990	1.9	
1991	–0.2	
1992	3.4	
2008	0.0	
2009	–2.6	
2010	2.9	

Use Table B-35 from the Economic Report of the President to fill in the annual values of the unemployment rate in the table above and consider these questions.

a. When is the unemployment rate in a recession higher, during the year of declining output or the following year? Explain why.

b. Explain the pattern of the unemployment rate after a recession if discouraged workers return to the labor force as the economy recovers.

c. The rate of unemployment remains substantially higher after the crisis-induced recession in 2009. In that recession, unemployment benefits were extended in length from 6 months to 12 months. What does the model predict the effect of this policy will be on the natural rate of unemployment? Do the data support this prediction in any way?

11. *A closer look at changes in state labor markets*

There is a lot of discussion of the decline of the "Rust Belt" and the differences between labor markets at the state level. The table below is a snapshot of the labor market in California, Ohio, and Texas in 2003 before the Great Financial Crisis, in 2009 at the height of the crisis, and in 2018 after the crisis. Ohio is considered a Rust Belt state.

Source: www.bls.gov/lau/ex14tables.htma.

State	California	Ohio	Texas
Variable	Participation Rate		
2003	65.9	67.4	68.0
2009	65.1	66.0	60.8
2018	62.4	62.4	61.7
	Employment Rate		
2003	61.5	63.3	63.4
2009	57.8	59.2	60.8
2018	59.8	59.5	61.7
	Unemployment Rate		
2003	6.7	6.1	6.8
2009	11.3	11.8	7.5
2018	4.2	4.5	3.8

a. In which state did participation rates fall by the largest amount between 2003 and 2018? Is this consistent with the story of economic decline in the Rust Belt?

b. Using the increase in the unemployment rate from 2003 to 2009 to measure economic stress, in what state did the crisis hit hardest?

c. Using the decline in the participation rate from 2003 to 2009 to measure economic stress, in what state did the financial crisis hit hardest?

d. As of 2018, which state has the weakest labor market using all three statistics?

FURTHER READINGS

- A further discussion of unemployment along the lines of this chapter is given by Richard Layard, Stephen Nickell, and Richard Jackman in *The Unemployment Crisis* (1994, Oxford University Press).
- For more detail on how recessions affect various groups, read "Who Suffers During Recessions?" by Hilary Hoynes, Douglas Miller, and Jessamyn Schaller, *Journal of Economic Perspectives*, 2012, 26(3): pp. 27–48.

APPENDIX: Wage- and Price-Setting Relations versus Labor Supply and Labor Demand

If you have taken a microeconomics course, you probably saw a representation of labor market equilibrium in terms of labor supply and labor demand. You may therefore be asking yourself: How does the representation in terms of wage setting and price setting relate to the representation of the labor market I saw in that course?

In an important sense, the two representations are similar.

To see why, let's redraw Figure 7-6 in terms of the real wage on the vertical axis, and the level of *employment* (rather than the unemployment rate) on the horizontal axis. We do this in Figure 1.

Employment, N, is measured on the horizontal axis. The level of employment must be somewhere between zero and L, the labor force. Employment cannot exceed the number of people available for work (i.e., the labor force). For any employment level N, unemployment is given by $u = L - N$. Knowing this, we can measure unemployment by starting from L and *moving to the left* on the horizontal axis. Unemployment is given by the distance between L and N. The lower is employment, N, the higher is unemployment, and by implication the higher the unemployment rate, u.

Let's now draw the wage-setting and price-setting relations and characterize the equilibrium.

- An increase in employment (a move to the right along the horizontal axis) implies a decrease in unemployment and therefore an increase in the real wage chosen in wage setting. Thus, the wage-setting relation is now *upward sloping*. Higher employment implies a higher real wage.
- The price-setting relation is still a horizontal line at $W/P = 1/(1 + m)$.
- The equilibrium is given by point A, with "natural" employment level N_n (and an implied natural unemployment rate equal to $u_n = (L - N_n)/L$).

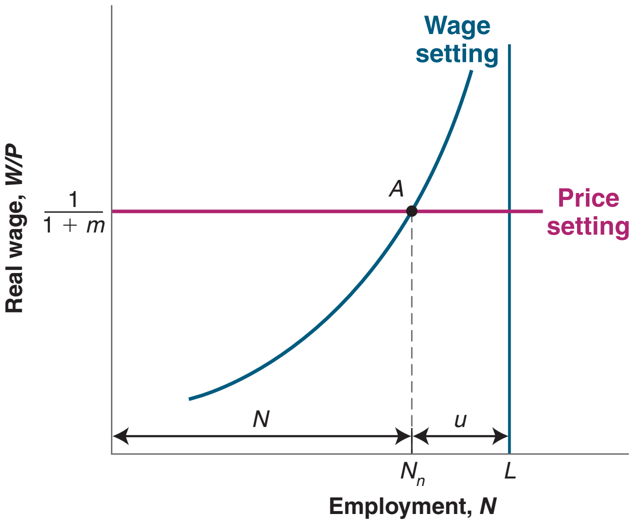


Figure 1

Wage and Price Setting and the Natural Level of Employment

In this figure the wage-setting relation looks like a labor supply relation. As the level of employment increases, the real wage paid to workers increases as well. For that reason, the wage-setting relation is sometimes called the "labor supply" relation (in quotes).

What we have called the *price-setting relation* looks like a flat labor demand relation. The reason it is flat rather than downward sloping has to do with our simplifying assumption of constant returns to labor in production. Had we assumed, more conventionally, that there were decreasing returns to labor in production, our price-setting curve would, like the standard labor demand curve, be downward sloping: As employment increased, the marginal cost of production would increase, forcing firms to increase their prices given the wages they pay. In other words, the real wage implied by price setting would decrease as employment increased.

But in a number of ways, the two approaches are different:

- The standard labor supply relation gives the wage at which a given number of workers are willing to work. The higher the wage, the larger the number of workers who are willing to work.

  In contrast, the wage corresponding to a given level of employment in the wage-setting relation is the result of a process of bargaining between workers and firms or unilateral wage setting by firms. Factors like the structure of collective bargaining or the use of wages to deter quits affect the wage-setting relation. In the real world, they seem to play an important role. Yet they play no role in the standard labor supply relation.
- The standard labor demand relation gives the level of employment chosen by firms at a given real wage. It is derived under the assumption that firms operate in competitive goods and labor markets and therefore take wages and prices—and by implication the real wage—as given.

  In contrast, the price-setting relation takes into account the fact that in most markets firms actually set prices. Factors such as the degree of competition in the goods market affect the price-setting relation by affecting the markup. But these factors aren't considered in the standard labor demand relation.
- In the labor supply–labor demand framework, those who are unemployed are *willingly unemployed*. At the equilibrium real wage, they prefer to be unemployed rather than work.

  In contrast, in the wage- and price-setting framework, unemployment is likely to be involuntary. For example, if firms pay an efficiency wage—a wage above the reservation wage—workers would rather be employed than unemployed. Yet, in equilibrium, there is still involuntary unemployment. This also seems to capture reality better than does the labor supply–labor demand framework.

  These are the three reasons why we have relied on the wage-setting and price-setting relations rather than on the labor supply–labor demand approach to characterize equilibrium in this chapter.

8

The Phillips Curve, the Natural Rate of Unemployment, and Inflation

In 1958, A. W. Phillips drew a diagram plotting the inflation rate against the unemployment rate in the United Kingdom for each year from 1861 to 1957. He found clear evidence of an inverse relation between inflation and unemployment. When unemployment was low, inflation was high, and when unemployment was high, inflation was low, often even negative.

A. W. Phillips was a New Zealander who taught at the London School of Economics. He had been, among other things, a crocodile hunter in his youth. He also built a hydraulic machine to describe the behavior of the macroeconomy. A working version of the machine is still on display in Cambridge, England.

Two years later, two US economists, Paul Samuelson and Robert Solow, replicated Phillips's exercise for the United States, using data from 1900 to 1960. Figure 8-1 reproduces their findings using consumer price index (CPI) inflation as a measure of the inflation rate. Apart from the period of high unemployment during the 1930s (the years from 1931 to 1939 are denoted by triangles and are well to the right of the other points in the figure), there was a clear negative relation between inflation and unemployment in the United States. This relation, which Samuelson and Solow labeled the **Phillips curve**, rapidly became central to macroeconomic thinking and policy. It appeared to imply that countries could choose between different combinations of unemployment and inflation. A country could achieve low unemployment if it were willing to tolerate higher inflation, or it could achieve price level stability—zero inflation—if it were willing to tolerate higher unemployment. Much of the discussion about macroeconomic policy became a discussion about which point to choose on the Phillips curve.

During the 1970s, however, this relation broke down. In the United States and most OECD countries, there was both high inflation *and* high unemployment, clearly contradicting the original Phillips curve. A relation reappeared, but as a relation between the *change* in the inflation rate and the unemployment rate. In the 1990s, the relation changed once again, and the old relation between inflation and unemployment reappeared. The purpose of this chapter is to explore these mutations of the Phillips curve and, more generally, to understand the relation between inflation and unemployment. We shall derive the Phillips curve from the model of the labor market we saw in Chapter 7. And you will see how the mutations of the Phillips curve have come from changes in the way people and firms have formed expectations.

The chapter has four sections:

Section 8-1 shows how the model of the labor market we saw previously implies a relation between inflation, expected inflation, and unemployment.

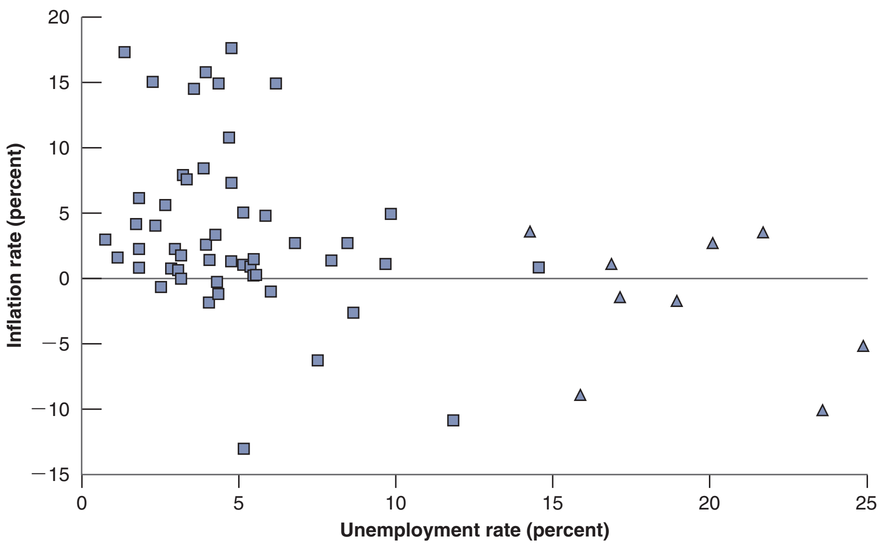
Section 8-2 uses this relation to interpret the mutations of the Phillips curve over time.

Figure 8-1

Inflation versus Unemployment in the United States, 1900–1960

During the period 1900–1960 in the United States, a low unemployment rate was typically associated with a high inflation rate, and a high unemployment rate was typically associated with a low or negative inflation rate.

Source: Based on Historical Statistics of the United States. http://hsus.cambridge.org/HSUSWeb/index.do.



Section 8-3 shows the relation between the Phillips curve and the natural rate of unemployment.

Section 8-4 further discusses the relation between unemployment and inflation across countries and over time.

If you remember a basic message from this chapter, it should be: Low unemployment puts upward pressure on inflation, but the form of the relation depends very much on how people and firms form expectations.

8-1 INFLATION, EXPECTED INFLATION, AND UNEMPLOYMENT

In Chapter 7, we derived the following equation for wage determination (equation (7.1)):

$$W = P^e F(u, z)$$

The nominal wage W, set by wage setters, depends on the expected price level, P^e, the unemployment rate, u, and a variable, z, that captures all the other factors that affect wage determination, from unemployment benefits to the form of collective bargaining.

It will be convenient to assume a specific form for the function, F:

$$F(u, z) = 1 - \alpha u + z$$

This captures the notion that the higher the unemployment rate, the lower the nominal wage; and the higher z (for example, the more generous unemployment benefits are), the higher the nominal wage. The parameter α (the Greek lowercase letter alpha) captures the strength of the effect of unemployment on the wage. Replacing the function, F, by this specific form in the equation above gives:

$$W = P^e(1 - \alpha u + z)$$

Also in Chapter 7, we derived the following equation for price determination (equation (7.3)):

$$P = (1 + m)\, W$$

The price, P, set by firms (equivalently, the price level) is equal to the nominal wage, W, times 1 plus the markup, m.

We then used these two relations together with the additional assumption that the actual price level was equal to the expected price level. Under this additional assumption, we then derived the natural rate of unemployment. We now explore what happens when we do not impose this additional assumption.

Replacing the nominal wage in the second equation by its expression from the first gives:

$$P = P^e(1 + m)(1 - \alpha u + z) \quad (8.1)$$

This gives us *a relation between the price level, the expected price level, and the unemployment rate.*

Let π denote the inflation rate and π^e the expected inflation rate. Then equation (8.1) can be rewritten as *a relation between inflation, expected inflation, and the unemployment rate:*

From now on, to lighten your reading, I shall often refer to the inflation rate simply as *inflation*, and to the unemployment rate simply as *unemployment*.

$$\pi = \pi^e + (m + z) - \alpha u \quad (8.2)$$

Deriving equation (8.2) from equation (8.1) is not difficult, but it is tedious, so it is left to an appendix at the end of this chapter. Equation (8.2) is one of the most important equations in macroeconomics. It is important that you understand each of the effects at work:

- *An increase in expected inflation, π^e, leads to an increase in actual inflation, π.*
  To see why, start from equation (8.1). An increase in the expected price level P^e leads, one for one, to an increase in the actual price level, P: If wage setters expect a higher price level, they set a higher nominal wage, which leads in turn to an increase in the price level.

  Now note that, given last period's price level, a higher price level this period implies a higher rate of increase in the price level from last period to this period—that is, higher inflation. Similarly, given last period's price level, a higher expected price level this period implies a higher expected rate of increase in the price level from last period to this period—that is, higher expected inflation. Thus the fact that an increase in the expected price level leads to an increase in the actual price level can be restated as: An increase in expected inflation leads to an increase in inflation.

  Increase in $\pi^e \Rightarrow$ Increase in π.

- *Given expected inflation, π^e, an increase in the markup, m, or an increase in the factors that affect wage determination—an increase in z—leads to an increase in actual inflation, π.*
  From equation (8.1): Given the expected price level, P^e, an increase in either m or z increases the price level, P. Using the same argument as in the previous bullet to restate this proposition in terms of inflation and expected inflation: Given expected inflation, π^e, an increase in either m or z leads to an increase in inflation π.

  Increase in m or $z \Rightarrow$ Increase in π.

- *Given expected inflation, π^e, a decrease in the unemployment rate, u, leads to an increase in actual inflation π.*

  From equation (8.1): Given the expected price level, P^e, a decrease in the unemployment rate, u, leads to a higher nominal wage, which leads to a higher price level, P. Restating this in terms of inflation and expected inflation: Given expected inflation, π^e, an increase in the unemployment rate, u, leads to an increase in inflation, π.

  Decrease in $u \Rightarrow$ Increase in π.

We need to take one more step before we return to a discussion of the Phillips curve. When we look at movements in inflation and unemployment in the rest of the chapter, it will be convenient to use time indexes so that we can refer to variables such as inflation, expected inflation, or unemployment in a specific year. So we rewrite equation (8.2) as:

$$\pi_t = \pi_t^e + (m + z) - \alpha u_t \quad (8.3)$$

The variables π_t, π_t^e, and u_t refer to inflation, expected inflation, and unemployment in year t. Note that there are no time indexes on m and z. This is because although m and z may move over time, they are likely to move slowly, especially relative to movement in inflation and unemployment. Thus, for the moment, we shall treat them as constant.

Equipped with equation (8.3), we can now return to the Phillips curve and its mutations.

8-2 THE PHILLIPS CURVE AND ITS MUTATIONS

Let's start with the relation between unemployment and inflation as it was first discovered by Phillips, Samuelson, and Solow.

The Original Phillips Curve

Assume that inflation varies from year to year around some value $\overline{\pi}$. Assume also that inflation is not persistent, so that inflation this year is not a good predictor of inflation next year. This happens to be a good characterization of the behavior of inflation over the period that Phillips or Solow and Samuelson were studying. In such an environment, it makes sense for wage setters in particular to expect that, whatever inflation was last year, inflation this year will simply be equal to $\overline{\pi}$. In this case, $\pi_t^e = \overline{\pi}$ and equation (8.3) becomes:

$$\pi_t = \overline{\pi} + (m + z) - \alpha u_t \tag{8.4}$$

In this case, we should observe a negative relation between unemployment and inflation. This is precisely the negative relation between unemployment and inflation that Phillips found for the United Kingdom and Solow and Samuelson found for the United States. When unemployment was high, inflation was low, even sometimes negative. When unemployment was low, inflation was positive.

When these findings were published, they suggested that policymakers faced a trade-off between inflation and unemployment. If they were willing to accept more inflation, they could achieve lower unemployment. This looked like an attractive trade-off, and starting in the early 1960s, US macroeconomic policy aimed at steadily decreasing unemployment. Figure 8-2 plots the combinations of the inflation rate and the unemployment rate in the United States for each year from 1961 to 1969. Note how well the negative relation between unemployment and inflation corresponding to equation (8.4) held during the long economic expansion that lasted throughout most of the 1960s. From 1961 to 1969, the unemployment rate declined steadily from 6.8% to 3.4%, and the inflation rate steadily increased, from 1.0% to 5.5%. Put informally, the US economy moved up along the Phillips curve. It indeed appeared that, if policymakers were willing to accept higher inflation, they could achieve lower unemployment.

The De-anchoring of Expectations

Around 1970, however, the relation between the inflation rate and the unemployment rate, so clear in Figure 8-2, broke down. Figure 8-3 shows the combination of the inflation rate and the unemployment rate in the United States for each year from 1970 to

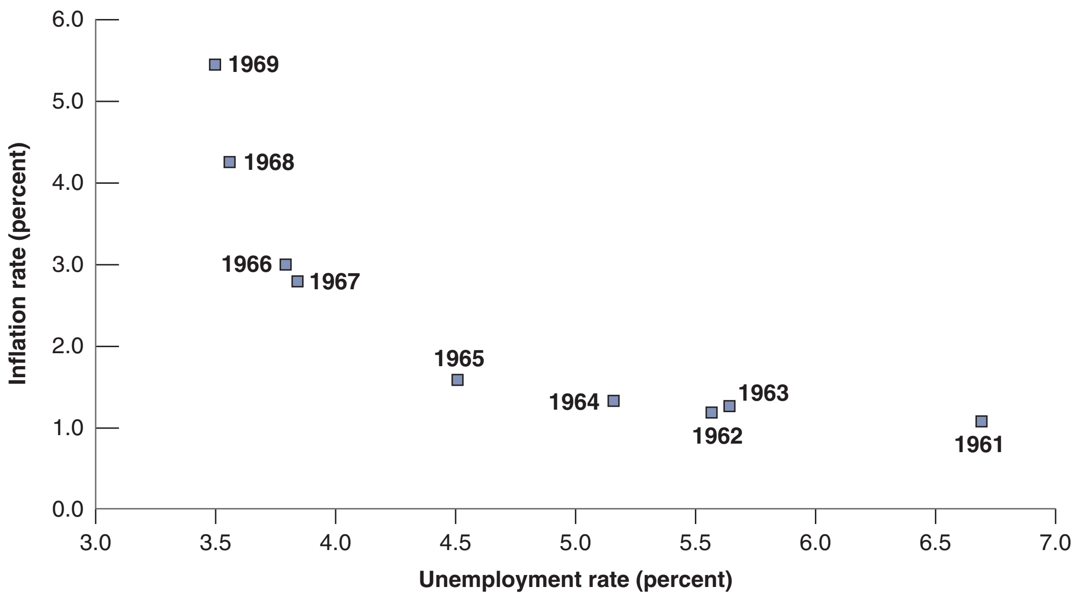


Figure 8-2

Inflation versus Unemployment in the United States, 1961–1969

The steady decline in the US unemployment rate throughout the 1960s was associated with a steady increase in the inflation rate.

Source: FRED: Series UNRATE, CPIAUSCL

1995. The points are scattered in a roughly symmetric cloud. There is no longer any visible relation between the unemployment rate and the inflation rate.

Why did the original Phillips curve vanish? *Because wage setters changed the way they formed their expectations about inflation.*

This change came from a change in the behavior of inflation. The rate of inflation became more persistent. High inflation in one year became more likely to be followed by high inflation the next year. As a result, people, when forming expectations, started to take into account the persistence of inflation, and this change in expectation formation

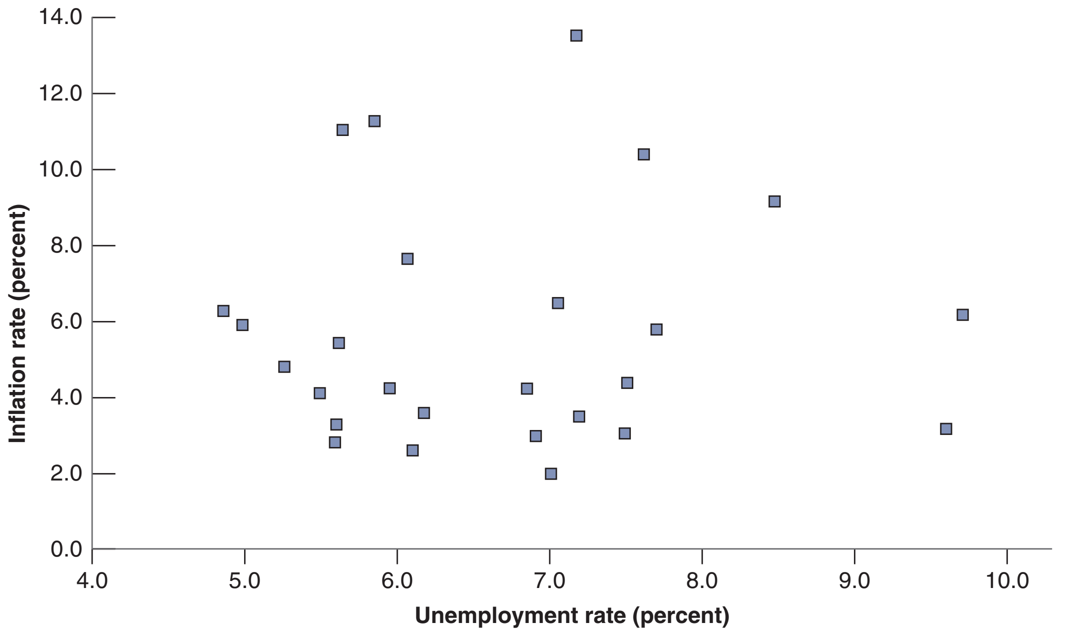


Figure 8-3

Inflation versus Unemployment in the United States, 1970–1995

Beginning in 1970 in the United States, the relation between the unemployment rate and the inflation rate disappeared.

Source: FRED: UNRATE, CPIAUSCL

changed the nature of the relation between unemployment and inflation. In macroeconomic jargon, expectations that had been **anchored** (i.e., roughly constant) became **de-anchored**.

Let's look at the argument in the previous paragraph more closely. Suppose expectations of inflation are formed according to:

$$\pi_t^e = (1 - \theta)\overline{\pi} + \theta\pi_{t-1} \tag{8.5}$$

In words: Expected inflation this year depends partly on a constant value, $\overline{\pi}$, with weight $1 - \theta$, and partly on inflation last year, which we denote by π_{t-1}, with weight θ. The higher the value of θ, the more last year's inflation leads workers and firms to revise their expectations of what inflation will be this year, and so the higher the expected inflation rate.

We can then think of what happened in the 1970s as an increase in the value of θ over time:

Go back to Figure 8-2 and look at the points corresponding to 1966 to 1969. If you had been a worker in 1969, what would you have assumed inflation was going to be in 1970? What nominal wage increase would you have asked for?

- So long as inflation was stable, it was reasonable for workers and firms to just ignore past inflation and to assume a constant value for inflation. For the period that Phillips and Samuelson and Solow had looked at, θ was close to zero and expectations were roughly given by $\pi^e = \overline{\pi}$. The Phillips curve was given by equation (8.4).
- But as inflation started to increase, workers and firms started changing the way they formed expectations. They started assuming that, if inflation had been high last year, it was likely to be high this year as well. The parameter θ, the effect of last year's inflation rate on this year's expected inflation rate, increased. The evidence suggests that, by the mid-1970s, people expected this year's inflation rate to be the same as last year's inflation rate—in other words, that θ was now equal to 1.

Now turn to the implications of different values of θ for the relation between inflation and unemployment. Substitute equation (8.5) for the value of π_t^e into equation (8.2):

$$\pi_t = \overbrace{(1 - \theta)\overline{\pi} + \theta\pi_{t-1}}^{\pi^e} + (m + z) - \alpha u_t$$

- When θ equals zero, we get the original Phillips curve, a relation between the inflation rate and the unemployment rate:

$$\pi_t = \overline{\pi} + (m + z) - \alpha u_t$$

- When θ is positive, the inflation rate depends not only on the unemployment rate but also on last year's inflation rate:

$$\pi_t = [(1 - \theta)\overline{\pi} + (m + z)] + \theta\pi_{t-1} - \alpha u_t$$

- When θ equals 1, the relation becomes (moving last year's inflation rate to the left side of the equation)

$$\pi_t - \pi_{t-1} = (m + z) - \alpha u_t \tag{8.6}$$

So, when $\theta = 1$, the unemployment rate affects not *the inflation rate* but rather the *change in the inflation rate*. High unemployment leads to *decreasing* inflation; low unemployment leads to *increasing* inflation.

This discussion is the key to what happened after 1970. As θ increased from 0 to 1, the simple relation between the unemployment rate and the inflation rate disappeared. This disappearance is what we saw in Figure 8-3.

But a new relation emerged, this time between the unemployment rate and the *change in the inflation rate*, as predicted by equation (8.6). This relation is shown in Figure 8-4,

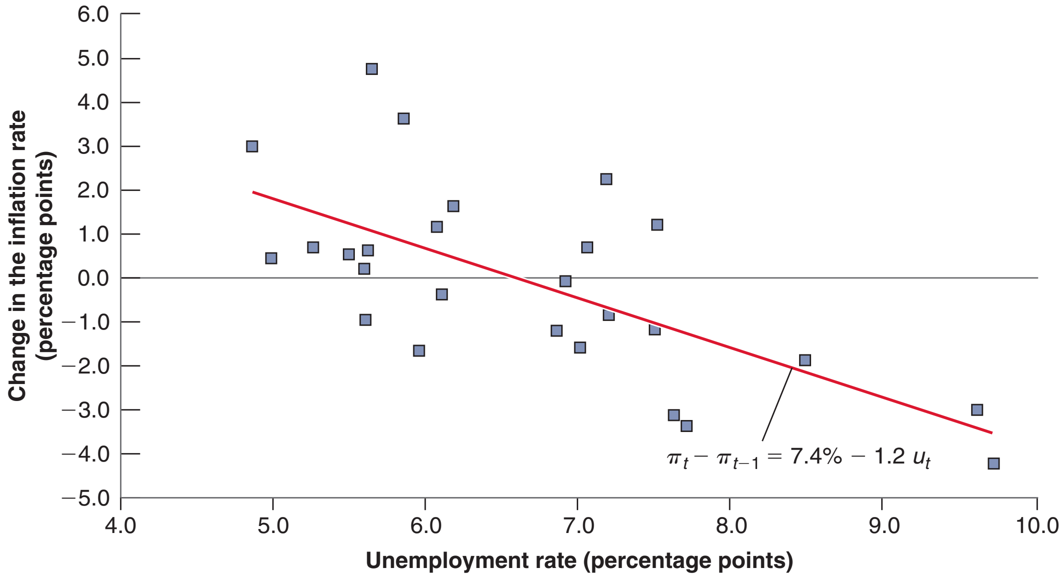


Figure 8-4

Change in Inflation versus Unemployment in the United States, 1970–1995

From 1970 to 1995, there was a negative relation between the unemployment rate and the change in the inflation rate in the United States.

Source: FRED: CPIAUCSL, UNRATE

which plots the *change* in the inflation rate versus the unemployment rate observed for each year from 1970 to 1995 and shows a clear negative relation between the unemployment and the change in the inflation rate. The figure also plots the relation that best fits the scatter of points for that period and is given by:

$$\pi_t - \pi_{t-1} = 7.4\% - 1.2\, u_t \qquad (8.7)$$

The lower the unemployment rate, the larger is the increase in the inflation rate. We shall return to this equation below.

So, instead of a relation between the inflation rate and the unemployment rate, the Phillips curve took the form of a relation between the change in the inflation rate and the unemployment rate. To distinguish it from the original Phillips curve, it became known as the **accelerationist Phillips curve** (to indicate that a low unemployment rate leads to an increase in the inflation rate and thus *an acceleration* of the price level).

Original Phillips curve: Low $u_t \Rightarrow$ High inflation.

Accelerationist Phillips curve: Low $u_t \Rightarrow$ Increasing inflation.

The Re-anchoring of Expectations

In the 1990s, the Phillips curve relation changed yet again, because of a change in monetary policy: From the early 1980s on, many central banks, including the Fed, increasingly emphasized their commitment to maintaining low and stable inflation. Many of them indicated that they would maintain inflation close to a given target, typically around 2%.

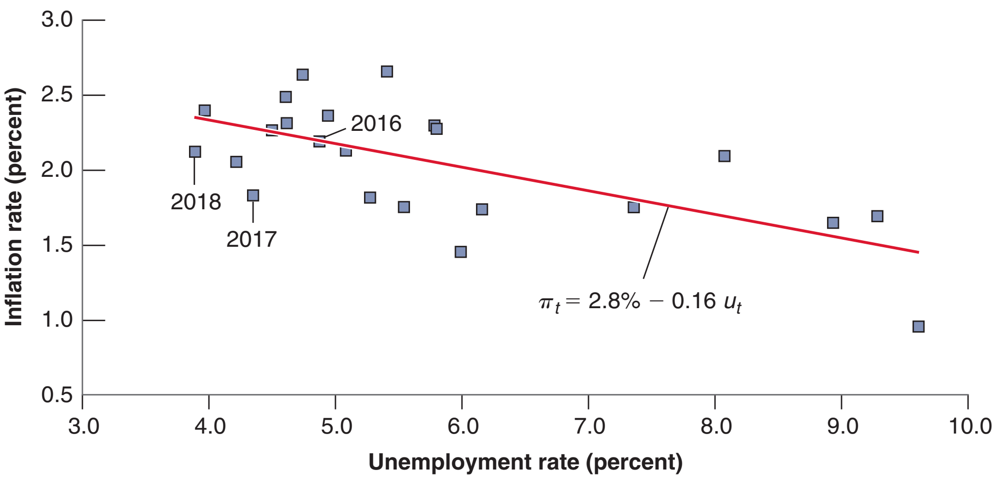
By the mid-1990s, the Fed had largely achieved its goal. Inflation had been stable for more than a decade, and so the way people formed expectations changed again. Even if inflation was higher than the target in a given year, people assumed that the central bank would take measures to return inflation to its target value in the future, and expected inflation became roughly constant, equal to the target inflation set by the central bank. Expectations of inflation that had become *de-anchored* during the 1970s and the 1980s became *re-anchored,* reacting little or not at all to movements in actual inflation.

Figure 8-5

Inflation versus Unemployment in the United States, 1996–2018

Since the mid-1990s, the Phillips curve has taken the form of a relation between the inflation rate and the unemployment rate.

Source: FRED: CPILFESL, UNRATE



In terms of equation (8.5), θ went back down to zero, and the Phillips curve returned to the relation between inflation and unemployment given by equation (8.4):

$$\pi_t = \overline{\pi} + (m + z) - \alpha u_t$$

You can see the negative relation between inflation and unemployment from 1996 to 2018 in Figure 8-5. Higher unemployment has been associated with lower inflation, lower unemployment associated with higher inflation. The figure also plots the line that best fits the scatter of points for the period 1996–2018 and is given by:

$$\pi_t = 2.8\% - 0.16\ u_t \qquad (8.8)$$

So are we back to the original Phillips curve? Yes, but having learned an important lesson.

The Phillips curve relation is a relation between inflation, expected inflation, and unemployment. What this implies for the relation between inflation and unemployment thus depends very much on how people form expectations. And this in turn depends very much on the behavior of inflation.

Today, because inflation has been stable for a long time, expectations of inflation are roughly constant and the Phillips curve again takes the form of a relation between the inflation rate and the unemployment rate. But we know that, were inflation to deviate again substantially from target, the way people would form expectations would change, and the relation between inflation and unemployment would change again, perhaps returning to the accelerationist Phillips curve of the 1970s and 1980s.

8-3 THE PHILLIPS CURVE AND THE NATURAL RATE OF UNEMPLOYMENT

The history of the Phillips curve is closely related to the discovery of the concept of the natural rate of unemployment that we introduced in Chapter 7.

The original Phillips curve was interpreted at the time as implying that there was no such thing as a natural unemployment rate. If policymakers were willing to tolerate a higher inflation rate, they could maintain a lower unemployment rate forever. And, indeed, throughout the 1960s, it looked as though they were right.

Theory ahead of Facts: Milton Friedman and Edmund Phelps

FOCUS

Economists are usually not good at predicting major changes before they happen, and most of their insights are derived *after* the fact. Here is an exception.

In the late 1960s—precisely as the original Phillips curve was working like a charm—two economists, Milton Friedman and Edmund Phelps, argued that the appearance of a trade-off between inflation and unemployment was an illusion.

Here are a few quotes from Milton Friedman about the Phillips curve:

> "Implicitly, Phillips wrote his article for a world in which everyone anticipated that nominal prices would be stable and in which this anticipation remained unshaken and immutable whatever happened to actual prices and wages. Suppose, by contrast, that everyone anticipates that prices will rise at a rate of more than 75% a year—as, for example, Brazilians did a few years ago. Then, wages must rise at that rate simply to keep real wages unchanged. An excess supply of labor [by this, Friedman means high unemployment] will be reflected in a less rapid rise in nominal wages than in anticipated prices, not in an absolute decline in wages."

He went on:

> "To state [my] conclusion differently, there is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off. The temporary trade-off comes not from inflation per se, but from a rising rate of inflation."

He then tried to guess how much longer the apparent trade-off between inflation and unemployment would last in the United States:

> "But how long, you will say, is 'temporary'?... I can at most venture a personal judgment, based on some examination of the historical evidence, that the initial effect of a higher and unanticipated rate of inflation lasts for something like two to five years; that this initial effect then begins to be reversed; and that a full adjustment to the new rate of inflation takes as long for employment as for interest rates, say, a couple of decades."

Friedman could not have been more right. A few years later, the original Phillips curve started to disappear, in exactly the way Friedman had predicted.[1]

In the late 1960s, however, although the original Phillips curve still gave a good description of the data, two economists, Milton Friedman and Edmund Phelps, questioned the existence of such a trade-off between unemployment and inflation. They questioned it on logical grounds, arguing that such a trade-off could exist only if wage setters systematically underpredicted inflation and that they were unlikely to make the same mistake forever. Friedman and Phelps also argued that if the government attempted to sustain lower unemployment by accepting higher inflation, the trade-off would ultimately disappear; the unemployment rate could not be sustained below a certain level that they called the *natural rate of unemployment*. Events proved them right, and, as we saw in Figure 8-3, the trade-off between the inflation rate and the unemployment rate indeed disappeared. (See the Focus Box "Theory ahead of Facts: Milton Friedman and Edmund Phelps.") Today, most economists accept the notion of a natural rate of unemployment—that is, subject to many caveats we shall see in the next section.

◀ Friedman was awarded the Nobel Prize in 1976. Phelps was awarded the Nobel Prize in 2006.

Let's make explicit the connection between the Phillips curve and the natural rate of unemployment.

By definition (see Chapter 7), the natural rate of unemployment is the unemployment rate at which the actual price level is equal to the expected price level. Equivalently, and more conveniently here, the natural rate of unemployment is the unemployment rate such that the actual inflation rate is equal to the expected inflation rate. Denote the natural unemployment rate by u_n (n stands for "natural"). Then, imposing the condition that actual inflation and expected inflation be the same ($\pi = \pi^e$) in equation (8.3) gives:

$$0 = (m + z) - \alpha u_n$$

[1]Source: *Milton Friedman, "The Role of Monetary Policy,"* American Economic Review, *1968, 58(1): pp. 1–17. The article by Edmund Phelps, "Money-Wage Dynamics and Labor-Market Equilibrium,"* Journal of Political Economy, *1968, 76(4–part 2): pp. 678–711, made many of the same points more formally.*

Note that under our assumption that m and z are constant, the natural rate is also constant, so we can drop the time index. We shall return to a discussion of what happens if m and z change over time.

Solving for the natural rate u_n,

$$u_n = \frac{m + z}{\alpha} \tag{8.9}$$

The higher the markup, m, or the higher the factors that affect wage setting, z, the higher is the natural rate of unemployment.

Now rewrite equation (8.3) as

$$\pi_t - \pi_t^e = -\alpha\left(u_t - \frac{m + z}{\alpha}\right)$$

Note from equation (8.9) that the fraction on the right side is equal to u_n, so we can rewrite the equation as

$$\pi_t - \pi_t^e = -\alpha(u_t - u_n) \tag{8.10}$$

This is an important equation, and one you must remember. It links the inflation rate, the expected inflation rate, and the deviation of the unemployment rate from the natural rate. It says that, if unemployment is at the natural rate, then inflation will be equal to expected inflation. If unemployment is below the natural rate, inflation will be higher than expected. If unemployment is instead above the natural rate, inflation will be lower than expected.

The equation also gives us a way of getting an estimate of the natural rate.

Take, for example, the period from 1970 to 1995, during which π^e was equal to last year's inflation rate, π_{t-1}, so the natural rate of unemployment was the rate at which $\pi_t - \pi_t^e = \pi_t - \pi_{t-1} = 0$. Using equation (8.7) and putting $\pi_t - \pi_{t-1} = 0$, the natural rate was thus given by:

$$0 = 7.4\% - 1.2u_n \Rightarrow u_n = 7.4\%/1.2 = 6.2\%$$

The natural rate was thus equal to roughly 6.2%. Why "roughly"? Because, as you can see from Figure 8-3, the fit of the regression is not tight, so we cannot be sure about the exact coefficients in the estimated equation. A more appropriate statement may be that the natural rate was probably between 6% and 7% during that period.

Turn to the Phillips curve relation that has prevailed since the mid-1990s. From survey evidence, expected inflation remained close to the target inflation of the Fed, 2%, so the natural rate of unemployment during this period has been the rate of unemployment at which inflation was equal to 2%. Using equation (8.8) and putting $\pi_t = 2\%$ gives:

$$2\% = 2.8\% - 0.16u_n \Rightarrow u_n = 0.8/0.16 = 5.0\%$$

Suppose, for example, that the true coefficient on the unemployment rate was 0.20 instead of 0.16. This would imply a natural unemployment rate of $0.8/0.2 = 4\%$.

This suggests that the natural unemployment rate during this period has been roughly 5.0%. Again, the "roughly" qualification is important. The fit of the regression line is not very good, and we cannot be sure of the exact values of the coefficients. By implication, we cannot be sure of the exact value of the natural rate. This uncertainty has important implications for monetary policy, a point to which we shall return later in the book.

8-4 A SUMMARY AND MANY WARNINGS

Let's take stock of what we have learned:

- The relation between inflation and unemployment depends on how wage setters form expectations of inflation.
- If expectations are anchored, as they were in the 1960s and have been again since the mid-1990s, then the Phillips curve takes the form of a relation between

inflation and unemployment. Inflation is higher than expected if unemployment is below the natural rate; it is lower than expected if unemployment is above the natural rate.

- If, however, expectations are unanchored—as they became in the 1970s and 1980s—and expected inflation this year is equal to inflation last year, the Phillips curve relation becomes a relation between the change in inflation and unemployment. Inflation increases if unemployment is below the natural rate, and it decreases if unemployment is above the natural rate.

In short, there is a relation between inflation and unemployment, but it is a complex one. And, unfortunately, further warnings are in order: The natural unemployment rate itself moves over time; it varies across countries; the relation between unemployment and inflation can completely disappear when inflation becomes very high; and it can disappear when inflation becomes low and turns into deflation. Let's take these issues in turn.

Variations in the Natural Rate over Time

In estimating equation (8.7) or equation (8.8), we implicitly treated $m + z$ as a constant. But there are good reasons to believe that m and z vary over time. The degree of monopoly power of firms, the costs of inputs other than labor, the structure of wage bargaining, the system of unemployment benefits, and so on are likely to change over time, leading to changes in either m or z and, by implication, changes in the natural rate of unemployment.

Indeed, we saw earlier that the estimate for the natural rate of unemployment for the period 1970–1995 was 6.2%, but the estimate for the period starting in 1995 was down to 5%. And, even during this more recent period, it is not clear that it has remained constant. As you can see from Figure 8-5, the unemployment rate has been lower than 5% since 2016, and yet inflation has remained close to 2%. As I am writing, there is an intense debate on whether the natural rate today is actually lower than 5%, perhaps as low as 3.5%. Why this might be is explored in the Focus Box "Changes in the US Natural Rate of Unemployment since 1990."

Will the natural rate of unemployment remain low in the future? Globalization, aging, prisons, temporary help agencies, the growth of platform companies like Uber, and the increasing role of the internet are probably all here to stay, suggesting that the natural rate will indeed remain low for the foreseeable future.

Variations in the Natural Unemployment Rate across Countries

Recall from equation (8.9) that the natural rate of unemployment depends on all the factors that affect wage setting, represented by the catchall variable, z; the markup set by firms, m; and the response of inflation to unemployment, represented by α. If these factors differ across countries, there is no reason to expect all countries to have the same natural rate of unemployment. And natural rates indeed differ across countries, sometimes considerably.

Take, for example, the unemployment rate in the euro area, where it has averaged close to 10% since 1990. A high unemployment rate for a few years may well reflect a deviation of the unemployment rate from the natural rate. A high average unemployment rate for 29 years almost surely reflects a high natural rate. This tells us where we should look for explanations, namely in the factors determining the wage-setting and the price-setting relations.

◀ Go back and look at Table 1-3 in Chapter 1.

Is it easy to identify the relevant factors? One often hears the statement that one of the main problems of Europe is its **labor market rigidities**. These rigidities, the

FOCUS

Changes in the US Natural Rate of Unemployment since 1990

As we discussed in the text, the natural rate of unemployment appears to have decreased in the United States from 6% or 7% in the 1970s and 1980s to perhaps less than 4% today. Researchers have offered a number of explanations.

- The fact that firms can more easily move some of their operations abroad makes them stronger when bargaining with their workers. Unions are becoming weaker. The unionization rate in the United States, which stood at 25% in the mid-1970s, is around 10% today. As we saw in Chapter 7, weaker bargaining power on the part of workers is likely to lead to a lower natural rate of unemployment.
- The nature of the labor market has changed. The number of workers under "alternative employment arrangements" (this includes on-call workers, temporary help agencies, independent contractors) has steadily increased, and now accounts for around 10% of the labor force. These workers are likely to have little bargaining power. The growth of alternative arrangements also allows many workers to look for jobs while being employed rather than unemployed. The increasing role of internet-based job sites has made the matching of jobs and workers easier. All these changes lead to lower unemployment.

  Some of the other explanations may surprise you. For example, researchers have also pointed to:
- The aging of the US population. The proportion of young workers (those between the ages of 16 and 24) has fallen from 24% in 1980 to 15% today. This reflects the end of the baby boom, which ended in the mid-1960s. Young workers tend to start their working life by going from job to job and typically have a higher unemployment rate. A decrease in the proportion of young workers leads to a decrease in the overall unemployment rate.
- An increase in the incarceration rate. The proportion of the US population in prison or in jail has tripled in the last 40 years. In 1980, 0.3% of the US population of working age was in prison; today the proportion is 0.9%. Because many of those in prison would likely have been unemployed were they not incarcerated, this is likely to have decreased the unemployment rate.
- The increase in the number of workers on disability. A relaxation of eligibility criteria since 1984 has led to a steady increase in the number of workers receiving disability insurance, from 2.2% of the working-age population in 1984 to 3.9% today. It is again likely that, absent changes in the rules, some of the workers now on disability insurance would have been unemployed instead.
- During the 2008–2009 crisis, there was the worry that the large increase in the actual unemployment rate (close to 10% in 2010) might eventually translate into an increase in the natural unemployment rate. The mechanism through which this might happen is known as **hysteresis** (in economics, hysteresis is used to mean that, "after a shock, a variable does not return to its initial value, even when the shock has gone away"). Workers who have been unemployed for a long time may lose their skills, or their morale, and become, in effect, unemployable, leading to a higher natural rate. This was a relevant concern: In 2010, the average duration of unemployment was 33 weeks, an exceptionally high number by historical standards. Of the unemployed, 43% had been unemployed for more than six months, and 28% for more than a year. When the economy picked up, how many of them would be scarred by their unemployment experience and hard to reemploy? Today, the verdict seems to be in. The return to a very low unemployment rate without high inflation and the recovery in labor force participation suggest that this worry was not justified, at least at the macroeconomic level.

argument goes, are responsible for its high unemployment. Although there is some truth to this statement, the reality is more complex. The Focus Box "What Explains European Unemployment?" discusses these issues further.

High Inflation and the Phillips Curve

Recall how, in the 1970s, the US Phillips curve changed as inflation became more persistent and wage setters changed the way they formed inflation expectations. The lesson is a general one. The relation between unemployment and inflation is likely to change with the level and persistence of inflation. Evidence from countries with high inflation confirms this. Not only does the way workers and firms form their expectations change, but so do institutional arrangements.

What Explains European Unemployment?

FOCUS

What do critics have in mind when they talk about the "labor market rigidities" afflicting Europe? They have in mind in particular:

- A generous system of unemployment insurance. The replacement rate—that is, the ratio of unemployment benefits to the after-tax wage—is often high in Europe, and the duration of benefits—the period during which the unemployed are entitled to receive benefits—often runs in years.

  Unemployment insurance is clearly desirable. But generous benefits are likely to increase unemployment in at least two ways. They decrease the incentives the unemployed have to search for jobs. They may also increase the wage that firms have to pay. Recall our discussion of efficiency wages in Chapter 7. The higher unemployment benefits are, the higher the wages firms must pay to motivate and keep workers.
- A high degree of employment protection. By employment protection, economists have in mind the set of rules that increase the cost of layoffs for firms. These include high severance payments, the need for firms to justify layoffs, and the possibility for workers to appeal the decision and have it reversed.

  The purpose of employment protection is to decrease layoffs, and thus to protect workers from the risk of unemployment. It indeed does that. What it also does, however, is make firms more reluctant to hire workers and thus make it harder for the unemployed to get jobs. The evidence suggests that, although employment protection does not necessarily increase unemployment, it changes its nature. The flows in and out of unemployment decrease, but the average duration of unemployment increases. Long durations increase the risk that the unemployed lose skills and morale, decreasing their employability.
- Minimum wages. Most European countries have national minimum wages. And in some countries, the ratio of the minimum wage to the median wage can be quite high. High minimum wages clearly risk limiting employment for the least-skilled workers, thus increasing their unemployment rate.
- Bargaining rules. In most European countries, labor contracts are subject to **extension agreements**. A contract agreed to by a subset of firms and unions can be automatically extended to all firms in the sector. This considerably reinforces the bargaining power of unions because it reduces the scope for competition by nonunionized firms. As we saw in Chapter 7, stronger bargaining power on the part of the unions may result in higher unemployment. Higher unemployment is the economic mechanism through which the higher demands of workers are reconciled with the wages paid by firms.

Do these labor market institutions really explain high unemployment in Europe? Is the case open and shut? Not quite. Here it is important to recall two important facts.

Fact 1: Unemployment was not always high in Europe. In the 1960s, the unemployment rate in the four major continental European countries (France, Germany, Italy, and Spain) was lower than that in the United States, around 2% to 3%. US economists would cross the ocean to study the "European unemployment miracle"! The natural rate in these countries today is around 8% to 9%. How do we explain this increase?

One hypothesis is that institutions were different then, and that labor market rigidities have appeared only in the last 40 years. This turns out not to be the case, however. It is true that, in response to the adverse shocks of the 1970s (in particular two recessions following large increases in the price of oil), many European governments increased the generosity of unemployment insurance and the degree of employment protection. But even in the 1960s, European labor market institutions looked nothing like US labor market institutions. Social protection was much higher in Europe; yet unemployment was lower.

A more convincing line of explanation focuses on the interaction between institutions and shocks. Some labor market institutions may be benign in some environments, yet costly in others. Take employment protection. If competition between firms is limited, the need to adjust employment in each firm may be limited as well, and so the cost of employment protection may be low. But if competition, either from other domestic firms or from foreign firms, increases, the cost of employment protection may become high. Firms that cannot adjust their labor force quickly may be unable to compete and go out of business.

Fact 2: There are large differences between natural rates across European countries. This is shown in Figure 1, which gives the unemployment rate for 15 European countries (the 15 members of the European Union before the increase in membership to 27) in 2006. I chose 2006 for two reasons: First, after 2006, the financial crisis triggered a large increase in unemployment above the natural rate; second, in most of these countries, inflation was stable and low in 2006, suggesting that the unemployment rate was roughly equal to the natural rate.

As you can see, the unemployment rate was indeed high in the four large continental countries: France, Spain, Germany, and Italy. (The German natural unemployment rate appears to have decreased a lot since then; in 2018, the actual unemployment rate was 3.5%, with little sign of inflation—another example of changes in the natural rate over time.)

But note also how low the unemployment rate was in some of the other countries, in particular Denmark, Ireland, and the Netherlands. Did these low unemployment countries have low benefits, low employment protection, and weak unions? Things are unfortunately not so simple. Spain has a generous safety net and has had a very high average unemployment rate. But so does the Netherlands, which has low average unemployment.

So, what is one to conclude? An emerging consensus among economists is that the devil is in the details. Generous social protection is consistent with low unemployment, but

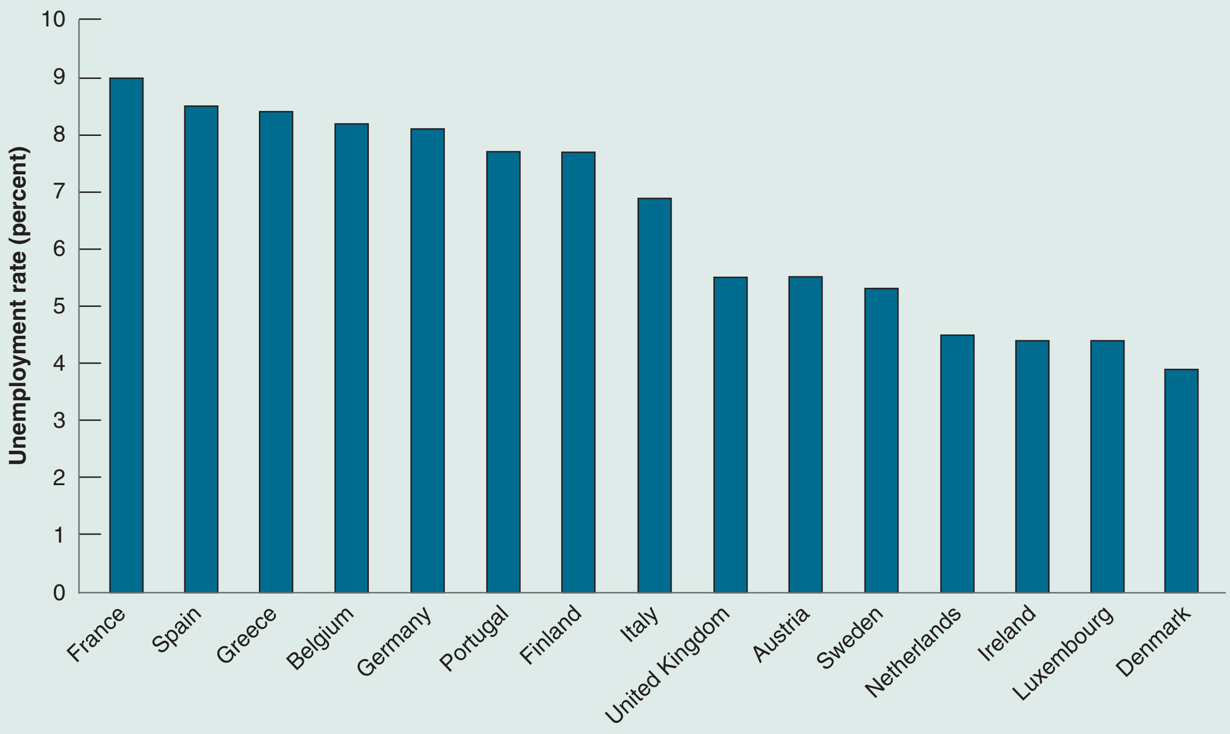


Figure 1

Unemployment Rates in 15 European Countries, 2006

Source: WEO database.

it must be provided efficiently. Unemployment benefits can be generous so long as the unemployed are induced to take jobs if jobs are available. Employment protection (e.g., in the form of generous severance payments) is consistent with low unemployment, so long as firms do not face the prospect of long administrative or judicial uncertainty when they lay off workers. Countries such as Denmark appear to have been more successful in achieving these goals. Creating incentives for the unemployed to take jobs and simplifying the rules of employment protection are on the reform agenda of many European governments. One may hope they will lead to a decrease in the natural rate in the future.

For more on European unemployment, read Olivier Blanchard, "European Unemployment: The Evolution of Facts and Ideas," ***Economic Policy*****, 2006, 21(45): pp. 1–54.**

More concretely, when inflation runs on average at 2% a year, wage setters can be reasonably confident that inflation will be between 1% and 3%. When inflation runs on average at 30% a year, wage setters can be confident that it will be between 20% and 40%. In the first case, the real wage may end up 1% higher or lower than expected when they set the nominal wage. In the second case, it may end up 10% higher or lower than they expected. There is much more uncertainty in the second case.

When the inflation rate becomes high, inflation tends to become more variable. As a result, workers and firms become more reluctant to enter into labor contracts that set nominal wages for a long period of time. If inflation turns out higher than expected, real wages may plunge, and workers will suffer a large cut in their living standard. If inflation turns out lower than expected, real wages may sharply increase, firms may not be able to pay their workers, and some firms may go bankrupt.

For this reason, the terms of wage agreements change with the level of inflation. Nominal wages are set for shorter periods of time, down from a year to a month or even less. **Wage indexation**, which is a provision that automatically increases wages in line with inflation, becomes more prevalent.

These changes lead in turn to a stronger response of inflation to unemployment. To see this, an example based on wage indexation will help. Imagine an economy that has two types of labor contracts. A proportion λ (the Greek lowercase letter lambda) of labor contracts is indexed. Nominal wages in those contracts move one-for-one with variations in the actual price level. A proportion $1 - \lambda$ of labor contracts is not indexed. Nominal wages are set on the basis of expected inflation.

Under this assumption, equation (8.10) becomes:

$$\pi_t = [\lambda\pi_t + (1 - \lambda)\pi_t^e] - \alpha(u_t - u_n)$$

The term in brackets on the right reflects the fact that a proportion λ of contracts is indexed and thus responds to actual inflation π_t, and a proportion, $1 - \lambda$, responds to expected inflation, π_t^e. If we assume that this year's expected inflation is equal to last year's actual inflation, $\pi_t^e = \pi_{t-1}$, we get:

$$\pi_t = [\lambda\pi_t + (1 - \lambda)\pi_{t-1}] - \alpha(u_t - u_n) \qquad (8.11)$$

When $\lambda = 0$, all wages are set on the basis of expected inflation—which is equal to last year's inflation, π_{t-1}—and the equation reduces to:

$$\pi_t - \pi_{t-1} = -\alpha(u_t - u_n)$$

When λ is positive, however, a proportion λ of wages is set on the basis of actual inflation rather than expected inflation. To see what this implies, reorganize equation (8.11). Move the term in brackets to the left, factor $(1 - \lambda)$ on the left of the equation, and divide both sides by $1 - \lambda$ to get:

$$\pi_t - \pi_{t-1} = -\frac{\alpha}{(1 - \lambda)}(u_t - u_n)$$

Wage indexation increases the effect of unemployment on inflation. The higher the proportion of wage contracts that are indexed—the higher λ—the larger the effect the unemployment rate has on the change in inflation—the higher the coefficient $\alpha/(1 - \lambda)$.

The intuition is as follows: Without wage indexation, lower unemployment increases wages, which in turn increase prices. But because wages do not respond to prices right away, there is no further increase in prices within the year. With wage indexation, however, an increase in prices leads to a further increase in wages within the year, which leads to a further increase in prices, and so on, so that the effect of unemployment on inflation within the year is higher.

If, and when, λ gets close to 1—which is when most labor contracts allow for wage indexation—small changes in unemployment can lead to large changes in inflation. Put another way, there can be large changes in inflation with nearly no change in unemployment. This is what happens in countries where inflation is high. The relation between inflation and unemployment becomes more and more tenuous and eventually disappears altogether.

For more on the dynamics of inflation when inflation is very high, see Chapter 22.

Deflation and the Phillips Curve

We have just looked at what happens to the Phillips curve when inflation is high. Another issue is what happens when inflation is low and possibly negative—when there is deflation.

One motivation for asking this question is given by an aspect of Figure 8-1 we mentioned at the start of the chapter but then left aside. In that figure, note that the points corresponding to the 1930s (denoted by triangles) lie to the right of the others. Not only was unemployment unusually high—this is no surprise because we are looking at the years corresponding to the Great Depression—but, *given the high unemployment rate*, the inflation rate was surprisingly high. In other words, given the high unemployment rate, we would have expected not merely deflation, but a large rate of deflation. In fact, deflation was limited, and from 1934 to 1937, despite high unemployment, inflation actually turned positive.

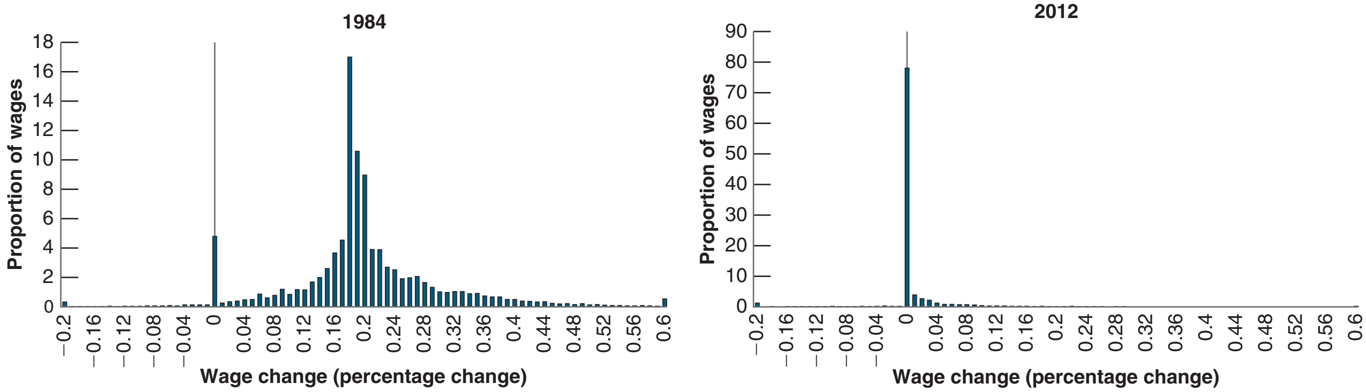


Figure 8-6

Distribution of Wage Changes in Portugal, in Times of High and Low Inflation

Source: Pedro Portugal, based on Portuguese household survey.

How do we interpret this fact? There are two potential explanations.

If u_n increases with u, then $u - u_n$ may remain small, and thus downward pressure on inflation may also remain small, even if u is high.

One is that the Great Depression was associated with an increase not only in the actual unemployment rate but also in the natural unemployment rate. This seems unlikely. Most economic historians see the Great Depression primarily as the result of a large adverse shift in demand (think of a large shift to the left of the IS curve) leading to an increase in the actual unemployment rate over the natural rate of unemployment, rather than an increase in the natural rate of unemployment itself.

Consider two scenarios. In one, inflation is 4% and your nominal wage goes up by 2%. In the other, inflation is 0% and your nominal wage is cut by 2%. Which do you dislike most? You should be indifferent between the two. In both cases, your real wage goes down by 2%. The evidence, however, is that many people find the first scenario less painful, and thus suffer from *money illusion*, a term made more explicit in Chapter 24.

The other is that, when the economy experiences deflation, the Phillips curve relation breaks down. One possible reason is the reluctance of workers to accept decreases in their nominal wages. Workers may unwittingly accept a cut in their real wages that occurs when their nominal wages increase more slowly than inflation. However, they are likely to fight the same cut in their real wages if it results from an overt cut in their nominal wages. This mechanism was clearly at work in some countries during the financial crisis.

Figure 8-6, for example, plots the distribution of wage changes in Portugal in two different years: 1984 when the inflation rate was a high 27%, and 2012, when it was just 2.1%. Note how the distribution of wage changes is roughly symmetric in 1984, but bunched at zero in 2012, with nearly no negative wage changes. To the extent that this mechanism is at work, this implies that the Phillips curve relation between inflation and unemployment may disappear, or at least become weaker, when the economy is close to zero inflation.

SUMMARY

- Labor market equilibrium delivers a relation between inflation, expected inflation, and unemployment. Given unemployment, higher expected inflation leads to higher inflation. Given expected inflation, higher unemployment leads to lower inflation.
- If expectations of inflation are anchored and expected inflation is roughly constant, the Phillips curve takes the form of a relation between the inflation rate and the unemployment rate. This is what Phillips in the United Kingdom and Solow and Samuelson in the United States discovered when they looked, in the late 1950s, at the joint behavior of unemployment and inflation.
- As inflation became higher and more persistent in the 1970s, expectations of inflation became de-anchored, increasingly reflecting past inflation. The Phillips curve took the form of a relation between the change in the inflation rate and the unemployment rate. High unemployment led to decreasing inflation; low unemployment led to increasing inflation.

- Starting in the 1980s, the Fed committed to keeping inflation low and stable. By the 1990s, expectations of inflation had become re-anchored, and expected inflation had become roughly constant. The Phillips curve became, again, a relation between the inflation rate and the unemployment rate.
- The natural unemployment rate is the unemployment rate at which the inflation rate is equal to expected inflation. If expectations are anchored and expected inflation is equal to the target rate of the central bank, this implies that, at the natural unemployment rate, actual inflation is equal to the target rate. When the actual unemployment rate exceeds the natural rate of unemployment, the inflation rate is lower than the target rate; when the actual unemployment rate is less than the natural unemployment rate, the inflation rate is higher than the target rate.
- The natural rate of unemployment is not a constant. It varies across countries, and is higher in Europe than in the United States. It also varies over time. In Europe, the natural rate has greatly increased since the 1960s. In the United States, it appears to have decreased since the early 1990s.
- Changes in the way the inflation rate varies over time affect the way wage setters form expectations and how much they use wage indexation. When wage indexation is widespread, small changes in unemployment can lead to large changes in inflation. At high rates of inflation, the relation between inflation and unemployment disappears altogether.
- At very low or negative rates of inflation, the relation between inflation and unemployment also becomes weaker. During the Great Depression even high unemployment led only to limited deflation. And this happened again in some European countries during the Great Recession.

KEY TERMS

Phillips curve, 155
original Phillips curve, 160
anchored (de-anchored, re-anchored, unanchored) expectations, 160
accelerationist Phillips curve, 161
hysteresis, 166
labor market rigidities, 165
extension agreements, 167
wage indexation, 168

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The original Phillips curve is the negative relation between unemployment and inflation that was first observed in the United Kingdom.
b. The original Phillips curve relation has proven to be very stable across countries and over time.
c. For some periods of history, inflation has been very persistent between adjacent years. In other periods of history, this year's inflation has been a poor predictor of next year's inflation.
d. Policy makers can only exploit the inflation–unemployment trade-off temporarily.
e. Expected inflation always equals actual inflation.
f. In the late 1960s, the economists Milton Friedman and Edmund Phelps said that policy makers could achieve as low a rate of unemployment as they wanted.
g. If people assume that inflation will be the same as last year's inflation, the Phillips curve relation will be a relation between the change in the inflation rate and the unemployment rate.
h. The natural rate of unemployment is constant over time within a country.
i. The natural rate of unemployment is the same in all countries.
j. Deflation means that the rate of inflation is negative.
k. When inflation expectations are anchored, the Phillips curve relation is a relation between the change in the inflation rate and the unemployment rate.
l. If inflation expectations are anchored, real wages will be constant since actual inflation will equal expected inflation.

2. Discuss the following statements.

a. The Phillips curve implies that when unemployment is high, inflation is low, and vice versa. Therefore, we may experience either high inflation or high unemployment, but we will never experience both together.
b. As long as we do not mind having high inflation, we can achieve as low a level of unemployment as we want. All we have to do is increase the demand for goods and services by using, for example, expansionary fiscal policy.
c. In periods of deflation, workers resist reductions in their nominal wages in spite of the fact prices are falling.

3. The natural rate of unemployment

a. The Phillips curve is $\pi_t = \pi_t^e + (m + z) - \alpha u_t$.
Rewrite this relation as a relation between the deviation of the unemployment rate from the natural rate, inflation, and expected inflation.
b. In the previous chapter, we derived the natural rate of unemployment. What condition on the price level and the expected price level was imposed in that derivation? How does it relate to the condition imposed in part a?

c. How does the natural rate of unemployment vary with the markup?
d. How does the natural rate of unemployment vary with the catchall term z?
e. Identify two important sources of variation in the natural rate of unemployment across countries and across time.

4. The formation of expected inflation

The text proposes the following model of expected inflation

$$\pi_t^e = (1 - \theta)\,\overline{\pi} + \theta\pi_{t-1}$$

a. Describe the process of the formation of expected inflation when $\theta = 0$
b. Describe the process of the formation of expected inflation when $\theta = 0$.
c. How do you form your own expectation of inflation—more like part a, or more like part b?

5. Mutations of the Phillips curve

Suppose that the Phillips curve is given by

$$\pi_t = \pi_t^e + 0.1 - 2u_t$$

and expected inflation is given by

$$\pi_t^e = (1 - \theta)\,\overline{\pi} + \theta\pi_{t-1}$$

and suppose that θ is initially equal to 0, and that $\overline{\pi}$ is given and does not change. It could be zero or any positive value. Suppose that the rate of unemployment is initially equal to the natural rate. In year t, the authorities decide to bring the unemployment rate down to 3% and hold it there forever.

a. Determine the rate of inflation in periods $t + 1$, $t + 2$, $t + 3$, $t + 4$, $t + 5$. How does $\overline{\pi}$ compare to π?
b. Do you believe the answer given in part a? Why or why not? (*Hint*: Think about how people are more likely to form expectations of inflation.)
Now suppose that in year $t + 6$, θ increases from 0 to 1. Suppose that the government is still determined to keep u at 3% forever.
c. Why might θ increase in this way?
d. What will the inflation rate be in years $t + 6$, $t + 7$, and $t + 8$?
e. What happens to inflation when $\theta = 1$ and unemployment is kept below the natural rate of unemployment?
f. What happens to inflation when $\theta = 1$ and unemployment is kept at the natural rate of unemployment?

DIG DEEPER

6. The macroeconomic effects of the indexation of wages

Suppose that the Phillips curve is given by

$$\pi_t - \pi_t^e = 0.1 - 2u_t$$

where

$$\pi_t^e = \pi_{t-1}$$

Suppose that inflation in year $t + 1$ is zero. In year t, the central bank decides to keep the unemployment rate at 4% forever.

a. Compute the rate of inflation for years t, $t + 1$, $t + 2$ and $t + 3$.

Now suppose that half the workers have indexed labor contracts.

b. What is the new equation for the Phillips curve?
c. Based on your answer to part b, recompute your answer to part a.
d. What is the effect of wage indexation on the relation between π and u?

7. Exploring the natural rate of unemployment

a. The equation of the Phillips curve from 1970 to 1995 is $\pi_t - \pi_{t-1} = 7.4\% - 1.2u_t$.
Calculate and define the natural rate of unemployment using this curve.
b. The equation of the Phillips curve from 1996 to 2018 is $\pi_t = 2.8\% - 0.16u_t$.
Here the natural rate of unemployment cannot immediately be calculated from this Phillips curve. Explain why.
c. Graph the Phillips relation $\pi_t = 2.8\% - 0.16u_t$ with inflation on the vertical axis and unemployment on the horizontal axis. Calculate and interpret the intercept on the vertical axis. Why might this be an undesirable economic outcome? Calculate and interpret the intercept on the horizontal axis. Why might this be an undesirable economic outcome? What is the unemployment rate if inflation is 2%?
d. What is the natural rate of unemployment using the relation $\pi_t = 2.8\% - 0.16u_t$ under the assumption that the value of $\overline{\pi} = 2.0\%$. Explain the logic of the calculation.
e. How has the natural rate of unemployment changed between 1970–1995 and 1996–2018?
f. What are the possible explanations for the change in the natural rate of unemployment between the 1980s and the 2000s?

8. Consider each table below. Is the data presented consistent with the Phillips curve model of wage determination? Each table has a point A and a point B. Start your answer with true/false/uncertain.

a. The natural rate of unemployment is 5%.

Point	Unemployment rate	Expected inflation (percent)	Percent increase in wages
A	6%	3%	3%
B	6%	2%	2%

b. The natural rate of unemployment is 5%.

Point	Unemployment rate	Expected inflation (percent)	Percent increase in wages
A	4%	2%	3%
B	3%	2%	2%

c. The natural rate of unemployment is 4%.

Point	Unemployment rate	Expected inflation (percent)	Percent increase in wages
A	4%	6%	7%
B	4%	2%	3%

d. The natural rate of unemployment is 5%.

Point	Unemployment rate	Expected inflation (percent)	Percent increase in wages
A	12%	2%	0%
B	12%	–2%	0%

EXPLORE FURTHER

9. Using the rate of unemployment to predict inflation between 1996 and 2018

The estimated Phillips curve from Figure 8.5 is

$$\pi_t = 2.8\% - 0.16\, u_t$$

Fill in the table below using the Phillips curve above after collecting the data from the FRED database. The monthly series are UNRATE, the unemployment rate and CPIAVCSL, the consumer price index. FRED allows you to download annual averages of these series. You will need to construct the inflation rate as the percentage change in the annual level of the CPI. You might choose to use a spreadsheet. Then answer the questions.

a. Assess the ability of the Phillips curve to predict inflation over the time period after 2006. What is the average prediction error?

b. Assess the ability of the Phillips curve to predict inflation during the crisis years 2009 and 2010? What do you think may be going on?

c. You will be able to add years after 2018 to your table. Assess the ability of the expectations augmented Phillips curve estimated with the data ending in 2018 to predict inflation after 2018.

Year	Inflation	Unemployment	Predicted inflation	Inflation minus predicted inflation
2006				
2007				
2008				
2009				
2010				
2011				
2012				
2013				
2014				
2015				
2016				
2017				
2018				
Future years				

10. The rate of inflation and expected inflation in different decades

Fill in the values in the table below for inflation and expected inflation using the 1960s. The data will come from FRED as they did in the last question. You will have the most success using a spreadsheet.

From the 1960s:

Year	π_t Actual inflation	π_{t-1} Lagged actual inflation	π_t^e Expected inflation under different assumptions		$\pi_t^e - \pi_t$ Difference: expected minus actual inflation under different assumptions	
Year			*Assume* $\theta = 0$ *and* $\overline{\pi} = 0$	*Assume* $\theta = 1.0$	*Assume* $\theta = 0$ *and* $\overline{\pi} = 0$	*Assume* $\theta = 1.0$
1963						
1964						
1965						
1966						
1967						
1968						
1969						

From the 1970s and 1980s:

Year	π_t Actual inflation	π_{t-1} Lagged actual inflation	π_t^e Expected inflation under different assumptions		$\pi_t^e - \pi_t$ Difference: expected minus actual inflation under different assumptions	
Year			*Assume* $\theta = 0$ *and* $\overline{\pi} = 0$	*Assume* $\theta = 1.0$	*Assume* $\theta = 0$ *and* $\overline{\pi} = 0$	*Assume* $\theta = 1.0$
1973						
1974						
1975						
1976						
1977						
1978						
1979						
1980						
1981						

From the 2010s (which you have partly done in Question 9):

Year	π_t Actual inflation	π_{t-1} Lagged actual inflation	π_t^e Expected inflation under different assumptions		$\pi_t^e - \pi_t$ Difference: expected minus actual inflation under different assumptions	
Year			*Assume* $\theta = 0$ *and* $\overline{\pi} = 2.0$	*Assume* $\theta = 1.0$	*Assume* $\theta = 0$ *and* $\overline{\pi} = 2.0$	*Assume* $\theta = 1.0$
2013						
2014						
2015						
2016						
2017						
2018						
Future years						

a. Is zero a good choice for the value of θ in the 1960s? Is $\overline{\pi} = 0$ a good choice for a value of $\overline{\pi}$? How are you making these judgments?

b. Is 1 a good choice for the value of θ in the 1960s? How are you making that judgment?

c. Is zero a good choice for the value of θ or $\overline{\pi}$ in the 1970s? How are you making that judgment?

d. Is 1 a good choice for the value of θ in the 1970s? How are you making that judgment?

e. How does the model where the anchored rate of inflation is 2% fit the data after 2012? How are you making that judgment?

f. Is 1 a good choice for the value of θ in the 2010s? How are you making that judgment?

g. How do you compare the behavior of inflation, its average level, and its persistence across these three time periods?

APPENDIX: Derivation of the Relation Between Inflation, Expected Inflation, and Unemployment

This appendix shows how to go from the relation between the price level, the expected price level, and the unemployment rate given by equation (8.1),

$$P = P^e(1 + m)(1 - \alpha u + z)$$

to the relation between inflation, expected inflation, and the unemployment rate given by equation (8.2),

$$\pi = \pi^e + (m + z) - \alpha u$$

First, introduce time subscripts for the price level, the expected price level, and the unemployment rate, so P_t, P_t^e, and u_t refer to the price level, the expected price level, and the unemployment rate in year t. Equation (8.1) becomes:

$$P_t = P_t^e(1 + m)(1 - \alpha u_t + z)$$

Next, go from an expression in terms of price levels to an expression in terms of inflation rates. Divide both sides by last year's price level, P_{t-1}:

$$\frac{P_t}{P_{t-1}} = \frac{P_t^e}{P_{t-1}}(1 + m)(1 - \alpha u_t + z) \qquad (8A.1)$$

Take the fraction P_t/P_{t-1} on the left side and rewrite it as:

$$\frac{P_t}{P_{t-1}} = \frac{P_t - P_{t-1} + P_{t-1}}{P_{t-1}} = 1 + \frac{P_t - P_{t-1}}{P_{t-1}} = 1 + \pi_t$$

where the first equality follows from actually subtracting and adding P_{t-1} in the numerator of the fraction, the second equality follows from the fact that $P_{t-1}/P_{t-1} = 1$, and the third follows from the definition of the inflation rate ($\pi_t \equiv (P_t - P_{t-1})/P_{t-1}$).

Do the same for the fraction P_t^e/P_{t-1} on the right side, using the definition of the expected inflation rate ($\pi_t^e \equiv (P_t^e - P_{t-1})/P_{t-1}$):

$$\frac{P_t^e}{P_{t-1}} = \frac{P_t^e - P_{t-1} + P_{t-1}}{P_{t-1}} = 1 + \frac{P_t^e - P_{t-1}}{P_{t-1}} = 1 + \pi_t^e$$

Replacing P_t/P_{t-1} and P_t^e/P_{t-1} in equation (8A.1) by the expressions we have just derived,

$$(1 + \pi_t) = (1 + \pi_t^e)(1 + m)(1 - \alpha u_t + z)$$

This gives us a relation between inflation, π_t, expected inflation, π_t^e, and the unemployment rate, u_t. The remaining steps make the relation look more friendly.

Divide both sides by $(1 + \pi_t^e)(1 + m)$:

$$\frac{(1 + \pi_t)}{(1 + \pi_t^e)(1 + m)} = 1 - \alpha u_t + z$$

So long as inflation, expected inflation, and the markup are not too large, a good approximation to the left side of this equation is given by $1 + \pi_t - \pi_t^e - m$ (see Propositions 3 and 6 in Appendix 2 at the end of the book). Replacing in the previous equation and rearranging gives:

$$\pi_t = \pi_t^e + (m + z) - \alpha u_t$$

Dropping the time indexes, this is equation (8.2) in the text. With the time indexes, this is equation (8.3) in the text.

The inflation rate, π_t, depends on the expected inflation rate π_t^e and the unemployment rate u_t. The relation also depends on the markup, m, on the factors that affect wage setting, z, and on the effect of the unemployment rate on wages, α.

This page is intentionally left blank

9

From the Short to the Medium Run: The IS-LM-PC Model

In Chapters 3 through 6 we looked at equilibrium in the goods and financial markets and saw how, in the short run, output is determined by demand. In Chapters 7 and 8, we looked at equilibrium in the labor market and derived how unemployment affects inflation. We now put the two parts together to characterize the behavior of output, unemployment, and inflation, in both the short run and the medium run. When confronted with a macroeconomic question about a particular shock or a particular policy, this model, which we shall call the IS-LM-PC (PC for Phillips curve), is typically the model I start from. I hope you find it as useful as I do.

The chapter is organized as follows.

Section 9-1 develops the IS-LM-PC model.

Section 9-2 shows how the economy adjusts from its short-run equilibrium to its medium-run equilibrium.

Section 9-3 discusses complications and how things can go wrong.

Section 9-4 looks at the dynamic effects of fiscal consolidation.

Section 9-5 looks at the dynamic effects of an increase in the price of oil.

Section 9-6 concludes the chapter.

If you remember one basic message from this chapter, it should be: In the short run, demand determines output. In the medium run, with the help of policy, output returns to potential. ▶▶▶

9-1 THE IS-LM-PC MODEL

In Chapter 6, we looked at the implications of equilibrium in the goods and financial markets.

From *equilibrium in the goods market*, we derived the following equation (equation 6.5) for the behavior of output in the short run:

$$Y = C(Y - T) + I(Y, r + x) + G \quad (9.1)$$

In the short run, output is determined by demand. Demand is the sum of consumption, investment, and government spending. Consumption depends on disposable income, which is equal to income net of taxes. Investment depends on output and on the real borrowing rate. The real borrowing rate is the sum of real rate, r, chosen by the central bank, and a risk premium, x. Government spending is exogenous.

As we did in Chapter 6, we can draw the IS curve implied by equation (9.1) between output, Y, and the real rate, r, taking as given taxes T, the risk premium x, and government spending G. This is done in the top half of Figure 9-1. The IS curve is downward sloping: The lower the real rate, r, the higher the equilibrium level of output. The mechanism behind the relation should be familiar by now: A lower real rate increases investment. Higher investment leads to higher demand. Higher demand leads to higher output. The increase in output further increases consumption and investment, leading to a further increase in demand, and so on.

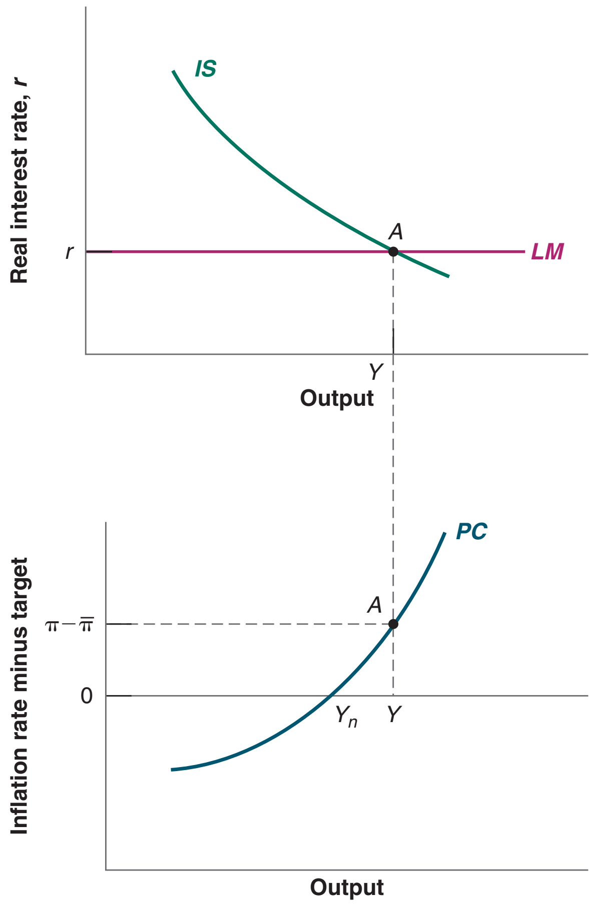
From goods and financial market equilibrium: A lower real interest rate increases demand and output.

From *equilibrium in financial markets*, we derived a flat LM curve, showing the real rate, r, chosen by the central bank. (We saw in Chapter 4 how, behind the scenes, the central bank adjusts the supply of money so as to achieve its desired interest rate.)

Figure 9-1

The IS-LM-PC Model: Output and Inflation

Top graph: A lower real rate leads to higher output. Bottom graph: Higher output leads to higher inflation.



It is drawn as the horizontal line at r in the top half of the figure. The intersection of the downward sloping IS curve and the horizontal LM curve gives us the equilibrium level of output in the short run.

From *equilibrium in labor markets* in Chapter 8, we derived the following equation (equation 8.10) for the relation between inflation and unemployment, a relation we called the *Phillips* curve (I am dropping the time indexes, which are not needed here):

$$\pi - \pi^e = -\alpha(u - u_n) \tag{9.2}$$

When the unemployment rate is lower than the natural rate, inflation is higher than expected. If the unemployment rate is higher than the natural rate, inflation is lower than expected.

Given that the first relation (equation (9.1)) is in terms of output, we must express this Phillips curve relation in terms of output and inflation rather than unemployment and inflation. It is easy, but it takes a few steps.

Start by looking at the relation between the unemployment rate and employment. By definition, the unemployment rate is equal to unemployment divided by the labor force:

◄ For a refresher, see Chapter 2.

$$u \equiv U/L = (L - N)/L = 1 - N/L$$

where N denotes employment and L denotes the labor force. The first equality is simply the definition of the unemployment rate. The second equality follows from the definition of unemployment, and the third equality is obtained through simplification. The unemployment rate is equal to one minus the ratio of employment to the labor force. Reorganizing to express N as a function of u gives:

$$N = L(1 - u)$$

Employment is equal to the labor force times one minus the unemployment rate.

Turning to output, we shall maintain for the moment the simplifying assumption we made in Chapter 7, that output is equal to employment, so:

$$Y = N = L(1 - u)$$

where the second equality follows from the previous equation.

Thus, when the unemployment rate is equal to the natural rate, u_n, employment is given by $N_n = L(1 - u_n)$ and output is equal to $Y_n = L(1 - u_n)$. Call N_n the natural level of employment (natural employment for short), and Y_n the natural level of output (natural output for short). Y_n *is* also called **potential output** and I shall often use that expression in what follows.

It follows that we can express the deviation of output from its natural level as:

$$Y - Y_n = L((1 - u) - (1 - u_n)) = -L(u - u_n)$$

This gives us a simple relation between the deviation of output from potential and the deviation of unemployment from its natural rate. The difference between output and potential output is called the **output gap**. If unemployment is equal to the natural rate, output is equal to potential, and the output gap is equal to zero; if unemployment is above the natural rate, output is below potential and the output gap is negative; and if unemployment is below the natural rate, output is above potential and the output gap is positive. (The relation of this equation to the actual relation between output and unemployment, known as Okun's law, is explored further in the Focus Box "Okun's Law across Time and Countries.")

Replacing $u - u_n$ in equation (9.2) gives:

$$\pi - \pi^e = (\alpha/L)(Y - Y_n) \tag{9.3}$$

FOCUS

Okun's Law across Time and Countries

How does the relation between output and unemployment that we derived in the text relate to the empirical relation between the two, known as Okun's law, which we saw in Chapter 2?

To answer this question, we must first rewrite the relation in the text in a way that makes the comparison easy between the two. Before giving you the derivation, which takes a few steps, let me give you the bottom line. The relation between unemployment and output derived in the text can be rewritten as shown in equation (9B.1) (the notation (–1) after a variable simply means last year's value of the variable):

$$u - u(-1) \approx -g_Y \qquad (9B.1)$$

The change in the unemployment rate is approximately equal (≈) to the negative of the growth rate (g) of output.

Here is the derivation. Start from the relation between employment, the labor force, and the unemployment rate $N = L(1 - u)$. Write the same relation for the year before, assuming a constant labor force L, so $N(-1) = L(1 - u(-1))$. Put the two relations together to get:

$$\begin{aligned} N - N(-1) &= L(1 - u) - L(1 - u(-1)) \\ &= -L(u - u(-1)) \end{aligned}$$

The change in employment is equal to minus the change in the unemployment rate, times the labor force. Divide both sides by $N(-1)$ to get

$$(N - N(-1))/N(-1) = -(L/N(-1))(u - u(-1))$$

Note that the expression on the left-hand side gives the rate of growth of employment, call it g_N. Given our assumption that output is proportional to employment, the rate of growth of output, call it g_Y, is equal to g_N. Note also that $L/N(-1)$ is a number close to 1. If the unemployment rate is equal to 5%, for example, then the ratio of the labor force to employment is 1.05. So, rounding it to 1, we can rewrite the expression as:

$$g_Y \approx -(u - u(-1)),$$

Reorganizing gives us the equation we want:

$$u - u(-1) \approx -g_Y \qquad (9B.1)$$

Now turn to the *actual relation* between the change in the unemployment rate and output growth, which is plotted in Figure 1 using annual data since 1960. (Figure 2-5 in Chapter 2 showed the same relation, but for the period starting in 2000 and using quarterly data.) The regression line that fits the points best is given by:

$$u - u(-1) = -0.4\,(g_Y - 3\%) \qquad (9B.2)$$

Like equation (9B.1), equation (9B.2) shows a negative relation between the change in unemployment and output growth. But it differs from equation (9B.1) in two ways.

- First, annual output growth must be at least 3% to prevent the unemployment rate from rising. This is because of two factors we ignored in our derivation: labor force growth and labor productivity growth. To maintain a constant unemployment rate, employment must grow at the same rate as the labor force. Suppose the labor force grows at 1.7% per year; then employment must grow at 1.7% per year. If, in addition, labor productivity (i.e., output per worker) grows at 1.3% per year, this implies that output must grow at 1.7% + 1.3% = 3% per year. In other words, to maintain a constant unemployment rate, output growth must be equal to the sum of labor force growth and labor productivity growth. In the United States, the sum of the rate of labor force growth and of labor productivity growth has been equal to 3% per year on average since 1960, and this explains why the number 3% appears on the right side of equation (9B.2). (There is some evidence, which we discussed in Chapter 1 and shall come back to in later chapters, that productivity growth has declined over time and that the growth rate needed to maintain a constant unemployment rate is now closer to 2%.)
- Second, the coefficient on the right side of equation (9B.2) is −0.4, compared to −1.0 in equation (9B.1). Put another way, output growth 1% above normal leads only to a 0.4% reduction in the unemployment rate in equation (9B.2) rather than the 1% reduction in equation (9B.1). There are two reasons why:

Firms adjust employment less than one-for-one in response to deviations of output growth from normal. More specifically, output growth 1% above normal for one year leads to only a 0.6% increase in the employment rate. One reason is that some workers are needed no matter what the level of output. The accounting department of a firm, for example, needs roughly the same number of employees whether the firm is selling more or less than normal. Another reason is that training new employees is costly, so firms prefer to keep current workers rather than lay them off when output is lower than normal and ask them to work overtime rather than hire new employees when output is higher than normal. In bad times, firms in effect hoard the labor they will need when times are better, a behavior called **labor hoarding**.

An increase in the employment rate does not lead to a one-for-one decrease in the unemployment rate. More specifically, a 0.6% increase in the employment rate leads to only a 0.4% decrease in the unemployment rate. The reason is that labor force participation increases. When employment increases, not all the new jobs are filled by the unemployed. Some of the jobs go to people who were classified as *out of the labor force*, meaning they were not actively looking for a job. Also, as labor market prospects improve for the unemployed, some discouraged workers, who were previously classified as out of the labor force, decide to start actively looking for a job and become classified as unemployed. For these reasons, unemployment decreases less than employment increases.

Putting the two steps together: Unemployment responds less than one-for-one to movements in employment, which itself responds less than one-for-one to movements in output. The coefficient giving the effect of output growth on the change in the unemployment rate, here 0.4, is called the **Okun coefficient**. Given the factors that determine this coefficient, one would expect the coefficient to differ across countries, and indeed it does. For example, in Japan, which has a tradition of lifetime employment, firms adjust employment much less in response to movements in output, leading to an Okun coefficient of only 0.1. Fluctuations in output are associated with much smaller fluctuations in unemployment in Japan than in the United States.

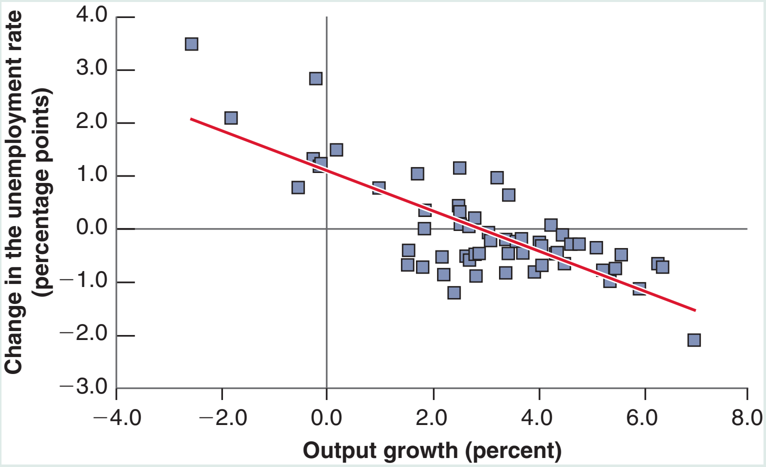
For more on Okun's law across countries and time, read "Okun's Law: Fit at 50?" by Laurence Ball, Daniel Leigh, and Prakash Loungani, working paper 606, The Johns Hopkins University, 2012.

Figure 1

Changes in the Unemployment Rate versus Output Growth in the United States, 1960–2018

High output growth is associated with a reduction in the unemployment rate; low output growth is associated with an increase in the unemployment rate.

Source: FRED: *Series* GDPCA, UNRATE



We need to take one last step. We saw in Chapter 8 that the way wage setters form expectations about inflation has changed through time. The evidence in Chapter 8 suggests that, in the United States today, inflation expectations are anchored and a reasonable assumption is that wage setters expect inflation to be equal to the target set by the Fed, denoted $\bar{\pi}$. This is an important assumption and I shall return to what happens when it does not hold at various points in the chapter. With this assumption, the relation between inflation and output is given by:

If you live in a country other than the United States, it may well be that this assumption is not the right one. You may want to explore the implications of other assumptions about expectations, for example, that wage setters expect inflation this year to be equal to last year's inflation.

$$\pi - \bar{\pi} = (\alpha/L)(Y - Y_n) \tag{9.4}$$

We have derived two relations in this section. The first, derived from goods and financial market equilibrium links the real rate to output. The second, derived from labor market equilibrium, links output to inflation. With these two equations, we can now describe what happens in the short and the medium run. This is what we do in the next section.

9-2 FROM THE SHORT TO THE MEDIUM RUN

Return to Figure 9-1. Assume that the real rate chosen by the central bank is equal to r. The top part of the figure tells us that, associated with this interest rate, the level of output is given by Y. The bottom part of the figure tells us that this level of output Y implies an inflation rate equal to π. Given the way I have drawn the figure, Y is larger than Y_n, so output is above potential. There is a positive output gap, which implies that inflation is higher than the target inflation, $\bar{\pi}$. Put less formally, the economy is overheating, putting pressure on inflation. This is the *short-run equilibrium.*

What happens over time? Suppose the central bank leaves the real rate unchanged at r. Then, output remains above potential and inflation remains above target. At some point, however, policy is likely to react to this higher inflation, for two reasons.

The first is that the mandate of the central bank is to keep inflation close to the target.

The second is that, if it does not react, what we saw happen in the 1970s and 1980s will happen again: Inflation expectations will de-anchor. Seeing that inflation continues to exceed the target, wage setters will change the way they form expectations. If, for example, they expect inflation to equal past inflation, then the Phillips curve relation will become a relation between the *change in inflation* and the output gap. If the output gap remains positive, inflation not only will be higher than the target but will start increasing. The central bank will have to act to stop it from increasing, and eventually return it to target. Better to react early before expectations de-anchor.

So, for both reasons, the central bank will react to the positive output gap by increasing the real rate so as to reduce inflation and, in so doing, reduce output. The process of adjustment is described in Figure 9-2. Let the initial equilibrium be denoted by point A in both the top and bottom graphs. The central bank increases the real rate over time, shifting the LM curve up, so the economy moves along the IS curve up from A to A'. Output decreases. Now turn to the bottom graph. As output decreases and the output gap shrinks, the economy moves down the PC curve from A to A'.

PC curve is redundant because the *C* stands for *curve*. But it will do.

At point A', the economy reaches its *medium-run equilibrium*. Let's look at it more closely.

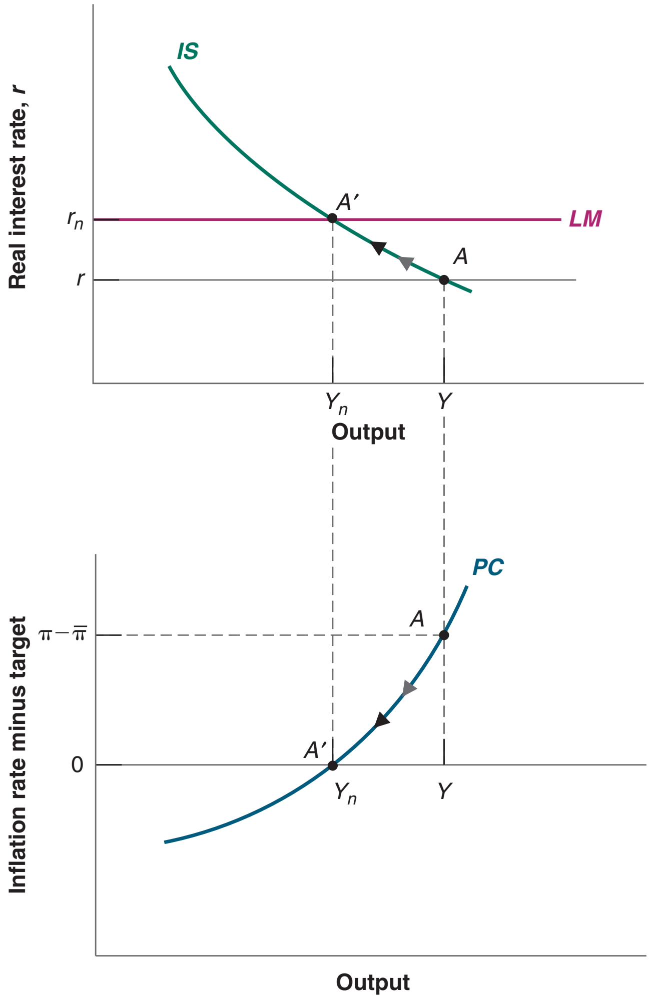
In the medium run, output returns to its natural level: $Y = Y_n$. In parallel, unemployment returns to the natural unemployment rate: $u = u_n$. With unemployment at the natural rate, the inflation rate returns to the target rate $\pi = \overline{\pi}$.

Turning to interest rates: The real rate must be such that the demand for goods is equal to potential output, and is thus given by $r = r_n$ in Figure 9-2. This interest rate is

Figure 9-2

Medium-Run Output and Inflation

Over the medium run, the economy converges to potential output and inflation converges to target inflation.



often called the **natural rate of interest** (to reflect the fact that it is the rate of interest associated with the natural rate of unemployment, or the natural level of output); it is also sometimes called the **neutral rate of interest** or the **Wicksellian rate of interest** (because the concept was introduced by Knut Wicksell, a Swedish economist who characterized it at the end of the 19th century). The real borrowing rate is given in turn by $r_n + x$, where x is the risk premium.

What about the nominal interest rate, i? Recall the relation between the nominal and the real rate from Chapter 6 (equation 6.4): The real rate is equal to the nominal rate minus expected inflation. Equivalently, the nominal rate is equal to the real rate plus expected inflation: $i = r + \pi^e$. Given that, in the medium run, the real rate is equal to the neutral rate, and that expected inflation is equal to actual inflation, which is itself equal to target inflation, this implies that the nominal rate is equal to $i = r_n + \overline{\pi}$. The higher the target inflation rate, the higher the nominal interest rate.

Finally, what about money and money growth? Recall the characterization of the equilibrium condition in Chapter 5 (equation 5.3) that the real supply of money equals real money demand, $M/P = YL(i)$. Given that output is equal to potential, and the nominal interest rate is determined from above, we can rewrite this equation as:

Recall that we are ignoring output growth for the time being.

$$\frac{M}{P} = Y_n L(r_n + \overline{\pi})$$

Note that all three variables on the right-hand side, natural output, natural rate of interest, and target inflation, are constant in steady state, so the real demand for money is constant in steady state. This implies that the left-hand side, the real supply of money, must be constant. This in turn implies that the price level, P, must be growing at the same rate as nominal money, M. Writing the rate of growth of money as g_M, this gives $\pi = g_M$. Replacing in the equation for the nominal interest rate, and using the fact that $\overline{\pi} = \pi$, we can rewrite the nominal rate as $i = r_n + g_M$. In the medium run, the nominal interest rate is equal to the real neutral rate plus the rate of nominal money growth.

A summary of the medium-run results: $Y = Y_n$, $u = u_n$, $\pi = \overline{\pi} = g_M$, $r = r_n$, $i = r_n + g_M$

To summarize: In the medium run, the real variables, be it output, unemployment, or the real rate of interest, are independent of monetary policy. What monetary policy determines is the rate of inflation and the nominal interest rate. In the medium run, a higher rate of money growth leads only to higher inflation and higher nominal interest rates. The fact that monetary policy does not affect real variables in the medium run is referred to as the **neutrality of money**.

9-3 COMPLICATIONS AND HOW THINGS CAN GO WRONG

The adjustment to the medium-run equilibrium in Section 9-2 seemed smooth and easy. Indeed, as you were reading, you may have asked yourself why, if the central bank wanted to return inflation to $\overline{\pi}$ and output to Y_n, didn't it simply increase the real rate from r to r_n right away, so that the medium-run equilibrium was reached without delay?

The answer is that central banks would indeed like to keep the economy at Y_n but, although it looks easy to do in Figure 9-2, reality is more complicated.

First, it is difficult for the central bank to know where potential output is exactly and thus how far output is from potential. The change in inflation provides a signal about the output gap (the distance between actual and potential output), but in contrast to the simple equation (9.4), the signal is noisy. The central bank may thus want to adjust the real rate slowly and see what happens. We saw in Chapter 8 that this is indeed one of the issues facing the Fed today. At 3.9%, the unemployment rate is below previous estimates of the natural rate. Yet, inflation, at 2.3%, is barely above 2%, the target inflation rate. Is the unemployment rate well below the natural rate, in which case the Fed should increase the interest rate in order to slow down the economy and allow the unemployment rate to increase? Or has the natural rate decreased, and there is no reason for the Fed to tighten? This is the main topic of discussion at the Fed and in the economic section of newspapers today.

Second, it takes time for the economy to respond, as discussed in Chapter 3 in the context of fiscal policy. Adjustments do not take place instantaneously. Firms take time to adjust their investment decisions. As investment spending slows down in response to the higher real rate, leading to lower demand, lower output, and lower income, it takes time for consumers to adjust to the decrease in income and for firms to adjust to the decrease in sales. In short, even if the central bank acts quickly, it takes time for the economy to go back to potential output.

Note the difficult problem facing the central bank in Figure 9-2. Output and inflation are too high. It may be difficult or even counterproductive to decrease demand and output too quickly. But going too slowly involves another risk, namely that, if inflation remains above target for too long, inflation expectations will de-anchor, leading to increasing inflation and requiring a costlier adjustment later on by the central bank. The next subsection explores a related and very relevant danger, which arises when the central bank is facing the zero lower bound.

The Zero Lower Bound and Deflation Spirals

In Figure 9-2, we considered the case where output was above potential and inflation was above target. Consider instead the case, represented in Figure 9-3, where the economy is in a recession. Given the real rate r and the position of the IS curve, the equilibrium is at point A. Output is equal to Y, which is lower than Y_n, so there is a negative output gap, implying that inflation is lower than target inflation. If the output gap is very large and so inflation is much lower than target inflation, this may imply that inflation is negative, that the economy is experiencing deflation.

For example, if the output gap is such that inflation is 3% below target ($\pi - \overline{\pi} = -3\%$), and the target inflation rate $\overline{\pi}$ is 2%, this implies that inflation is equal to $2\% - 3\% = -1\%$.

What the central bank should do in this case appears straightforward. It should decrease the real rate until output has increased back to its natural level. In terms of Figure 9-3, it should decrease the real rate from r to r_n. At r_n, output is equal to Y_n and inflation is equal to the target. Note that, if the economy is sufficiently depressed, the real rate, r_n, needed to return output to its natural level may be very low, perhaps even negative. I have drawn it so it is negative in the figure.

Recall that a negative real rate does not necessarily imply that people and firms, who borrow at a real rate equal to $r + x$, also face a negative real rate. If x is sufficiently large, the real rate at which they can borrow is positive even if the real rate is negative.

The zero lower bound may, however, make it impossible to achieve this negative real rate because the lowest nominal rate monetary policy can achieve is a rate of 0%. If there is deflation, however, then this implies that the lowest real rate monetary policy can achieve is actually positive, and equal to the deflation rate.

Suppose, for example, that the nominal rate is equal to zero and inflation is equal to –2%. Then the real rate $r = i - \pi = 0\% - (-2\%) = 2\%$.

This means that the central bank may not be able to decrease the real rate enough to return output to potential. To avoid cluttering the figure, assume that the central bank simply cannot decrease the real rate below r. (You could assume instead that it can decrease the real rate somewhat, but not all the way to r_n.) What happens then?

The first answer is that the economy remains at point A, with both a large output gap and deflation, not a good outcome. But the right answer is that things are likely to go from bad to worse. As people see that inflation is below target and the economy is experiencing deflation, they start changing the way they form expectations and start anticipating deflation. Expectations become de-anchored, and the negative output gap now leads not just to deflation but to larger and larger deflation. Not only that, but as deflation becomes larger, the real interest rate, which is equal to the deflation rate, increases. The larger the deflation, the higher the real interest rate, the lower output, and so on, a situation known as a **deflation spiral** or a **deflation trap**. As indicated by the two arrows in the figure,

Suppose, for example, that the nominal rate is equal to zero. With deflation at 2%, the real rate is equal to 2%. As deflation increases, say from 2% to 3%, the real rate increases from 2% to 3%, further decreasing spending and output, which leads to even larger deflation, and so on.

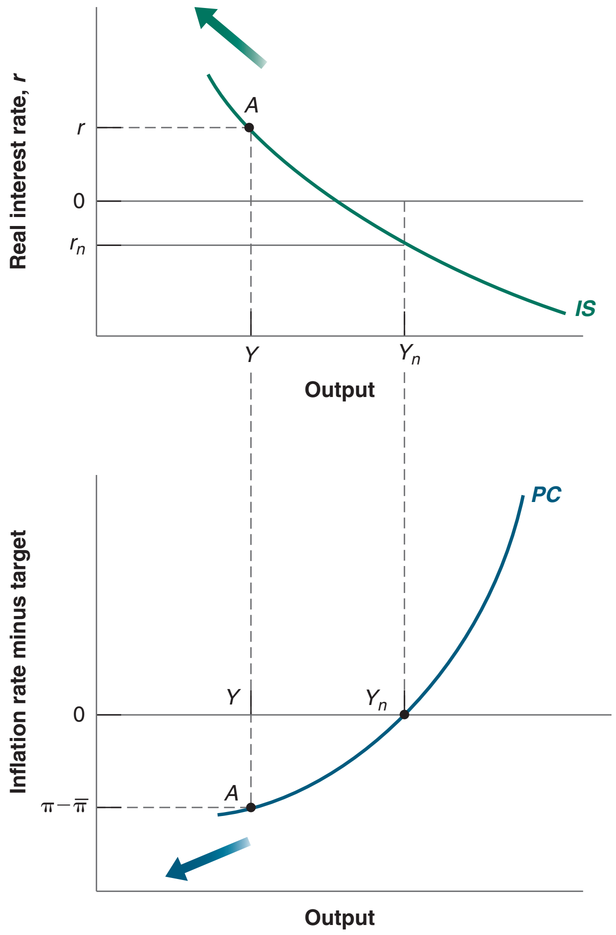


Figure 9-3

The Deflation Spiral

If the zero lower bound prevents monetary policy from increasing output back to potential, the result may be a deflation spiral. More deflation leads to a higher real rate, which in turn leads to lower output and more deflation.

instead of returning to its medium-run equilibrium, the economy moves further away, with lower and lower output and larger and large deflation.

This nightmare scenario is not just a theoretical concern. It played out during the Great Depression. As shown in the Focus Box "Deflation in the Great Depression," from 1929 to 1933 inflation turned into larger and larger deflation, steadily increasing the real rate and decreasing spending and output, until other policy measures were taken and the economy started turning around.

More recently, the Great Recession gave rise to a similar worry. With the nominal rate down to zero in the major advanced countries, the worry was that inflation would turn negative and start a similar spiral. Thankfully, it did not happen. Inflation decreased and in some countries—Greece, Spain, and Portugal, for example—turned to deflation. This limited the ability of the central bank to decrease the real rate and increase output, but deflation remained limited, and the deflation spiral did not happen. One reason for this, which connects to our previous discussion of expectation formation, is that inflation expectations remained largely anchored. Low output led to low inflation and in some cases mild deflation, but not to steadily larger deflation, as was the case during the Great Depression.

9-4 FISCAL CONSOLIDATION REVISITED

We can now take the IS-LM-PC model through its paces. In this section, we go back to the fiscal consolidation we discussed in Chapter 5 and look at not only its short-run effects but its medium-run effects as well.

FOCUS

Deflation in the Great Depression

After the collapse of the stock market in 1929, the US economy plunged into an economic depression. As the first two columns of Table 1 show, the unemployment rate soared from 3.2% in 1929 to 24.9% in 1933, and output growth was strongly negative for four years in a row. From 1933 on, the economy recovered slowly, but by 1940 the unemployment rate was still high at 14.6%.

The Great Depression has many elements in common with the Great Recession. A large increase in asset prices before the crash: housing prices in the recent crisis, stock market prices in the Great Depression, and the amplification of the shock through the banking system. There are also important differences. As you can see by comparing the output growth and unemployment numbers in Table 1 to the numbers for the Great Recession in Chapter 1, the decrease in output and the increase in unemployment were much larger then than in the Great Recession. In this box, we focus on just one aspect of the Great Depression: the evolution of the nominal and real interest rates and the dangers of deflation.

As you can see in the third column of the table, in the face of rising unemployment, the Fed decreased the nominal rate, measured in the table by the one-year T-bill rate, although it did this slowly and did not quite go all the way to zero. The nominal rate decreased from 5.3% in 1929 to 2.6% in 1933. At the same time, as shown in the fourth column, the decline in output and the increase in unemployment led to a sharp decrease in inflation. Inflation, equal to zero in 1929, turned negative in 1930, reaching −9.2% in 1931, and −10.8% in 1932. If we make the assumption that expected deflation was equal to actual deflation in each year, we can construct a series for the real rate. This is done in the last column of the table and gives a hint for why output continued to decline until 1933. The real rate reached 12.3% in 1931, 14.8% in 1932, and still a high 7.8% in 1933! It is no great surprise that, at those interest rates, both consumption and investment demand remained very low, and the depression worsened.

In 1933, the economy seemed to be in a deflation trap, with low activity leading to more deflation, a higher real interest rate, lower spending, and so on. Starting in 1934, however, deflation gave way to inflation, leading to a large decrease in the real interest rate, and the economy began to recover. Why, despite a high unemployment rate, the US economy was able to avoid further deflation remains a hotly debated issue. Some point to a change in monetary policy, a large increase in the money supply, leading to a change in inflation expectations. Others point to the policies of the New Deal, in particular the establishment of a minimum wage, thus limiting further wage decreases. Whatever the reason, this was the end of the deflation trap and the beginning of a long recovery.

For more on the Great Depression:

Lester Chandler, *America's Greatest Depression* (1970, HarperCollins), gives the basic facts. So does the book by John A. Garraty, *The Great Depression* (1986, Anchor).

Did Monetary Forces Cause the Great Depression? (1976, W. W. Norton), by Peter Temin, looks more specifically at the macroeconomic issues. So do the articles from a symposium on the Great Depression in the *Journal of Economic Perspectives*, Spring 1993.

For a look at the Great Depression in countries other than the United States, read Peter Temin's *Lessons from the Great Depression* (1989, MIT Press).

Table 1 Unemployment, Output Growth, Nominal Interest Rate, Inflation, and Real Interest Rate, 1929–1933

Year	Unemployment Rate (%)	Output Growth Rate (%)	One-Year Nominal Interest Rate (%), i	Inflation Rate (%), π	One-Year Real Interest Rate (%), r
1929	3.2	−9.8	5.3	0.0	5.3
1930	8.7	−7.6	4.4	−2.5	6.9
1931	15.9	−14.7	3.1	−9.2	12.3
1932	23.6	−1.8	4.0	−10.8	14.8
1933	24.9	9.1	2.6	−5.2	7.8

Suppose that output is at potential, so the economy is at point A in both the top and bottom graphs of Figure 9-4. Output is equal to Y_n, the real rate is equal to r_n, and inflation is equal to target. Now, assume that the government, which was running a deficit, decides to reduce it by, say, increasing taxes. The increase in taxes shifts the IS curve to the left, from IS to IS'. The new short-run equilibrium is given by point A' in the top and bottom graphs. At the given real rate r_n, output decreases from Y_n to Y' and inflation

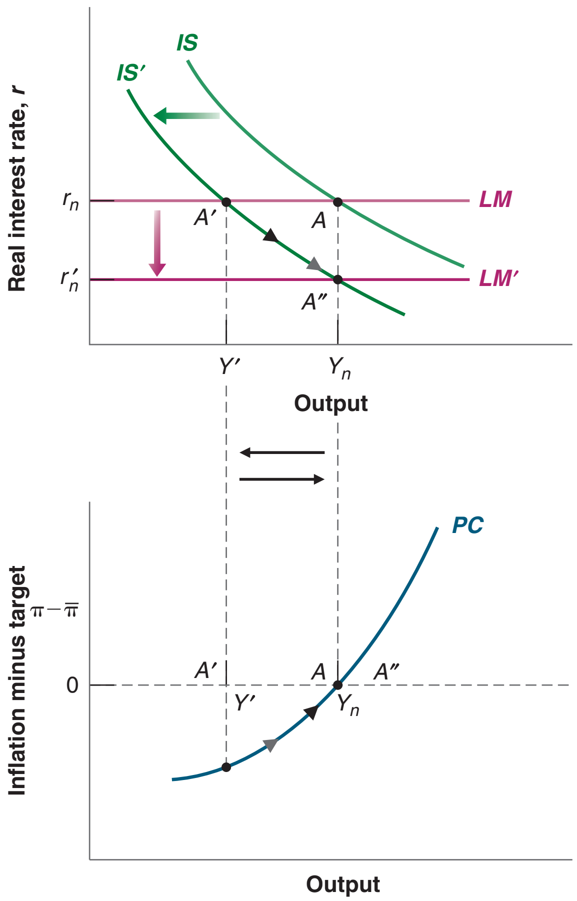


Figure 9-4

Fiscal Consolidation in the Short and the Medium Run

Fiscal consolidation leads to a decrease in output in the short run. In the medium run, output returns to potential, and the interest rate is lower.

is lower. In other words, if output was equal to potential output to start with, the fiscal consolidation, as desirable as it may be on other grounds, leads to a recession. This is the short-run equilibrium we characterized in Section 5-3 of Chapter 5. Note that, as income comes down and taxes increase, consumption decreases on both counts. Note also that, as output decreases, so does investment. In the short run, on macroeconomic grounds, fiscal consolidation looks rather unappealing: Both consumption and investment go down.

Let's, however, turn to the dynamics and to the medium run. As output is too low, and inflation is below target, the central bank is likely to react and decrease the policy rate until output is back to potential. In terms of Figure 9-4, the economy moves down the *IS*′ curve in the top graph and output increases. As output increases, the economy moves up the PC curve in the bottom graph until output is back to potential. Thus, the medium-run equilibrium is given by point *A*″ in both the top and bottom graphs. Output is back at Y_n and inflation is again equal to target inflation. The real rate needed to maintain output at potential is now lower than before, equal to r'_n rather than r_n.

Now look at the composition of output in this new equilibrium. As income is the same as it was before fiscal consolidation but taxes are higher, consumption is lower, although not as low as it was in the short run. As output is the same as before but the interest rate is lower, investment is higher than before. In other words, the decrease in consumption is offset by an increase in investment, so demand, and by implication output, is unchanged. This is in sharp contrast to what happened in the short run and makes fiscal consolidation look more attractive. Although consolidation may decrease investment in the short run, it increases investment in the medium run.

We have looked at a fiscal consolidation, or equivalently, at an increase in public saving. The same argument would apply to an increase in private saving. At a given real rate, such an increase would lead to a decrease in investment in the short run, but to an increase in investment in the medium run. (In light of these results, you may want to go back to the Focus Boxes "The Paradox of Saving" in Chapter 3 and "Deficit Reduction: Good or Bad for Investment?" in Chapter 5.)

This discussion raises some of the same issues we discussed in the previous section.

It looks as if fiscal consolidation could take place without a decrease in output in the short run. All that is needed is for the central bank and the government to coordinate carefully. As fiscal consolidation takes place, the central bank should decrease the real rate so as to maintain output at the natural level. In other words, the proper combination of fiscal and monetary policy could achieve the medium-run equilibrium outcome in the short run. Such coordination indeed happens sometimes; as we saw in Chapter 5, it happened in the United States in the 1990s, when a fiscal consolidation was accompanied with a monetary expansion. But it does not always happen. One reason is that the central bank may be unable to decrease the real rate sufficiently because of the zero lower bound. This was the case in the euro area during the Great Recession: with the nominal policy rate at zero, monetary policy was unable to offset the adverse effects of fiscal consolidation on output. The result was a stronger and longer-lasting adverse effect of fiscal consolidation on output than would have been the case had the European Central Bank been able to decrease the real rate further.

This has led to an intense debate as to whether "fiscal austerity," i.e., the fiscal consolidation that took place in Europe from 2010 on, was justified. More on this in Chapter 22.

9-5 THE EFFECTS OF AN INCREASE IN THE PRICE OF OIL

So far we have looked at shocks to demand, shocks that shifted the IS curve but left potential output and thus the position of the PC curve unaffected. There are other shocks, however, that affect both demand and potential output and play an important role in fluctuations. An obvious candidate is movements in the price of oil. To see why, turn to Figure 9-5.

Figure 9-5 plots two series. The first, represented by the blue line, is the price of a barrel of oil in dollars since 1970. It is measured on the vertical axis on the left. This is the series that is quoted in the newspapers every day. What matters, however, for economic decisions is not the dollar price, but the real price of oil; that is, the dollar price of oil divided by the price level. Thus, the second series in the figure, represented by the red line, shows the real price of oil, constructed as the dollar price of oil divided by the US consumer price index. Note that the real price, measured on the vertical axis on the right, is an index; it is normalized to equal 100 in 1970.

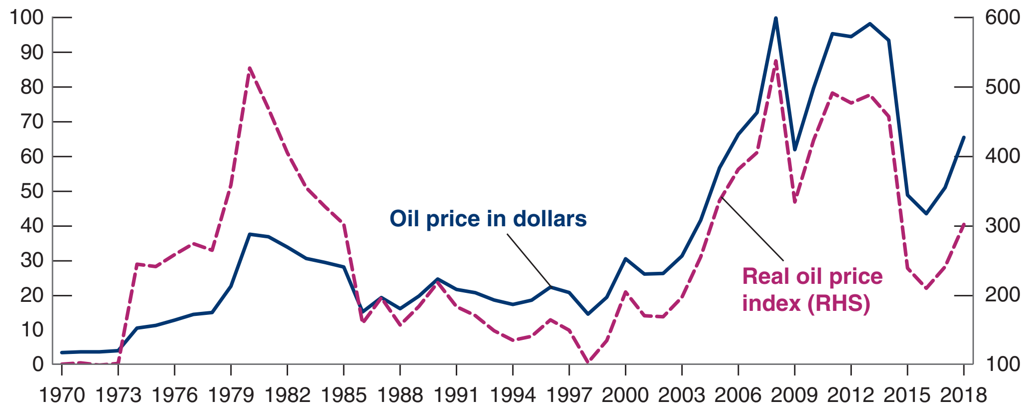


Figure 9-5

The Nominal and the Real Price of Oil, 1970–2018

Over the last 40 years, there have been two sharp increases in the real price of oil, the first in the 1970s and the second in the 2000s.

Source: FRED: OILPRICE, CPIAUSCL. Left vertical axis: nominal price of oil, dollars per barrel. Right vertical axis: real oil price index, 1970=100.

What is striking in the figure is the size of the movements in the real price of oil. Twice over the last 50 years, the US economy was hit with a fivefold increase in the real price of oil, the first time in the 1970s and the second time in the 2000s. The financial crisis then led to a dramatic drop in late 2008, followed by a partial recovery. What was behind the two large increases? In the 1970s, the main factors were the formation of **OPEC** (the **Organization of Petroleum Exporting Countries**), a cartel of oil producers that was able to act as a monopoly and increase prices, and disruptions due to wars and revolutions in the Middle East. In the 2000s, the main factor was quite different, namely the fast growth of emerging economies, in particular China, which led to a rapid increase in the world demand for oil and, by implication, a steady increase in real oil prices.

What was behind the recent two decreases in 2008 and 2014? The decrease at the end of 2008 was the result of the crisis, which led to a large recession and in turn a large and sudden decrease in the demand for oil. The causes of the more recent drop since 2014 are still being debated. Most observers believe that it is a combination of increased supply because of the increase in shale oil production in the United States and the partial breakdown of the OPEC cartel.

Let's focus on the two large increases. Although the causes were different, the implication for US firms and consumers was the same: more expensive oil. The question is: What would we expect the short- and medium-run effects of such increases to be? In answering the question, we face a problem: The price of oil appears nowhere in the model we have developed so far! The reason is that, until now, I have assumed that output was produced using only labor. One way to extend our model would be to recognize explicitly that output is produced using labor *and* other inputs (including energy), and then figure out what effect an increase in the price of oil has on the price set by firms and on the relation between output and employment. An easier way, and the way we shall go here, is simply to capture the increase in the price of oil by an increase in m—the markup of the price over the nominal wage. The justification is straightforward. Given wages, an increase in the price of oil increases the cost of production, forcing firms to increase prices to maintain the same profit rate.

Having made this assumption, we can then track the dynamic effects of an *increase in the markup* on output and inflation.

Effects on the Natural Rate of Unemployment

Let's start by asking what happens to the natural rate of unemployment when the real price of oil increases (for simplicity, I shall drop "real" in what follows). For this, we need to return to Chapter 7 and the characterization of labor market equilibrium. Figure 9-6 reproduces the characterization of labor market equilibrium from Figure 7-8 in Chapter 7.

The wage-setting relation (WS) is downward sloping; a higher unemployment rate leads to lower real wages. The price-setting relation (PS) is represented by the horizontal line at $W/P = 1/(1 + m)$. The initial equilibrium is at point A, and the initial natural unemployment rate is u_n. An increase in the markup leads to a downward shift of the price-setting line, from PS to PS'. The higher the markup, the lower the real wage implied by price setting. The equilibrium moves from A to A'. The real wage is lower and the natural unemployment rate is higher. Think of it this way: Because firms have to pay more for the oil, the wage they can pay is lower. Getting workers to accept the lower real wage requires an increase in unemployment.

Firms may, for some time, react by accepting a decrease in profit and so increase the price less than in the figure. But, if the price of oil remains high, eventually they will want to reestablish their profit margins.

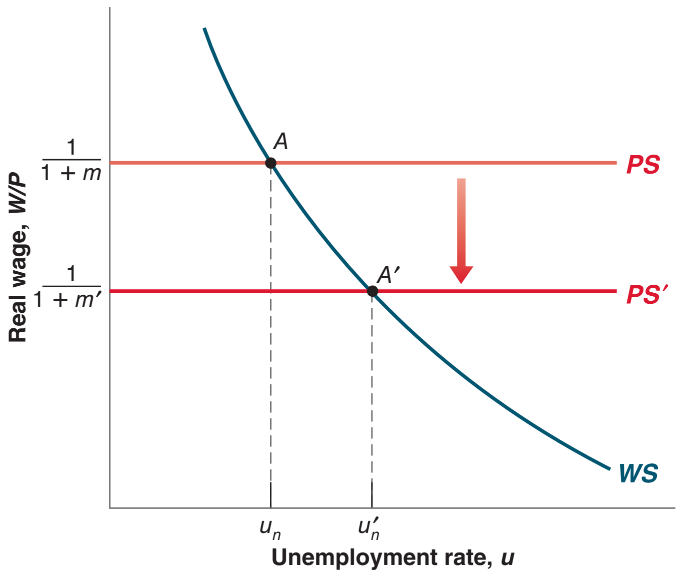
The increase in the natural rate of unemployment leads in turn to a decrease in the natural level of employment. If we assume that the relation between employment and output is unchanged—that is, that each unit of output still requires one worker in addition to the energy input—then the decrease in the natural level of employment leads to an identical decrease in potential output. Putting things together: An increase in the price of oil leads to a decrease in potential output.

This assumes that the increase in the price of oil is permanent. If, in the medium run, the price of oil goes back to its initial value, then the medium run natural rate of unemployment is clearly unaffected.

Figure 9-6

The Effects of an Increase in the Price of Oil on the Natural Rate of Unemployment

An increase in the price of oil is equivalent to an increase in the markup. It leads to lower real wages and a higher natural rate of unemployment.



We can now go back to the IS-LM-PC model, and this is done in Figure 9-7. Assume the initial equilibrium is at point A in both the top and bottom panels, with output at potential, so Y is equal to Y_n, inflation is equal to target, and the real rate is equal to r_n. As the price of oil increases, the natural level of output decreases (this is what we just saw), say from Y_n to Y'_n. The PC curve shifts up, from PC to PC'. If the IS curve does not shift (we return to this assumption later) and the central bank does not change the real rate, output does not change, but the same level of output is now associated with inflation above the target. The short-run equilibrium is given by point A' in the top and bottom panels. In the short run, output does not change, but inflation is higher.

Turn to dynamics. If the central bank were to leave the real rate unchanged, output would continue to exceed the now lower level of potential output, and inflation would remain high. Eventually, expectations would become de-anchored, wage setters would start expecting higher inflation, and, as we have seen, inflation would keep increasing. To avoid this, the central bank must increase the real rate to return inflation to target. As it does so, the economy moves from A' to A'' along the IS curve in the top panel, and from A' to A'' along the PC curve in the bottom panel. As output decreases to its lower level, inflation returns to target. Once the economy is at point A'', the economy is in its medium-run equilibrium. Because potential output is lower, the increase in the price of oil is reflected in a permanently lower level of output. Note that along the way, *lower output* is associated with *inflation higher* than the target, a combination that economists call **stagflation** (stag- for stagnation, and -flation for inflation).

Stagflation: Lower output, higher inflation. ►

As in the previous sections, this description raises many issues.

The first is our assumption that the IS curve does not shift. In fact, there are many channels through which the increase in the price of oil may affect demand and shift not only the PC curve but also the IS curve. The higher price of oil may lead firms to change their investment plans, canceling some investment projects, shifting to less energy-intensive equipment. The increase in the price of oil also redistributes income from oil buyers to oil producers. Oil producers may spend less than oil buyers, leading to a decrease in demand. This is especially true if the oil producers are in other countries than the oil buyers (which is the case when the United States buys oil from the Middle East, for example). As the price increases and their income increases, the oil producers are likely to spend most of it on their own goods, not on the goods produced by the oil buyers. Thus, demand for domestic goods is likely to go down. So it may well be that the IS curve shifts to the left, leading to a decrease in output not only in the medium run but in the short run as well.

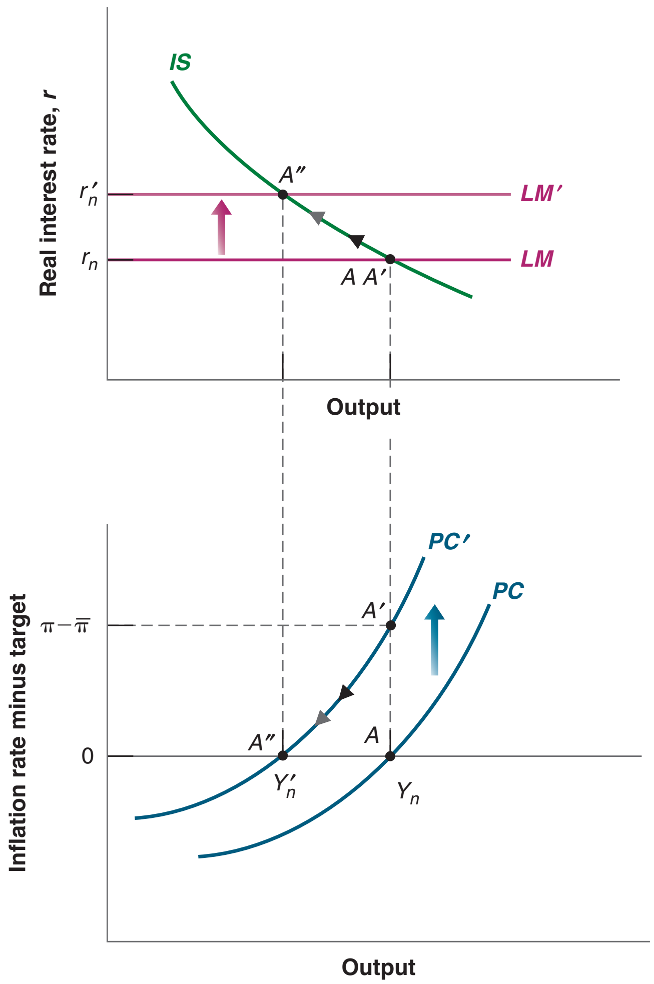


Figure 9-7

Short- and Medium-Run Effects of an Increase in the Price of Oil

In the short run, an increase in the price of oil leads to higher inflation. If the price increase is permanent, it leads to lower output in the medium run.

The second issue has to do with the persistence of the increase in the price of oil. What we looked at in Figure 9-7 was a permanent increase, leading in turn to a permanent decrease in potential output. The central bank had no alternative than to tighten and reduce inflation. But if the increase is temporary, then the central bank may want to let inflation be higher while the price of oil is high, anticipating that inflation will return to target by itself when the price of oil comes down. This may, however, be a risky strategy: If the price of oil does not decrease quickly, inflation will remain high for some time. Expectations of inflation may de-anchor, leading to higher and higher inflation and making the eventual adjustment more painful. You can see why this makes the job of the central bank much harder than suggested in Figure 9-7. Assessing whether the increase in the price of oil is temporary or permanent and whether and when inflation expectations may de-anchor is, in practice, very difficult.

This discussion again shows the importance of expectation formation on the dynamic effects of shocks. It also helps explain the difference between the effects of the price of oil in the 1970s, which led to high inflation and a large recession, and the much more benign effects of the price of oil in the 2000s. This is explored at more length in the Focus Box "Oil Price Increases: Why Were the 2000s So Different from the 1970s?"

9-6 CONCLUSIONS

This chapter has covered a lot of ground. Let me repeat some key ideas and develop some of the conclusions.

FOCUS

Oil Price Increases: Why Were the 2000s So Different from the 1970s?

Why is it that oil price increases were associated with stagflation in the 1970s but had little apparent effect on the US economy in the 2000s?

A first line of explanation is that shocks other than the increase in the price of oil were at work in the 1970s but not in the 2000s. In the 1970s, not only did the price of oil increase, but so did the price of many other raw materials. So the effect was stronger than would have been the case had only the price of oil increased.

Econometric studies suggest, however, that more was at work, and that, even after controlling for the presence of these other factors, the effects of the price of oil have changed since the 1970s. Figure 1 shows the effects of a 100% increase in the price of oil on output and on the price level, estimated using data from two different periods. The black and blue lines show the effects of an increase in the price of oil on the consumer price index (CPI) and on gross domestic product (GDP), based on data from 1970:1 to 1986:4; the green and red lines do the same, but based on data from 1987:1 to 2006:4 (the time scale on the horizontal axis is in quarters). The figure suggests two main conclusions. First, in both periods, as predicted by our model, the increase in the price of oil led to an increase in the CPI and a decrease in GDP. Second, the effects of the increase in the price of oil on both the CPI and on GDP have become smaller, roughly half of what they were previously.

Why have the adverse effects of the increase in the price of oil become smaller? This is still a topic of research. But, at this stage, two hypotheses appear plausible.

The first hypothesis is that partly because of globalization and foreign competition, US workers have less bargaining power today than they did in the 1970s. Thus, as the price of oil has increased, workers have had no choice than to accept a reduction in real wages, limiting the increase in the unemployment rate.

The second hypothesis has to do with monetary policy. As we saw in Chapter 8, in the 1970s, inflation expectations were not anchored. Seeing the initial increase in inflation due to the increase in the price of oil, wage setters assumed that inflation would continue to be high and thus asked for higher nominal wages, which led to further increases in inflation. In contrast, in the 2000s, inflation expectations were anchored. Seeing the initial increase in inflation, wage setters assumed it was a one-time increase and did not change their expectations of future inflation as much as they would have in the 1970s. Thus, the effect on inflation was much more muted, and the need for the Fed to control inflation through higher real rates and low output was much more limited.

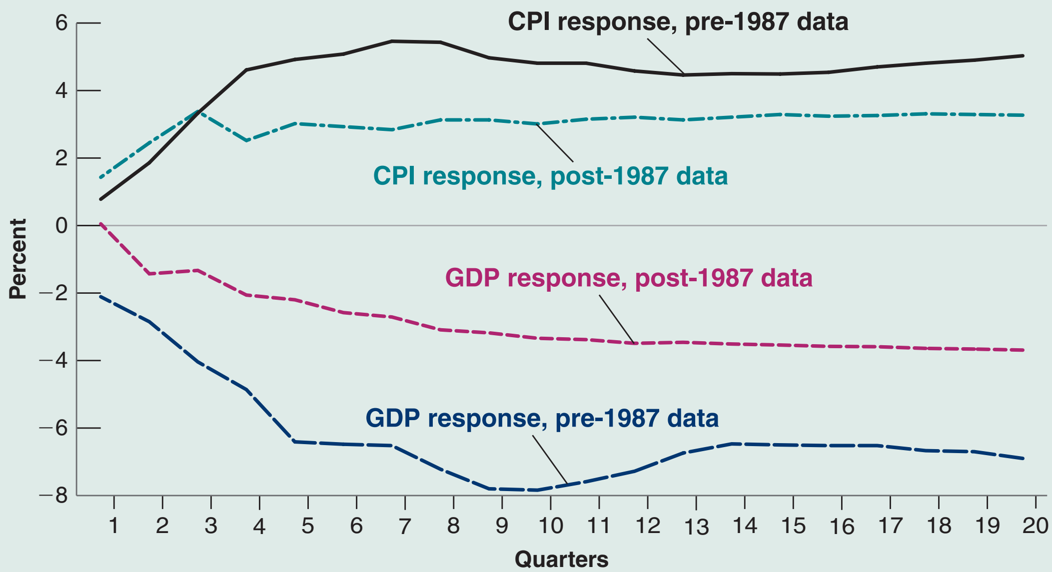


Figure 1

The Effects of a 100% Permanent Increase in the Price of Oil on the CPI and on GDP

The effects of an increase in the price of oil on output and the price level are smaller than they used to be.

The Short Run versus the Medium Run

One key message of this chapter is that shocks or changes in policy typically have different effects in the short run and in the medium run. Disagreements among economists about the effects of various policies often come from differences in the time frame they have in mind. If you are worried about output and investment in the short run, you might be reluctant to proceed with fiscal consolidation. But if your focus is on the medium and long run, you will see the consolidation as helping investment and eventually, through higher investment and thus capital accumulation, increasing output. One implication is that where you stand depends, among other things, on how fast you think the economy adjusts to shocks. If you believe that it takes a long time for output to return to potential you will naturally focus more on the short run and be willing to use policies that increase output in the short run, even if medium-run effects are nil or negative. If you believe instead that output returns to potential quickly, you will put more emphasis on the medium-run implications and will, by implication, be more reluctant to use those policies.

Shocks and Propagation Mechanisms

This chapter also gives you a general way of thinking about **output fluctuations** (sometimes called **business cycles**)—movements in output around its trend (a trend that we have ignored so far but on which we will focus in Chapters 10 through 13).

You can think of the economy as being constantly hit by **shocks**. These shocks may be shifts in consumption coming from changes in consumer confidence, or shifts in investment, or shifts in the risk premium due to a financial crisis, and so on. Or they may come from changes in policy—such as the introduction of a new tax law, a new program of infrastructure investment, a decision by the central bank to fight inflation, or a trade war.

How to define *shocks* is harder than it looks. Suppose a failed economic program in an Eastern European country leads to political chaos in that country, which leads to increased risk of nuclear war in the region, which leads to a fall in consumer confidence in the United States, which leads to a recession in the United States. What is the "shock"? The failed program? The fall of democracy? The increased risk of nuclear war? Or the decrease in consumer confidence? In practice, we have to cut the chain of causation somewhere. Thus, we may refer to the drop in consumer confidence as the shock and ignore its underlying causes.

Each shock has dynamic effects on output and its components. These dynamic effects are called the **propagation mechanism** of the shock. Propagation mechanisms are different for different shocks. The effects of a shock on activity may build up over time, affecting output in the medium run, or they may build up for a while and then decrease and disappear. Some shocks are sufficiently large or come in sufficiently bad combinations that they create a recession. The two recessions of the 1970s were due largely to increases in the price of oil; the recession of the early 1980s was due to a sharp tightening of monetary policy; the recession of the early 1990s was due primarily to a sudden decline in consumer confidence; the recession of 2001 was due to a sharp drop in investment spending. The sharp decrease in output in 2009 had its origins in the problems of the housing market, which led to a major financial shock and in turn to a sharp reduction in output. What we call *economic fluctuations* are the result of these shocks and their dynamic effects on output.

Typically, the economy returns over time to its medium-run equilibrium. A major issue is whether we can expect the economy to return to that equilibrium on its own, or whether it requires an active policy response. While this is a controversial issue, I believe that the economy does not self-stabilize, and that an active policy response, whether through monetary or through fiscal policy, is typically needed. Indeed, in all the examples we saw in this chapter, a reaction from the central bank was the mechanism through which output returned to potential output over time. In the absence of a policy reaction, inflation would have either steadily increased or decreased, eventually requiring a more painful policy adjustment. This was particularly clear in the discussion of the zero lower bound case: To the extent that monetary policy could no longer be used, there was a risk that the economy may enter a deflation spiral, getting steadily worse over time. During the Great Recession, the worst was avoided, partly because the financial system slowly recovered and partly because of the use of fiscal policy.

SUMMARY

- In the short run, output is determined by demand. The output gap, defined as the difference between output and potential output, affects inflation.
- A positive output gap leads to higher inflation, which leads the central bank to increase the real rate. The increase in the real rate leads to a decrease in output and thus to a decrease in the output gap. Symmetrically, a negative output gap leads to lower inflation, which leads the central bank to decrease the real rate. The decrease in the real rate increases output and thus decreases the output gap.
- In the medium run, output is equal to potential output. The output gap is equal to zero, and inflation is equal to target inflation. The real interest rate associated with output equal to potential is called the *natural interest rate*.
- When the output gap is negative, the combination of the zero lower bound and deflation may lead to a *deflation spiral*. Lower output leads to lower inflation, which leads to a higher real interest rate. If inflation remains low or negative, inflation expectations may de-anchor, leading to larger and larger deflation, higher and higher real interest rates, and further decreases in output.
- In the short run, a fiscal consolidation through higher taxes leads, at an unchanged real rate, to a decrease in output, a decrease in consumption, and a decrease in investment. In the medium run, output returns to potential. Consumption is lower, and investment is higher.
- An increase in the price of oil leads in the short run to higher inflation. Depending on the effect of the price of oil on demand, it may also lead to a decrease in output. The combination of higher inflation and lower output is called *stagflation*. In the medium run, the increase in the price of oil leads to lower potential output and thus lower actual output.
- The difference between short-run and medium-run effects of policies is one of the reasons economists disagree in their policy recommendations. Some economists believe the economy adjusts quickly to its medium-run equilibrium, so they emphasize medium-run implications of policy. Others believe the adjustment mechanism through which output returns to the natural level of output is a slow process at best, and so they put more emphasis on the short-run effects of policy.
- Economic fluctuations are the result of a continual stream of shocks and of the dynamic effects of each of these shocks on output. Sometimes the shocks are sufficiently adverse, alone or in combination, that they lead to a recession.

KEY TERMS

potential output, 179
output gap, 179
labor hoarding, 180
Okun coefficient, 181
natural rate of interest, 183
neutral rate of interest, 183
Wicksellian rate of interest, 183
neutrality of money, 183
deflation spiral, 184
deflation trap, 184
Organization of Petroleum Exporting Countries (OPEC), 189
stagflation, 190
output fluctuations, 193
business cycles, 193
shocks, 193
propagation mechanism, 193

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The *IS* curve shifts to the right with an increase in G, to the right with an increase in T, and to the right with an increase in x.
b. If $(u - u_n)$ is greater than zero, then $(Y - Y_n)$ is greater than zero.
c. If $(u - u_n)$ is equal to zero, the output is at potential.
d. If $(u - u_n)$ is less than zero, the output gap is negative.
e. If the output gap is positive, inflation is higher than expected inflation.
f. Okun's law says that if output growth increases by one percentage point, the rate of unemployment drops by one percentage point.
g. At the natural rate of unemployment, inflation is equal to expected inflation.
h. In a medium-run equilibrium, the rate of inflation is stable at zero.
i. The central bank can always act to keep output equal to potential output.
j. Active policy changes, either in monetary policy or fiscal policy, are required to return the economy to medium run equilibrium if a shock is permanent.
k. A permanent increase in government spending would require a higher real policy rate be put in place by the central bank in order to keep inflation stable.
l. A large increase in the price of oil increases the natural rate of unemployment.

2. Identifying if an economy is in medium run equilibrium and the necessary central bank action to return the economy to medium run equilibrium

Here are values for a hypothetical economy

$Y_n = 1000$; $u_n = 5\%$; $r_n = 2\%$; $x = 1\%$; $\pi^e = 2\%$ and a table describing this economy in various situations:

Situation	Y_n	Y	C	I	G	i (%)	π(%)	u (%)	x (%)
A	1000	1000	700	150	150	4	2	5	1
B	1000	1050	730	170	150	2	3	3	1
C	1000	950	670	130	150	4	1	8	3
D	1000	950	670	150	130	4	1	8	1
E	1000	1050	730	150	170	4	3	3	1

a. Explain why Situation A is a medium run equilibrium and Situation B, C, D and E are not a medium run equilibrium.
b. What is the action to be taken by the central bank to move from Situation B to medium run equilibrium?
c. What is the action to be taken by the central bank to move from Situation C to medium run equilibrium?
d. What is the action to be taken by the central bank to move from Situation D to medium run equilibrium?
e. What is the action to be taken by the central bank to move from Situation E to medium run equilibrium?

3. The medium-run equilibrium is characterized by four conditions:

Output is equal to potential output $Y = Y_n$ and the real policy rate r_n must be chosen by the central bank so:

The unemployment rate is equal to the natural rate $u = u_n$.

The real policy interest rate is equal to the natural rate of interest r_n where r_n is defined as the policy rate where $Y_n = C(Y_n - T) + I(Y_n, r_n + x) + G$.

The expected and actual rate of inflation π^e is equal to the anchored or target rate of inflation $\overline{\pi}$. This implies the nominal policy rate $i = r_n + \overline{\pi}$.

a. If the level of expected inflation is formed so π^e equals $\overline{\pi}$, characterize the behavior of inflation in a medium-run equilibrium.
b. Write the *IS* relation as $Y = C(Y - T) + I(Y, r + x) + G$. Suppose r_n is 2%. If x increases from 3 to 5%, how must the central bank change r_n to maintain the existing medium-run equilibrium. Explain in words.
c. Suppose G increases permanently. In what direction must the central bank change r to maintain the existing medium-run equilibrium? Explain in words.
d. Suppose T decreases permanently. In what direction must the central bank change r to maintain the existing medium-run equilibrium? Explain in words.
e. Discuss: In the medium run, a fiscal expansion leads to an increase in the natural rate of interest.

4. This chapter assumes that expected inflation remains equal to the central bank's target rate of inflation. In Chapter 8, in the discussion of the Phillips curve, it was noted that expected inflation was, for some time, equal to lagged inflation and was not anchored by the central bank's target rate of inflation. This question considers the implications of these two assumptions about expected inflation for the effects of a permanent change in demand, given unchanged monetary policy.

The permanent change in demand studied in this question is an increase in consumer confidence where the parameter "a" from Chapter 3 takes a larger value.

One assumption is that the level of expected inflation equals lagged inflation and so changes over time. The other assumption is that the level of expected inflation is anchored to the central bank's target rate of inflation and never changes

Begin in medium-run equilibrium where actual and expected inflation equals 2% in period t. Suppose there is an increase in consumer confidence in period $(t + 1)$.

Parts (a), (b) and (c) assume expected inflation in each period equals lagged inflation from the previous period. For example, in period $(t + 2)$ $\pi^e_{t+2} = \pi_{t+1}$ and in period $(t + 1)$, $\pi^e_{t+1} = \pi_t$.

a. How does the *IS* curve shift from period t to period $(t + 1)$? What is the value of expected inflation in period t? How does the short-run equilibrium output and inflation rate in period $(t + 1)$ compare to the equilibrium output and inflation rate in period *t* if the central bank does not change the real policy rate from period to period $(t + 1)$?
b. Now advance the economy to the period $(t + 2)$ equilibrium under the assumption that $\pi^e_{t+2} = \pi_{t+1}$ and consumer confidence remains high. If the central bank leaves the real policy rate unchanged will inflation be higher or lower in period $(t + 2)$ than in period $(t + 1)$?
c. What do you conclude about the central bank policy of keeping the real policy rate unchanged in period $(t + 2)$? Is it sustainable?

Parts (d), (e) and (f) assume expected inflation remains equal to target inflation, so $\pi^e = \overline{\pi}$ in all periods.

d. Consider the period $(t + 1)$ equilibrium given the assumption that $\pi^e = \overline{\pi}$. If the central bank leaves the real policy rate unchanged, how does inflation in period $(t + 1)$ compare to inflation in period t? Is output higher or lower in period $(t + 1)$ than in period t?
e. Consider the period $(t + 2)$ equilibrium given the assumption that $\pi^e_{t+2} = \overline{\pi}$. If the central bank leaves the real policy rate unchanged, how does actual inflation in period $(t + 2)$ compare to inflation in period $(t + 1)$? Is output higher or lower in period $(t + 2)$ than in period $(t + 1)$?
f. Explain why the policy choice to maintain the real policy rate at the original level in period $(t + 1)$ is not a sustainable policy?

Comparing the economic outcomes in parts (a), (b) and (c) to the economic outcomes in (d), (e) and (f)

g. Compare the inflation, expected inflation and output outcomes in part (a), (b) and (c) to that in parts (d), (e) and (f).
h. Which assumption about expected inflation, do you think is more realistic. Discuss.

5. A shock to aggregate supply will also have different outcomes when there are different assumptions about the formation of the level of expected inflation. As in Question 4, one assumption is *that the level of expected inflation equals lagged inflation. The level of expected inflation changes over time. The second assumption is that the level of expected inflation is anchored to the central bank's target value and never changes. Begin in medium-run equilibrium where actual and expected inflation equal 2% in period t. Then there is a permanent increase in the price of oil in period $(t + 1)$.*

Parts (a), (b) and (c) assume expected inflation in each period equals lagged inflation from the previous period. For example, in period $(t + 2)$ $\pi^e_{t+2} = \pi_{t+1}$ and in period $(t + 1)$, $\pi^e_{t+1} = \pi_t$.

a. How does the *PC* curve shift from period t to period $(t + 1)$? Assume that the central bank does not change the real policy rate. What happens to output in period $(t + 1)$ compared to period t? What happens to inflation in period $(t + 1)$ compared to period (t)?
b. Consider the period $(t + 2)$ equilibrium under the assumption that $\pi^e_{t+2} = \pi_{t+1}$ and that the central bank does not change the real policy rate. How does inflation in period $(t + 2)$ compare to inflation in period $(t + 1)$? How does output in period $(t + 2)$ compare to output in period $(t + 1)$?
c. Is the policy choice to maintain the real policy rate at its period t level sustainable?

Parts (d), (e) and (f) assume expected inflation remains equal to target inflation, so $\pi^e = \overline{\pi}$ in all periods.

d. Consider the period $(t + 1)$ equilibrium given the assumption that $\pi^e_{t+1} = \overline{\pi}$. If the central bank leaves the real policy rate unchanged, how does actual inflation in period $(t + 1)$ compare to actual inflation in period t?
e. Consider the period $(t + 2)$ equilibrium given the assumption that $\pi^e_{t+2} = \overline{\pi}$ and assuming that the central bank leaves the real policy rate unchanged. How does actual inflation in period $(t + 2)$ compare to inflation in period $(t + 1)$?
f. Is the policy choice to maintain the real policy rate at the period t level a sustainable policy?

Comparing the economic outcomes in parts (a), (b) and (c) to the economic outcomes in (d), (e) and (f)

g. Compare the inflation and output outcomes in part (a), (b) and (c) to that in parts (d), (e) and (f).
h. Which assumption about expected inflation, do you think is more realistic. Discuss.

DIG DEEPER

6. Okun's Law is written as $u - u(-1) = -0.4\,(g_Y - 3\%)$

a. What is the sign of $u - u(-1)$ in a recession? What is the sign of $u - u(-1)$ in a recovery?
b. Explain where the 3% number comes from?
c. Explain why the coefficient on the term $(g_Y - 3\%)$ is -0.4 and not -1.
d. Suppose the number of immigrants per year allowed entering the United States is sharply increased. How would Okun's law change?

7. Fiscal consolidation at the Zero Lower Bound

Suppose the economy is operating at the zero lower bound for the nominal policy rate; the initial equilibrium is at the positive target rate of inflation and the economy is resting at potential output but there is a large government deficit in period t. A newly elected government vows to cut spending and permanently reduces the deficit by cutting G and raising T in period $t + 1$ and subsequent periods.

a. Draw a sketch of the initial equilibrium in this situation, then work out the effects of the policy on output in period $t + 1$, $t + 2$ and so as long as expected inflation remains anchored at its initial level.
b. Suppose with inflation persistently less than the target rate of inflation, the expected rate of inflation falls. When would inflation become negative in this situation? If expected inflation changed to a negative value, what happens to the real policy rate? How will this affect output in in the period where expected inflation becomes negative?
c. How does the zero lower bound on nominal interest rates make a fiscal consolidation more difficult? Could this policy lead to a deflation spiral?

EXPLORE FURTHER

8. Consider the data in the Focus Box, "Deflation in the Great Depression."

a. Do you believe that output had returned to its potential level in 1933?
b. Which years suggest a deflation spiral as described in Figure 9-3?
c. Make the argument that if the expected level of inflation had remained anchored at the actual value of inflation in 1929, the Great Depression would have been less severe.
d. Make the argument that a substantial fiscal stimulus in 1930 would have made the Great Depression less severe.

9. Consider the data in the Focus Box, "Deflation in the Great Depression."

a. Calculate real interest rates in each year making the assumption that the expected level of inflation is last year's rate of inflation. The rate of inflation in 1928 was −1.7%. Do the changes in real interest rates explain the data on real output growth and unemployment better than when you make the assumption the expected rate of inflation is the current year's rate of inflation?
b. Calculate the Okun's law coefficient for each year from 1930 to 1933. To do so, assume potential output is not growing. Speculate on why firms did not take on additional workers in 1933 even though output growth was 9.1%. *Hint:* If potential output is not growing, Okun's law is $u - u(-1) = -\alpha g_Y$.

10. The Great Depression in the United Kingdom

Answer the following questions based on information found in the table below

a. Is there evidence of the deflation spiral from 1929 to 1933 in the United Kingdom?
b. Is there evidence of the effect of high real interest rates on output?
c. Is there evidence of a poor choice of the real policy interest rate by the central bank?

The Nominal Interest Rate, Inflation, and the Real Interest Rate in the United Kingdom, 1929–1933					
Year	Unemployment Rate (%)	Output Growth Rate (%)	One-Year Nominal Interest Rate (%), i	Inflation Rate (%), π	One-Year Real Interest Rate (%), r
1929	10.4	3.0	5.0	−0.9	5.9
1930	21.3	−1.0	3.0	−2.8	5.8
1931	22.1	−5.0	6.0	−4.3	10.3
1932	19.9	0.4	2.0	−2.6	4.6
1933	16.7	3.3	2.0	−2.1	4.1

11. President's Trump's complaints about Federal Reserve policy in 2018 made the news. On December 22, 2017 President Trump signed a bill, the Tax Cuts and Jobs Act that substantially reduced federal taxes in 2018.

a. *On November 20, 2018, Bloomberg reported President Trump as saying* The president tells reporters that the central bank is a "problem" and that he would "like to see the Fed with a lower interest rate."

How does this comment fit into our model of a permanent fiscal expansion?

b. *On September 25, 2018, according to Bloomberg, President Trump said after the The Fed raised interest rates by a quarter percentage point.*

"We are doing great as a country," Trump says at a press conference in New York. "Unfortunately they just raised interest rates a little bit because we are doing so well. I am not happy about that."

How does this comment fit into our model of the Phillips Curve?

This page is intentionally left blank

The Long Run

The next four chapters focus on the long run. In the long run, what dominates is not fluctuations, but growth. So now we need to ask: What determines growth?

Chapter 10

Chapter 10 looks at the facts of growth. It first documents the large increase in output that has taken place in rich countries since 1950. Then, taking a wider look, it shows that on the scale of human history, such growth is a recent phenomenon. And it is not a universal phenomenon: Some countries are catching up, but some poor countries are suffering from no or low growth.

Chapter 11

Chapter 11 focuses on the role of capital accumulation in growth. It shows that capital accumulation cannot by itself sustain growth, but that it does affect the level of output. A higher saving rate typically leads to lower consumption initially, but to more consumption in the long run.

Chapter 12

Chapter 12 turns to technological progress. It shows how, in the long run, the growth rate of an economy is determined by the rate of technological progress. It then looks at the role of research and development, and institutions in generating such progress.

Chapter 13

Chapter 13 looks at some of the challenges to growth. It discusses whether we can expect technological progress to continue at the same rate, whether robots might lead to mass unemployment, whether growth comes with rising inequality and what can be done about it, and whether we can mitigate climate change.

This page is intentionally left blank

The Facts of Growth

10

Our perceptions of how the economy is doing are often dominated by year-to-year fluctuations in economic activity. A recession leads to gloom, and an expansion to optimism. But if we step back to get a look at activity over longer periods—say, over many decades—the picture changes. Fluctuations fade. **Growth**, which is the steady increase in aggregate output over time, dominates the picture.

There is a famous quote from Robert Lucas, the 1995 Nobel Prize winner: "Once one starts thinking about growth, it is hard to think about anything else."

Figure 10-1 shows the evolution of US GDP (panel (a)) and of US GDP per person (panel (b)) (both in 2012 dollars) since 1890. (The scale used to measure GDP or GDP per person on the vertical axis is called a **logarithmic scale**. The defining characteristic of a logarithmic scale is that the same proportional increase in a variable is represented by the same distance on the vertical axis.)

For more on log scales, see Appendix 2 at the end of the book.

The vertical bars from 1929 to 1933 correspond to the large decrease in output during the Great Depression: the other sets of bars correspond to the 1980–1982 recession, which is the largest postwar recession before the financial crisis and the Great Recession of 2008–2010. Note how small these three episodes appear compared to the steady increase in output per person over the last 130 years. The cartoon on p203 makes the same point about growth and fluctuations, in an even more obvious way.

With this in mind, we now shift our focus from fluctuations to growth. Put another way, we turn from the study of the determination of output in the *short and medium run*—where fluctuations dominate—to the determination of output in the *long run*—where growth dominates. Our goal is to understand what determines growth, why some countries are growing while others are not, and why some countries are rich while many others are still poor.

Section 10-1 discusses a central measurement issue; namely, how to measure the standard of living.

Section 10-2 looks at growth in the United States and other rich countries over the last 50 years.

Section 10-3 takes a broader look, across both time and space.

Section 10-4 gives a primer on growth and introduces the framework that will be developed in the next three chapters.

If you remember one basic message from this chapter, it should be: Over long periods of time, growth dwarfs fluctuations. The (complex) key to sustained growth is technological progress. ▶▶▶

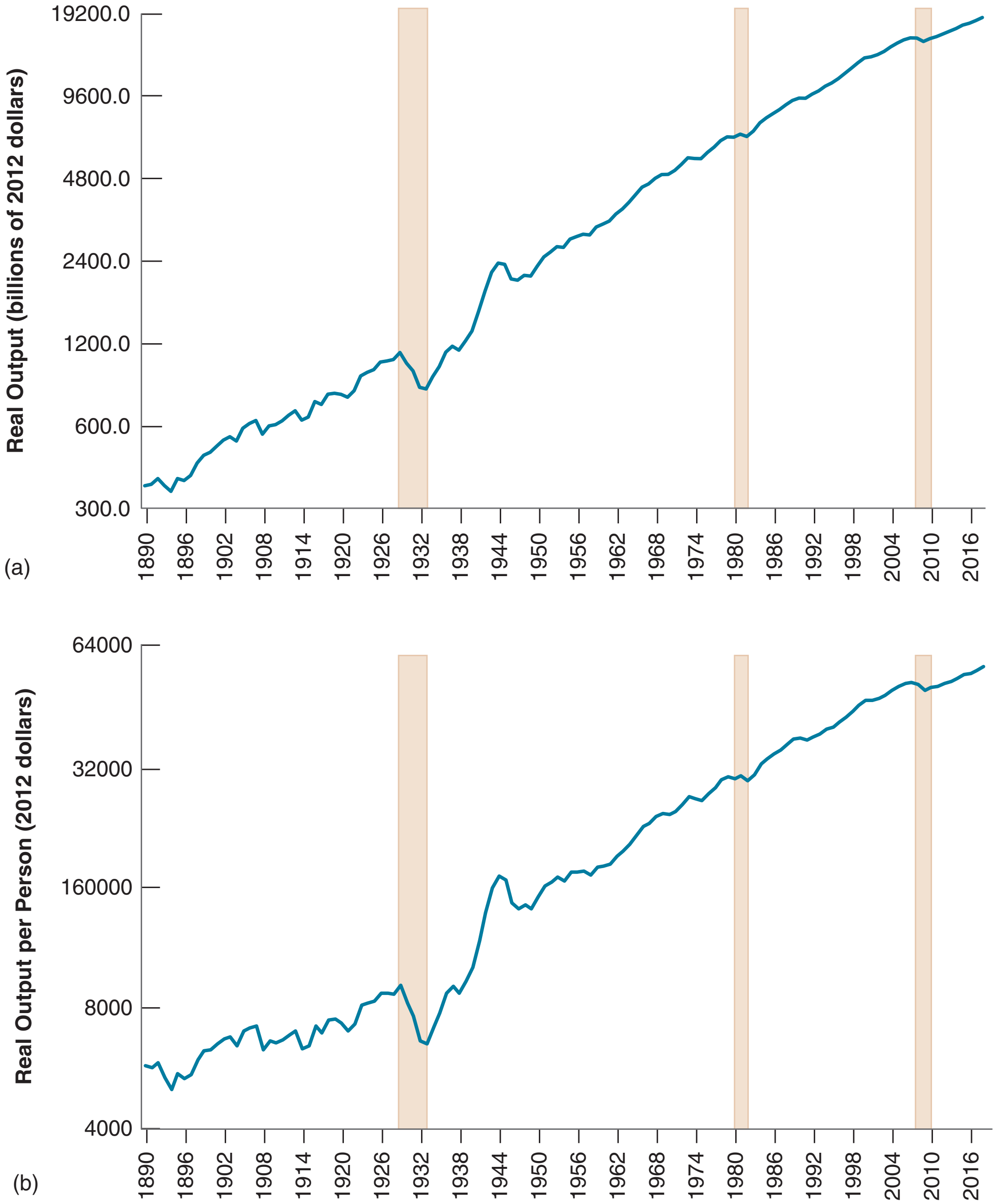


Figure 10-1

US GDP since 1890 and US GDP per Person since 1890

Panel (a) shows the enormous increase in US output since 1890, by a factor of 51. Panel (b) shows that the increase in output is not simply the result of the large increase in US population from 63 million to more than 320 million over this period. Output per person has risen by a factor of 10.

Sources: 1890–1947: *Historical Statistics of the United States*. http://hsus.cambridge.org/HSUSWeb/toc/hsusHome.do. 1948–2014: *National Income and Product Accounts*. Population estimates 1890–2014: Louis Johnston and Samuel H. Williamson, "What Was the US GDP Then?" Measuring Worth, 2015, www.measuringworth.com/datasets/usgdp/. More recent data: FRED: GDPC1, B230RC0Q173SBEA.

10-1 MEASURING THE STANDARD OF LIVING

We care about other dimensions of growth as well, in particular whether the increase in the standard of living benefits some more than others, whether growth comes with more or less inequality. More on this in Chapter 13.

Output per person is also called *output per capita* (*capita* means "head" in Latin). And given that output and income are always equal, it is also called *income per person* or *income per capita*.

The fundamental reason we care about the growth of a country is that we care about the average **standard of living** in the country. Looking across time, we want to know by how much the standard of living has increased. Looking across countries, we want to know how much higher the standard of living is in one country relative to another. Thus, the variable we want to focus on, and compare either over time or across countries, is **output per person**, rather than *output* itself.

A practical problem then arises: How do we compare output per person across countries? Countries use different currencies; thus, output in each country is expressed in terms of its own currency. A natural solution is to use exchange rates. When comparing, say, the

Dana Fradon / The New Yorker Collection/The Cartoon Bank

"It's true, Caesar. Rome is declining, but I expect it to pick up in the next quarter."

output per person of India to the output per person of the United States, we can compute Indian GDP per person in rupees, use the exchange rate to get Indian GDP per person in dollars, and compare it to the US GDP per person in dollars. This simple approach will not do, however, for two reasons.

- First, exchange rates can vary a lot (more on this in Chapters 17 to 20). For example, the dollar increased and then decreased in the 1980s by roughly 50% vis-à-vis the currencies of the trading partners of the United States. But the standard of living in the United States did not increase by 50% and then decrease by 50% compared to the standard of living of its trading partners during the decade. Yet this is the conclusion we would reach if we were to compare GDP per person using exchange rates.
- The second reason goes beyond fluctuations in exchange rates. In 2018, GDP per person in India, using the current exchange rate, was $2,016 compared to $62,517 in the United States. Surely no one could live on $2,016 a year in the United States. But people live on it—admittedly, not very well—in India, where the prices of basic goods (those needed for subsistence) are much lower than in the United States. The level of income of the average person in India, who consumes mostly basic goods, is certainly not 31 times smaller than that of the average person in the United States. This point applies to other countries as well. In general, the lower a country's output per person, the lower the prices of food and basic services in that country.

Recall a similar discussion in Chapter 1 where we looked at output per person in China.

So, when we focus on comparing standards of living, we get more meaningful comparisons by correcting for these two effects—variations in exchange rates and systematic differences in prices across countries. The details of construction are complicated, but the principle is simple. The numbers for GDP—and hence for GDP per person—are constructed using a common set of prices for all countries. Such adjusted real GDP numbers, which you can think of as measures of **purchasing power** across time or across countries, are called **purchasing power parity (PPP)** numbers. Further discussion is given in the Focus Box "The Construction of PPP Numbers."

When one is comparing rich versus poor countries, the differences between PPP numbers and the numbers based on current exchange rates can be large. Return to the

FOCUS

The Construction of PPP Numbers

Consider two countries—let's call them the United States and Russia, although we are not attempting to fit the characteristics of those two countries very closely.

Assume that, in the United States, annual consumption per person equals $20,000 and that people each buy two goods every year: they buy a new car for $10,000 and spend the rest on food. The price of a yearly bundle of food in the United States is $10,000.

Assume that, in Russia, annual consumption per person equals 120,000 rubles and people keep their cars for 15 years. The price of a car is 600,000 rubles, so individuals spend on average 40,000 rubles—600,000 divided by 15—a year on cars. They buy the same yearly bundle of food as their US counterparts, at a price of 80,000 rubles.

Russian and US cars are of identical quality, and so are Russian and US food. (You may dispute the realism of these assumptions. Whether a car in country X is the same as a car in country Y is the type of problem economists face when constructing PPP measures.) The exchange rate is such that one dollar is equal to 60 rubles. What is consumption per person in Russia relative to consumption per person in the United States?

One way to answer this question is by taking consumption per person in Russia and converting it into dollars using the exchange rate. Using this method, Russian consumption per person in dollars is $2,000 (120,000 rubles divided by the exchange rate, 60 rubles to the dollar). According to these numbers, consumption per person in Russia is only 10% of US consumption per person.

Does this answer make sense? It is true that Russians are poorer, but food is much cheaper in Russia. A US consumer spending all of his $20,000 on food would buy 2 bundles of food ($20,000/$10,000). A Russian consumer spending all of his 120,000 rubles on food would buy 1.5 bundles of food (120,000 rubles/80,000 rubles). In terms of food bundles, the difference between US and Russian consumption per person looks much smaller. And given that one-half of consumption in the United States and two-thirds of consumption in Russia go to spending on food, this seems like a relevant computation.

Can we improve on our initial answer? Yes. One way is to use the same set of prices for both countries and then measure the quantities of each good consumed in each country using this common set of prices. Suppose we use US prices. In terms of US prices, annual consumption per person in the United States is obviously still $20,000. What is it in Russia? Every year, the average Russian buys approximately 0.07 car (one car every fifteen years) and one bundle of food. Using US prices—$10,000 for a car and $10,000 for a bundle of food—gives Russian consumption per person as $[(0.07 \times \$10{,}000) + (1 \times \$10{,}000)] = [\$700 + \$10{,}000] = \$10{,}700$. So, using US prices to compute consumption in both countries puts annual Russian consumption per person at $\$10{,}700/\$20{,}000 = 53.5\%$ of annual US consumption per person, a better estimate of relative standards of living than we obtained using our first method (which put the number at only 10%).

This type of computation, namely the construction of variables across countries using a common set of prices, underlies PPP estimates. Rather than using US dollar prices as in our example (why use US rather than Russian or, for that matter, French prices?), these estimates use average prices across countries. These average prices are called *international dollar prices*. Many of the estimates we use in this chapter are the result of an ambitious project known as the Penn World Tables. (Penn stands for the University of Pennsylvania, where the project was initially located.) Led by three economists—Irving Kravis, Robert Summers, and Alan Heston—over more than 40 years, researchers working on the project have constructed PPP series not only for consumption (as we just did in our example), but more generally for GDP and its components, going back to 1950, for most countries in the world. See Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer, "The Next Generation of the Penn World Table," *American Economic Review*, 2015, 105(10): pp. 3150–3182. The PPP data are available for download at www.ggdc.net/pwt

---

comparison between India and the United States. We saw that, at current exchange rates, the ratio of GDP per person in the United States to GDP per person in India was 31.3. Using PPP numbers, the ratio is "only" 8. Although this is still a large difference, it is much smaller than the ratio we obtained using current exchange rates. Differences between PPP numbers and numbers based on current exchange rates are typically smaller when making comparisons among rich countries. For example, using current exchange rates, the ratio of GDP per person in the United States to GDP per person in Germany in 2018 was 128%; based instead on PPP numbers, the ratio was 118%. (More generally, PPP numbers suggest that the United States still has the highest GDP per person among the world's major countries.)

> The bottom line: When comparing standards of living across countries, make sure to use PPP numbers.

> Among advanced economies, only Ireland and Switzerland have a higher PPP GDP per capita than the United States (see the WEO database). Other countries with higher PPP GDP include Kuwait and Qatar. Can you guess why?

Let me end this section with three remarks before we move on and look at growth.

- What matters for people's welfare is their consumption rather than their income. One might therefore want to use *consumption per person* rather than output per person as a measure of the standard of living. (This is indeed what we do in the Focus Box "The Construction of PPP Numbers.") Because the ratio of consumption to output is rather similar across countries, the ranking of countries is roughly the same, whether we use consumption per person or output per person.

- Thinking about the production side, we may be interested in differences in productivity rather than differences in the standard of living across countries. In this case, the right measure is *output per worker*—or, even better, *output per hour worked* if the information about total hours worked is available—rather than output per person. Output per person and output per worker (or per hour) will differ to the extent that the ratio of the number of workers (or hours) to population differs across countries. Most of the difference we saw between output per person in the United States and in Germany comes, for example, from differences in hours worked per person (1,780 hours per year in the United States, compared to 1,360 hours per year in Germany, a difference largely due to a higher proportion of part-time employment in Germany) rather than from differences in productivity. Put another way, German workers are about as productive as their US counterparts, but they work fewer hours, so output per person is lower.
- The reason we ultimately care about the standard of living is presumably that we care about happiness. We may therefore ask the obvious question: Does a higher standard of living lead to greater happiness? The answer is given in the Focus Box "Does Money Buy Happiness?" The answer: a qualified yes.

10-2 GROWTH IN RICH COUNTRIES SINCE 1950

Let's start by looking, in this section, at growth in rich countries since 1950. In the next section, we shall look further back in time and across a wider range of countries.

Table 10-1 shows the evolution of output per person (GDP, measured at PPP prices, divided by population) for France, Japan, the United Kingdom, and the United States since 1950. We have chosen these four countries not only because they are some of the world's major economic powers, but also because what has happened to them is broadly representative of what has happened in other advanced countries over the last half-century or so.

Table 10-1 yields two main conclusions:

- There has been a large increase in output per person.
- There has been a convergence of output per person across countries.

Let's look at each of these points in turn.

Table 10-1 The Evolution of Output per Person in Four Rich Countries since 1950

	Annual Growth Rate Output per Person (%)	Real Output per Person (2011 dollars)		
	1950–2017	1950	2017	2017/1950
France	2.6	7,025	39,461	5.6
Japan	4.1	2,531	40.374	15.9
United Kingdom	2.1	9,354	39,128	4.2
United States	2.0	14,569	54,995	3.8
Average	2.7	8,370	43,490	5.2

The average in the last line is a simple unweighted average.

Source: Penn World Table Version 8.1./Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" forthcoming American Economic Review.

FOCUS

Does Money Buy Happiness?

Does money buy happiness? Or, put more accurately, does higher income per person lead to more happiness? The implicit assumption, when economists assess the performance of an economy by looking at its level of income per person or at its growth rate, is that this is indeed the case. But early examinations of data on the relation between income and self-reported measures of happiness suggested that this assumption may not be right. They yielded what is now known as the **Easterlin paradox** (named for Richard Easterlin, one of the first economists to look systematically at the evidence):

- Looking across countries, happiness in a country appeared to be higher the higher the average level of income per person. The relation, however, appeared to hold only in relatively poor countries. Looking at rich countries, say the OECD countries, there appeared to be little relation between average income per person and happiness.
- Looking at individual countries over time, average happiness in rich countries did not seem to increase much, if at all, with income. (There were no reliable data for poor countries.) In other words, in rich countries, higher income per person did not appear to increase happiness.
- Looking across people within a given country, happiness appeared to be strongly correlated with income. Rich people were consistently happier than poor people. This was true in both poor and rich countries.

The first two findings suggested that, once basic needs are satisfied, higher income per person did not increase happiness. The third fact suggested that what is important was not the absolute level of income but the level of income relative to others.

If these findings were correct, it would have major implications for the way we think about the world and about economic policies. In rich countries, policies aimed at increasing income per person might be misdirected because what matters is the distribution of income rather than its average level. Globalization and the diffusion of information, to the extent that they make people in poor countries compare themselves not to rich people in the same country but to people in richer countries, might actually decrease rather than increase happiness. So, as you can guess, these findings have led to an intense debate and further research. As new data have become available, better evidence has accumulated. The state of knowledge and the remaining controversies are analyzed in a recent article by Betsey Stevenson and Justin Wolfers. Their conclusions are well summarized in Figure 1.

The figure contains a lot of information. Let's go through it step by step.

The horizontal axis measures PPP GDP per person for 131 countries. The scale is a logarithmic scale, so a given size interval represents a given percentage increase in GDP per person. The vertical axis measures average life satisfaction in each country. The source for this variable is a 2006 Gallup

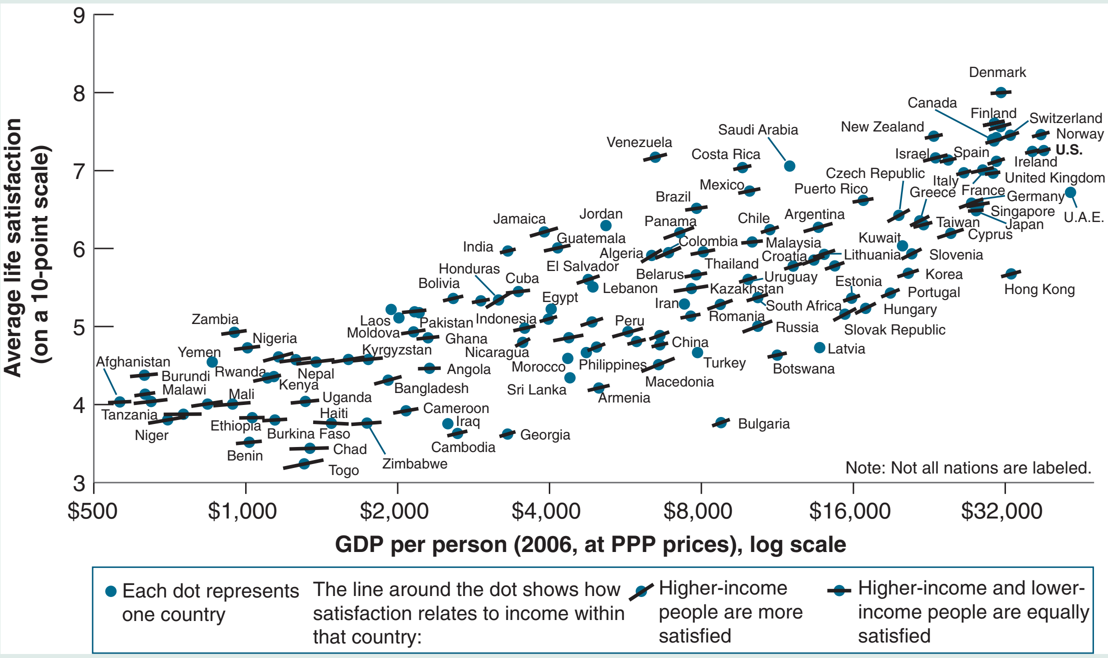


Figure 1

Life Satisfaction and Income per Person

Source: Betsey Stevenson and Justin Wolfers, Wharton School at the University of Pennsylvania.Used with Permission.

World Poll survey, which asked about a thousand individuals in each country the following question:

> "Here is a ladder representing the 'ladder of life.' Let's suppose the top of the ladder represents the best possible life for you; and the bottom, the worst possible life for you. On which step of the ladder do you feel you personally stand at the present time?"

The ladder went from 0 to 10. The variable measured on the vertical axis is the average of the individual answers in each country.

Focus first on the dots representing each country, ignoring for the moment the lines that cross each dot. The visual impression is clear. There is a strong relation across countries between average income and average happiness. The index is around 4 in the poorest countries, around 8 in the richest. And, more importantly in view of the Easterlin paradox, this relation appears to hold for both poor and rich countries; if anything, life satisfaction appears to increase faster, as GDP per person increases, in rich than in poor countries.

Focus now on the lines through each dot. The slope of each line reflects the estimated relation between life satisfaction and income *within* each country. Note first that all the lines slope upward. This confirms the third leg of the Easterlin paradox. In each country, rich people are happier than poor people. Note also that the slopes of most of these lines are roughly similar to the slope of the relation across countries. This goes against the Easterlin paradox. Individual happiness increases with income, whether this is because the country is getting richer or because the individual becomes relatively richer within the country.

Stevenson and Wolfers draw a strong conclusion from their findings. Although individual happiness depends on much more than income, it definitely increases with income. While the idea that there is some critical level of income beyond which income no longer impacts well-being is intuitively appealing, it is at odds with the data. Thus, it is not a crime for economists to focus first on levels and growth rates of GDP per person.

So, is the debate over? The answer is no. Even if we accept this interpretation of the evidence, clearly, many other aspects of the economy matter for welfare, income distribution surely being one of them. And not everyone is convinced by the evidence. In particular, the evidence on the relation between happiness and income per person over time within a country is not as clear as the evidence across countries or across individuals presented in Figure 1.

Given the importance of the question, the debate will continue. One aspect that has become clear, for example, from the work of Nobel Prize winners Angus Deaton and Daniel Kahneman is that, when thinking about "happiness," it is important to distinguish between two ways in which a person may assess her or his well-being. The first one is *emotional well-being*—the frequency and intensity of experiences such as joy, stress, sadness, anger, and affection that make one's life pleasant or unpleasant. Emotional well-being appears to rise with income because low income exacerbates the emotional pain associated with misfortunes such as divorce, ill health, and being alone. But only up to a threshold; there is no further progress beyond an annual income of about $75,000 (the experiment was run in 2009). The second is *life satisfaction*, a person's assessment of her or his life when they think about it. Life satisfaction appears more closely correlated with income. Deaton and Kahneman conclude that high income buys life satisfaction but not necessarily happiness. If measures of well-being are to be used to guide policy, their findings raise the question of whether life satisfaction or emotional well-being is better suited to these aims.[1]

The Large Increase in the Standard of Living since 1950

Look at the column on the far right of Table 1. Output per person has increased by a factor of 3.8 since 1950 in the United States, by a factor of 5.6 in France, and by a factor of 15.9 in Japan. These numbers show what is sometimes called the **force of compounding**. In a different context, you probably have heard that saving even a little while you are young will build to a large amount by the time you retire. For example, if the interest rate is 4.1% a year, an investment of $1, with the proceeds reinvested every year, will grow to about $15.9 65 years later. The same logic applies to growth rates. The average annual growth rate per person in Japan over the period 1950 to 2017 (67 years) was equal to 4.1%. This high growth rate led to a 15.9-fold increase in real output per person in Japan over the period.

Most of the increase in Japan took place before 1990. Since then, Japan has been in a prolonged economic slump, with much lower growth.

Clearly, a better understanding of growth, if it leads to the design of policies that stimulate growth, can have a large effect on the standard of living. Suppose we could find a policy measure that permanently increased the growth rate by 1% per year. This would lead, after 40 years, to a standard of living 48% higher than it would have been without the policy—a substantial difference.

$1.01^{40} - 1 = 1.48 - 1 = 48\%$

Unfortunately, policy measures with such magical results have proven difficult to discover!

[1]Sources: *Betsey Stevenson and Justin Wolfers, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox,"* Brookings Papers on Economic Activity, *2008: pp. 1–87 and "Subjective Well-Being and Income: Is There Any Evidence of Satiation?"* American Economic Review: Papers & Proceedings, *2013, 103(3): pp. 598–604; Daniel Kahneman and Angus Deaton, "High Income Improves Evaluation of Life But Not Emotional Well-Being,"* Proceedings of the National Academy of Sciences, *2010, 107(38): pp. 16,489–16,493. For a view closer to the Easterlin paradox and a fascinating discussion of policy implications, read Richard Layard,* Happiness: Lessons from a New Science *(2005, Penguin Books).*

The Convergence of Output per Person

The second and third columns of Table 10-1 show that the levels of output per person have converged over time. The numbers for output per person are much closer in 2017 than they were in 1950. Put another way, those countries that were behind have grown faster, reducing the gap between them and the United States.

In 1950, output per person in the United States was roughly twice the level of output per person in France and more than five times the level of output per person in Japan. From the perspective of people in Europe or Japan, the United States was the land of plenty, where everything was bigger and better.

As a child in France in the 1950s, I thought of the United States as the land of skyscrapers, big automobiles, and Hollywood movies.

Today these perceptions have faded, and the numbers explain why. Using PPP numbers, US output per person is still the highest, but in 2017 it was only 39% above average output per person in the other three countries, a much smaller difference than in the 1950s.

This **convergence** of levels of output per person across countries is not specific to these four countries. It extends to a set of OECD countries. This is shown in Figure 10-2, which plots the average annual growth rate of output per person since 1950 against the initial level of output per person in 1950 for the current OECD countries. There is a clear negative relation between the initial level of output per person and the growth rate since 1950: Countries that were behind in 1950 have typically grown faster. The relation is not perfect. Turkey, which had roughly the same low level of output per person as Japan in 1950, has had a growth rate equal to only about one-half that of Japan. But the relation is clearly there.

The figure includes only the OECD members for which we have a reliable estimate of the level of output per person in 1950.

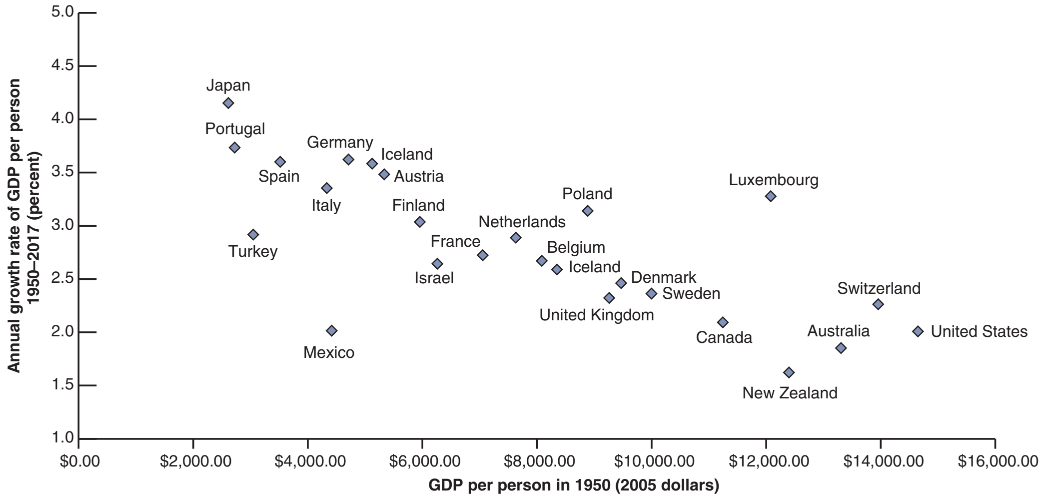
Some economists have pointed to a problem in graphs like Figure 10-2. By looking at the subset of countries that are members of the OECD today, we are in effect looking at a club of economic winners. OECD membership is not officially based on economic success, but economic success is an important determinant of membership. But when you look at a club whose membership is based on economic success, you will find that those who came from behind had the fastest growth. This is precisely why they made it into the club! So the finding of convergence could come in part from the way we selected the countries in the first place.

A better way of looking at convergence is to define the set of countries we look at not on the basis of where they are today—as we did in Figure 10-2 by taking today's OECD members—but on the basis of where they were in, say, 1950. For example, we can look at all countries that had an output per person of at least one-fourth of US output per

Figure 10-2

Growth Rate of GDP per Person since 1950 versus GDP per Person in 1950, OECD Countries

Countries with lower levels of output per person in 1950 have typically grown faster.



Source: Penn World Tables Version 9. /Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150–3182, available for download at www.ggdc.net/pwt.

person in 1950, and then look for convergence within that group. It turns out that most of the countries in that group have indeed converged, and therefore convergence is not solely an OECD phenomenon. However, a few countries—Uruguay, Venezuela among them—have not converged. In 1950, those two countries had roughly the same output per person as France. In 2017, they had fallen far behind; Uruguay's output per person stood at one-third of the French level, Venezuela's at one-fifth (and since then, although we do not have PPP-adjusted numbers, the evidence is that Venezuela's output has collapsed further).

10-3 A BROADER LOOK ACROSS TIME AND SPACE

In the previous section, we focused on growth since 1950 in rich countries. Let's now put this in context by looking at the evidence over a much longer time span and a wider set of countries.

Looking across Two Millennia

Has output per person in the currently rich economies always grown at rates similar to the growth rates shown in Table 10-1? The answer is no. Estimates of growth are clearly harder to construct as we look further back in time. But there is agreement among economic historians about the main evolutions over the last 2,000 years.

From the end of the Roman Empire in 476 to roughly 1500, there was essentially no growth of output per person in Europe. Most people worked in agriculture in which there was little technological progress. Because agriculture's share of output was so large, inventions with applications outside agriculture could contribute little to overall production and output. Although there was some output growth, a roughly proportional increase in population led to roughly constant output per person.

This period of stagnation of output per person is often called the *Malthusian era*. Thomas Robert Malthus, an English economist at the end of the 18th century, argued that this proportional increase in output and population was not a coincidence. Any increase in output, he argued, would lead to a decrease in mortality, leading to an increase in population until output per person was back to its initial level. Europe was in a **Malthusian trap**, unable to increase its output per person.

Eventually, Europe was able to escape this trap. From about 1500 to 1700, growth of output per person turned positive, but it was still small—only around 0.1% per year. It then increased to just 0.2% per year from 1700 to 1820. Starting with the Industrial Revolution, growth rates increased, but from 1820 to 1950 the growth rate of output per person in the United States was still only 1.5% per year. On the scale of human history, therefore, sustained growth of output per person—especially the high growth rates we have seen since 1950—is definitely a recent phenomenon. The Focus Box "The Reality of Growth: A US Workingman's Budget in 1851" gives a vivid sense of the progress over the last 150 years.

Looking across Countries

We have seen how output per person has converged among OECD countries. But what about the other countries? Are the poorest countries also growing faster? Are they converging toward the United States, even if they are still far behind?

The answer is given in Figure 10-3, which plots the average annual growth rate of output per person since 1960 against output per person for the year 1960, for the 85 countries for which we have data.

The numbers for 1950 are missing for too many countries to use 1950 as the initial year, as we did in Figure 10-2.

FOCUS

The Reality of Growth: A US Workingman's Budget in 1851

Data on GDP per capita do not fully convey the reality of growth and the accompanying increase in the standard of living. An examination of a "workingman's budget" in 1851 Philadelphia gives a much better sense of the improvement.

Table 1 Annual Workingman's Budget, Philadelphia, 1851, in current dollars

Item of expenditure	Amount ($)	(% of total)
Butcher's meat (2 lb a day)	72.80	13.5
Flour (6½ lb a year)	32.50	6.0
Butter (2 lb a week)	32.50	6.0
Potatoes (2 pk a week)	26.00	4.8
Sugar (4 lb a week)	16.64	3.0
Coffee and tea	13.00	2.4
Milk	7.28	1.3
Salt, pepper, vinegar, starch, soap, yeast, cheese, eggs	20.80	3.8
Total expenditures for food	221.52	41.1
Rent	156.00	28.9
Coal (3 tons a year)	15.00	2.7
Charcoal, chips, matches	5.00	0.9
Candles and oil	7.28	1.4
Household articles (wear, tear, and breakage)	13.00	2.4
Bedclothes and bedding	10.40	1.9
Wearing apparel	104.00	19.3
Newspapers	6.24	1.1
Total expenditures other than food	316.92	58.9

Note how much a worker spent on food, 41% of total expenditures. Today's corresponding share—as reflected in the composition of the consumption basket used to compute the Consumer Price Index—is only 15.2%. And food at home—as opposed to food in restaurants—accounts for only 9.4% of total consumption today. But perhaps more revealing is the composition of food consumption. Compare the food in the table to the richness and diversity of the food we eat today.

Sources: William Baumol, Sue Ann Batey Blackman, and Edward N. Wolff, *Productivity and American Leadership: The Long View* (1989, MIT Press), Chapter 3, Table 3.2. The composition of expenditures today comes from Appendix 9, Chapter 6, in the BLS *Handbook of Methods*.

The striking feature of Figure 10-3 is that there is no clear pattern. It is not the case that, in general, countries that were behind in 1960 have grown faster. Some have, but many have clearly not.

The cloud of points in Figure 10-3 hides, however, a number of interesting patterns that appear when we put countries into different groups, denoted by different symbols.

Figure 10-3

Growth Rate of GDP per Person since 1960, versus GDP per Person in 1960, 85 Countries

There is no clear relation between the growth rate of output since 1960 and the level of output per person in 1960.

Source: Penn World Tables Version 9. /Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150–3182, available for download at www.ggdc.net/pwt.

The diamonds represent OECD countries, the squares African countries, and the triangles Asian countries. Looking at patterns by groups yields three main conclusions.

1. The picture for the OECD countries (the rich countries) is much the same as in Figure 10-2, which looked at a slightly longer period (from 1950 onward, rather than from 1960). Nearly all start at high levels of output per person (say, at least one-third of the US level in 1960) and there is clear evidence of convergence.

2. Convergence is also visible for many Asian countries: Most of the countries with high growth rates over the period are in Asia. Japan was the first country to take off. A decade later, in the 1960s, four countries—Singapore, Taiwan, Hong Kong, and South Korea, a group of countries sometimes called the **four tigers**—started catching up as well. In 1960, their average output per person was about 15% of that of the United States; by 2017, the ratio had increased to 85%. More recently, the major story has been China—both because of its very high growth rates and because of its sheer size. Over the period 1960–2017 average growth of output per person per year in China has been 4.5%. But, because it started low, its output per person is still only about one-fourth of that of the United States.

3. The picture is different, however, for African countries. Most African countries (represented by squares) were very poor in 1960, and most have not done well over the period. Many have suffered from either internal or external conflicts. Four of them have had negative growth of output per person—an absolute decline in their standard of living between 1960 and 2017. For example, growth averaged—1.1% in the Central African Republic and, as a result, output per person in 2017 was only 52% of its level in 1960. Hope for Africa, however, comes from more recent numbers: Growth of output per person in sub-Saharan Africa has been close to 3% since 2000.

Paradoxically, the two fastest growing countries in Figure 10-3 are Botswana and Equatorial Guinea in Africa. In both cases, however, high growth reflects primarily favorable natural resources—diamonds in Botswana and oil in Guinea.

Looking further back in time, the following picture emerges. For much of the first millennium, and until the 15th century, China probably had the world's highest level of output per person. For a couple of centuries, leadership then moved to the cities of northern Italy. But until the 19th century, differences across countries were typically much

The distinction between *growth theory* and *development economics* is fuzzy. A rough distinction: Growth theory takes many of the institutions of a country (e.g., its legal system and its form of government) as given. Development economics asks what institutions are needed to sustain steady growth, and how they can be put in place.

smaller than they are today. Starting in the 19th century, a number of countries, first in Western Europe then in North and South America, started growing faster than others. Since then, a number of other countries, most notably in Asia, have started growing fast and are converging. Many others, mainly in Africa, are not.

Our main focus, in this and the next chapter, will be on growth in rich and emerging countries. We shall not take on some of the wider challenges raised by the facts we have just seen, such as why growth of output per person started in earnest in the 19th century or why Africa has remained so poor. Doing so would take us too far into economic history and development economics. But these facts put in perspective the two basic facts we discussed when looking at the OECD: Neither growth nor convergence is a historical necessity.

10-4 THINKING ABOUT GROWTH: A PRIMER

To think about growth, economists use a framework developed by Robert Solow, of the Massachusetts Institute of Technology (MIT), in the late 1950s. The framework has proven sturdy and useful, and we will use it here. This section provides an introduction; Chapters 11 and 12 will provide a more detailed analysis, first of the role of capital accumulation and then of the role of technological progress in the process of growth.

Solow was awarded the Nobel Prize in 1987 for his work on growth.

The Aggregate Production Function

The starting point for any theory of growth must be an **aggregate production function**, which is a specification of the relation between aggregate output and the inputs in production.

The aggregate production function we introduced in Chapter 7 to study the determination of output in the short run and the medium run took a particularly simple form. Output was simply proportional to the amount of labor used by firms; more specifically, proportional to the number of workers employed by firms (equation (7.2)). So long as our focus was on fluctuations in output and employment, the assumption was acceptable. But now that our focus has shifted to growth, this assumption will no longer do. It implies that output per worker is constant, ruling out growth of output per worker altogether. It is time to relax it. From now on, we will assume that there are two inputs—capital and labor—and that the relation between aggregate output and the two inputs is given by:

$$Y = F(K, N) \tag{10.1}$$

As before, Y is aggregate output. K is capital, the sum of all the machines, plants, and office buildings in the economy. N is labor, the number of workers in the economy. The function F, which tells us how much output is produced for given quantities of capital and labor, is the *aggregate production function.*

The aggregate production function is $Y = F(K, N)$. Aggregate output (Y) depends on the aggregate capital stock (K) and aggregate employment (N).

This way of thinking about aggregate production is an improvement on our treatment in Chapter 7. But it is still a dramatic simplification of reality. Machines and office buildings play different roles in production and should be treated as separate inputs. Workers with doctorate degrees are different from high school dropouts; yet, by constructing the labor input as simply the *number* of workers in the economy, we treat all workers as identical. We will relax some of these simplifications later. For the time being, equation (10.1), which emphasizes the role of both labor and capital in production, will do.

The next step must be to think about where the aggregate production function F, which relates output to the two inputs, comes from. In other words, what determines how much output can be produced for given quantities of capital and labor? The answer: the **state of technology**. A country with more advanced technology will produce more output from the same quantities of capital and labor than will an economy with primitive technology.

The function F depends on the state of technology. The higher the state of technology, the higher $F(K, N)$ for a given K and a given N.

Returns to Scale and Returns to Factors

Now that we have introduced the aggregate production function, the next question is: What restrictions can we reasonably impose on this function?

Consider first a thought experiment in which we double both the number of workers and the amount of capital in the economy. What do you expect will happen to output? A reasonable answer is that output will double as well. In effect, we have cloned the original economy, and the clone economy can produce output in the same way as the original economy. This property is called **constant returns to scale**. If the scale of operation is doubled—that is, if the quantities of capital and labor are doubled—then output will also double.

$$F(2K, 2N) = 2Y$$

Or, more generally, for any number x (this will be useful later)

Constant returns to scale: $F(xK, xN) = xY$.

$$F(xK, xN) = xY \tag{10.2}$$

We have just looked at what happens to production when *both* capital and labor are increased. Let's now ask a different question. What should we expect to happen if *only one* of the two inputs in the economy—say, capital—is increased?

Output will increase. That part is clear. But the same increase in capital will lead to smaller and smaller increases in output. In other words, if there is little capital to start with, a little more capital will help a lot. If there is a lot of capital to start with, a little more capital may make little difference. Why? Think, for example, of a secretarial pool, composed of a given number of secretaries. Think of capital as computers. The introduction of the first computer will substantially increase the pool's production because some of the more time-consuming tasks can now be done automatically by the computer. As the number of computers increases and more secretaries in the pool get their own computers, production will further increase, although by less per additional computer than was the case when the first one was introduced. Once every secretary has a computer, increasing the number of computers further is unlikely to increase production much, if at all. Additional computers might simply remain unused and left in their shipping boxes and lead to no increase in output.

I know, the example is a bit old-fashioned, but it is the most transparent example I could think of.

Output here is secretarial services. The two inputs are secretaries and computers. The production function relates secretarial services to the number of secretaries and the number of computers.

We shall refer to the property that increases in capital lead to smaller and smaller increases in output as **decreasing returns to capital** (a property that will be familiar to those who have taken a course in microeconomics).

A similar argument applies to the other input, labor. Increases in labor, given capital, lead to smaller and smaller increases in output. (Return to our example, and think of what happens as you increase the number of secretaries for a given number of computers.) There are **decreasing returns to labor** as well.

Even under constant returns to scale, there are decreasing returns to each factor, keeping the other factor constant. There are decreasing returns to capital. Given labor, increases in capital lead to smaller and smaller increases in output. There are decreasing returns to labor. Given capital, increases in labor lead to smaller and smaller increases in output.

Output per Worker and Capital per Worker

The production function we have written, together with the assumption of constant returns to scale, implies that there is a simple relation between *output per worker* and *capital per worker*.

To see this, set $x = 1/N$ in equation (10.2), so that

$$\frac{Y}{N} = F\left(\frac{K}{N}, \frac{N}{N}\right) = F\left(\frac{K}{N}, 1\right) \tag{10.3}$$

Make sure you understand what is behind the algebra. Suppose that capital and the number of workers both double. What happens to output per worker?

Y/N is output per worker, and K/N is capital per worker. So equation (10.3) tells us that the amount of output per worker depends on the amount of capital per worker. This relation between output per worker and capital per worker will play a central role in what follows, so let's look at it more closely.

This relation is drawn in Figure 10-4. Output per worker (Y/N) is measured on the vertical axis, and capital per worker (K/N) is measured on the horizontal axis. The relation between the two is given by the upward-sloping curve. As capital per worker increases, so does output per worker. Note that the curve is drawn so that increases in capital lead to smaller and smaller increases in output. This follows from the property that there are *decreasing returns to capital*: At point A, where capital per worker is low, an increase in capital per worker, represented by the horizontal distance AB, leads to an increase in output per worker equal to the vertical distance $A'B'$. At point C, where capital per worker is larger, the same increase in capital per worker, represented by the horizontal distance CD (where the distance CD is equal to the distance AB), leads to a much smaller increase in output per worker, only the distance $C'D'$. This is just like our secretarial pool example, in which additional computers had less and less impact on total output.

Increases in capital per worker lead to smaller and smaller increases in output per worker as the level of capital per worker increases.

The Sources of Growth

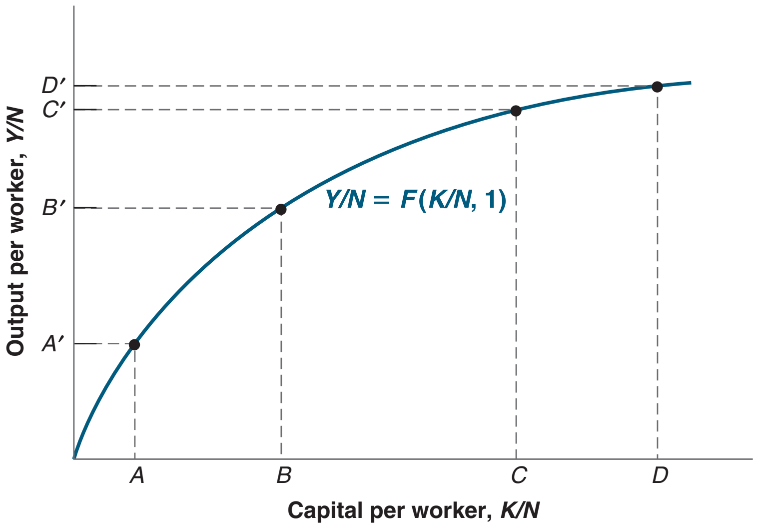
We are now ready to return to our basic question. Where does growth come from? Why does output per worker—or output per person, if we assume the ratio of workers to the population as a whole remains constant—go up over time? Equation (10.3) gives a first answer:

- Increases in output per worker (Y/N) can come from increases in capital per worker (K/N). This is the relation we just looked at in Figure 10-4. As (K/N) increases—that is, as we move to the right on the horizontal axis—(Y/N) increases.
- Or they can come from improvements in the state of technology that shift the production function, F, and lead to more output per worker *given* capital per worker.

Figure 10-4

Output and Capital per Worker

Increases in capital per worker lead to smaller and smaller increases in output per worker.



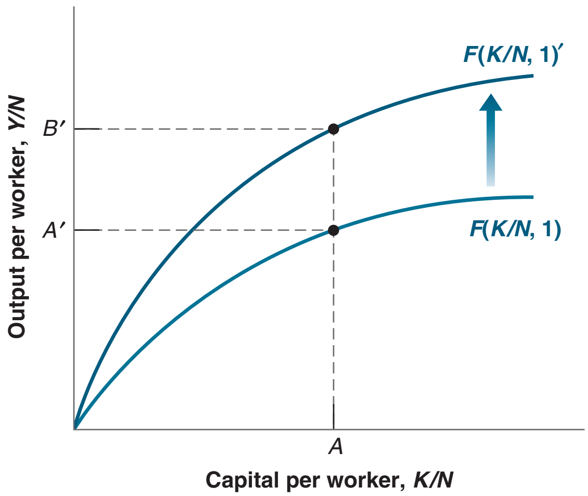


Figure 10-5

The Effects of an Improvement in the State of Technology

An improvement in technology shifts the production function up, leading to an increase in output per worker for a given level of capital per worker.

This is shown in Figure 10-5. An improvement in the state of technology shifts the production function up, from $F(K/N, 1)$ to $F(K/N, 1)'$. For a given level of capital per worker, the improvement in technology leads to an increase in output per worker. For example, for the level of capital per worker corresponding to point A, output per worker increases from A' to B'. (To go back to our secretarial pool example, a reallocation of tasks within the pool may lead to a better division of labor and an increase in the output per secretary.)

Increases in capital per worker: Movements along the production function.

Improvements in the state of technology: Shifts (up) of the production function.

Hence, we can think of growth as coming from **capital accumulation** and from **technological progress**—the improvement in the state of technology. We will see, however, that these two factors play different roles in the growth process.

- Capital accumulation *by itself* cannot sustain growth. A formal argument will have to wait until Chapter 11. But you can already see the intuition behind this in Figure 10-5. Because of decreasing returns to capital, sustaining a steady increase in output per worker will require larger and larger increases in the level of capital per worker. At some stage, the economy will be unwilling or unable to save and invest enough to further increase capital. At that point, output per worker will stop growing.

  Does this mean that an economy's **saving rate** (the proportion of income that is saved) is irrelevant? No. It is true that a higher saving rate cannot permanently increase the *growth rate* of output. But a higher saving rate can sustain a higher *level* of output. Let me state this in a slightly different way. Take two economies that differ only in their saving rates. The two economies will eventually grow at the same rate, but at any point in time, the economy with the higher saving rate will have a higher level of output per person than the other. How this happens, how much the saving rate affects the level of output, and whether a country like the United States (which has a low saving rate) should try to increase its saving rate will be discussed in Chapter 11.
- Sustained growth requires sustained technological progress. This really follows from the previous proposition. Given that the two factors that can lead to an increase in output are capital accumulation and technological progress, if capital accumulation cannot sustain growth forever, then technological progress must be the key to growth. And it is. We will see in Chapter 12 that an economy's rate of growth of output per person is eventually determined by its rate of technological progress.

This is important. It means that in the long run, an economy that sustains a higher rate of technological progress will eventually overtake all other economies. This, of course, raises the next question. What determines the rate of technological progress? One can think of the *state of technology* as the list of blueprints defining both the range of products that can be produced in the economy and the techniques available to produce them. But, in fact, how efficiently things are produced depends not only on the list of blueprints but also on the way the economy is organized—from the internal organization of firms, to the system of laws and the quality of their enforcement, to the political system, and so on. We shall review the evidence in Chapter 12.

SUMMARY

- Over long periods, fluctuations in output are dwarfed by growth, which is the steady increase of aggregate output over time.
- Looking at growth in four rich countries (France, Japan, the United Kingdom, and the United States) since 1950, two main facts emerge.
    1. All four countries have experienced strong growth and a large increase in the standard of living. Growth from 1950 to 2017 increased real output per person by a factor of 3.8 in the United States and by a factor of 15.9 in Japan.
    2. The levels of output per person across the four countries have converged over time. Put another way, the countries that were behind have grown faster, reducing the gap between them and the current leader, the United States.
- Looking at the evidence across a broader set of countries and a longer period, the following facts emerge.
    1. On the scale of human history, sustained output growth is a recent phenomenon.
    2. The convergence of levels of output per person is not a worldwide phenomenon. Many Asian countries are rapidly catching up, while many African countries have both low levels of output per person and low growth rates.
- To think about growth, economists start from an aggregate production function relating aggregate output to two factors of production: capital and labor. How much output is produced given these inputs depends on the state of technology.
- Under the assumption of constant returns, the aggregate production function implies that increases in output per worker can come either from increases in capital per worker or from improvements in the state of technology.
- Capital accumulation by itself cannot permanently sustain growth of output per person. Nevertheless, how much a country saves is important because the saving rate determines the *level* of output per person, if not its growth rate.
- Sustained growth of output per person is ultimately due to technological progress. Perhaps the most important question in growth theory is what the determinants of technological progress are.

KEY TERMS

growth, 201
logarithmic scale, 201
standard of living, 202
output per person, 202
purchasing power, 203
purchasing power parity (PPP), 203
Easterlin paradox, 206
force of compounding, 207
convergence, 208
Malthusian trap, 209
four tigers, 211
aggregate production function, 212
state of technology, 213
constant returns to scale, 213
decreasing returns to capital, 213
decreasing returns to labor, 213
capital accumulation, 215
technological progress, 215
saving rate, 215

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. On a logarithmic scale, a variable that increases at 5% per year will move along an upward-sloping line with a slope of 0.05.
b. The price of food is higher in poor countries than it is in rich countries.
c. Evidence suggests that happiness in rich countries increases with output per person.
d. In virtually all the countries of the world, output per person is converging to the level of output per person in the United States.
e. For about 1,000 years after the fall of the Roman Empire, there was essentially no growth in output per person in Europe because any increase in output led to a proportional increase in population.
f. Capital accumulation does not affect the level of output in the long run, only technological progress does.
g. The aggregate production function is a relation between output on one hand and labor and capital on the other.

2. Assume that the average consumer in Mexico and the average consumer in the United States buy the quantities and pay the prices indicated in the following table:

	Food		Transportation Services	
	Price	Quantity	Price	Quantity
Mexico	5 pesos	400	20 pesos	200
United States	$1	1,000	$2	2,000

a. Compute US consumption per person in dollars.
b. Compute Mexican consumption per person in pesos.
c. Suppose that 1 dollar is worth 10 pesos. Compute Mexico's consumption per person in dollars. This method of comparing consumption in Mexico to consumption in the United States uses market exchange rates.
d. Using the purchasing power parity method and US prices, compute Mexican consumption per person in dollars.
e. Under each method, how much lower is the standard of living in Mexico than in the United States? Does the choice of method make a difference?

3. Consider the production function

$$Y = \sqrt{K}\,\sqrt{N}$$

a. Compute output when $K = 49$ and $N = 81$.
b. If both capital and labor double, what happens to output?
c. Is this production function characterized by constant returns to scale? Explain.
d. Write this production function as a relation between output per worker and capital per worker.
e. Let $K/N = 4$. What is Y/N? Now double K/N to 8. Does Y/N double as a result?
f. Does the relation between output per worker and capital per worker exhibit constant returns to scale?
g. Is your answer to (f) the same as your answer to (c)? Why or why not?
h. Plot the relation between output per worker and capital per worker. Does it have the same general shape as the relation in Figure 10-4? Explain.

DIG DEEPER

4. The growth rates of capital and output

Consider the production function given in Problem 3. Assume that N is constant and equal to 1. Note that if $z = x^a$, then $g_z \approx a\, g_x$, where g_z and g_x are the growth rates of z and x (See Appendix 2 at the end of the book).

a. Given the growth approximation here, derive the relation between the growth rate of output and the growth rate of capital.
b. Suppose we want to achieve output growth equal to 2% per year. What is the required rate of growth of capital?
c. In part b, what happens to the ratio of capital to output over time?
d. Is it possible to sustain output growth of 2% forever in this economy? Why or why not?

5. Between 1950 and 1973, France, Germany, and Japan all experienced growth rates that were at least two percentage points higher than those in the United States. Yet the most important technological advances of that period were made in the United States. How can this be?

EXPLORE FURTHER

Questions 6, 7, and 8 all require you find the data in the Penn World Tables that measures real GDP per capita using purchasing power parity exchange rates as described in the text. As of the time of writing, version 9 is the latest version of this data. The variable rgdpo is total real output in millions of PPP 2011 US dollars. The variable pop is population in millions. You will have to isolate these variables in the spreadsheet and construct real GDP per person for each country in each year. In manipulating the data in a very large spreadsheet, you will be able to answer Questions 6, 7, and 8.

6. Convergence between Japan and the United States since 1960

Using the data in the Penn World Tables, find real GDP per capita in the United States and Japan in 1960, 1990, and the most recent year available.

a. Compute the average annual growth rates of GDP per person for the United States and Japan for two time periods: 1960 to 1990, and 1990 to the most recent year available. Did the level of real output per person in Japan tend to converge to the level of real output per person in the United States in both of these periods? Explain.
b. Suppose that in every year since 1990, Japan and the United States had each continued to have their average annual growth rates for the period 1960 to 1990. How would real GDP per person compare in Japan and the United States today?

c. What actually happened to growth in real GDP per capita in Japan and the United States from 1990 to the most recent year available?

7. Convergence in two sets of countries

Consider three rich countries: France, Belgium, and Italy, and four poor countries, Ethiopia, Kenya, Nigeria, and Uganda. Define for each country the ratio of its real GDP per person to that of the United States in 1970 and in the latest year available (2014 in version 9 of the Penn World Tables) so that this ratio will be equal to 1 for the United States for all years.

a. Calculate these ratios for France, Belgium, and Italy in 1970 and 2014 (or the latest year) over the period for which you have data. Does your data support the notion of convergence among France, Belgium, and Italy with the United States?
b. Calculate these ratios for Ethiopia, Kenya, Nigeria, and Uganda in 1970 and 2014. Does this data support the notion of convergence among Ethiopia, Kenya, Nigeria, and Uganda with the United States?

8. Growth successes and failures

Using the Penn World Tables, find the data on real GDP per person (chained series) for 1970 for all available countries. Do the same for a recent year of data where the data are available for most countries (it takes more time to produce this measure in some countries than in others).

a. Rank the countries according to GDP per person in 1970. List the countries with the 10 highest levels of GDP per person in 1970. Are there any surprises?
b. Carry out the analysis in part (a) for the most recent year for which you collected data. Has the composition of the 10 richest countries changed since 1970?
c. Use all the countries for which there are data in both 1970 and your latest year. Which 10 countries have the highest rate of growth of real GDP per capita after 1970?
d. Use all the countries for which there are data in both 1970 and your latest year. Which 10 countries have the lowest rate of growth of real GDP per capita after 1970?
e. Do a brief Internet search on either the country from part c with the greatest increase in GDP per capita or the country from part d with the smallest increase. Can you guess any reasons for the economic success, or lack of it, for this country?

9. Your institution may have a subscription to The Economist news magazine, or you may be able to find this graphic on the web. The March 23, 2019 issue, in a section entitled "Graphic Detail: Happiness Economics," makes the same point as Figure 1.

a. Does happiness rise with the level of real GDP per person?
b. Contrast the growth and happiness experience of China and India over the period from 2005–08 to 2015–18?

FURTHER READINGS

- Brad deLong has a number of fascinating articles on growth (http://web.efzg.hr/dok/MGR/vcavrak//Berkeley%20Faculty%20Lunch%20Talk.pdf). Read in particular "Berkeley Faculty Lunch Talk: Main Themes of Twentieth Century Economic History," which covers many of the themes of this chapter.
- A broad presentation of facts about growth is given by Angus Maddison in *The World Economy: A Millenium Perspective* (2001). The associated site, www.theworldeconomy.org, has a large number of facts and data on growth over the last two millenia.
- Chapter 3 in *Productivity and American Leadership*, by William Baumol, Sue Anne Batey Blackman, and Edward Wolff (1989), gives a vivid description of how life has been transformed by growth in the United States since the mid-1880s.
- The site https://ourworldindata.org/economic-growth run by Max Roser has a number of striking figures and facts about growth. You will enjoy going there.
- A rich description of the many dimensions of growth is given by Charles Jones, "The Facts of Economic Growth," *Handbook of Macroeconomics*, 2016, Vol. 2A, pp. 3–69. Jones has also written a textbook on growth (*Introduction to Economic Growth*, Charles Jones and Dietrich Vollrath, 3rd edition, W. W. Norton & Co, New York, 2013), which, if you want to learn more than there is in this book, I highly recommend.
- A broad presentation of facts about growth is given by Angus Maddison in *The World Economy: A Millennial Perspective* (2001, OECD). The associated site, www.theworldeconomy.org, has a large number of facts and data on growth over the last two millennia.

11

Saving, Capital Accumulation, and Output

Since 1970, the US **saving rate**—the ratio of saving to gross domestic product (GDP)—has averaged only 17%, compared to 23% in Germany and 29% in Japan. Would a higher US saving rate have led to a higher US growth rate over the period? Would increasing the US saving rate lead to sustained higher US growth in the future?

We gave the basic answer to these questions at the end of Chapter 10: no. Over long periods—an important qualification to which we will return—an economy's growth rate does not depend on its saving rate. It does not appear that a higher saving rate would have led to higher US growth rate over the last 50 years. Nor should we expect that an increase in the saving rate will lead to sustained higher US growth.

This conclusion does not mean, however, that we should not be concerned about the low US saving rate. Even if the saving rate does not permanently affect the growth rate, it does affect the level of output and the standard of living. An increase in the saving rate would lead to higher growth for some time and eventually to a higher standard of living in the United States.

This chapter focuses on the effects of the saving rate on the level and growth rate of output.

Sections 11-1 and 11-2 look at the interactions between output and capital accumulation and the effects of the saving rate.

Section 11-3 plugs in numbers to give a better sense of the magnitudes involved.

Section 11-4 extends our discussion to take into account not only physical but also human capital.

If you remember one basic message from this chapter, it should be: Capital accumulation increases output, but cannot by itself sustain growth. ▶▶▶

11-1 INTERACTIONS BETWEEN OUTPUT AND CAPITAL

At the center of the determination of output in the long run are two relations between output and capital:

- The amount of capital determines the amount of output being produced.
- The amount of output being produced determines the amount of saving and, in turn, the amount of capital being accumulated over time.

Together, these two relations, which are represented in Figure 11-1, determine the evolution of output and capital over time. The green arrow captures the first relation, from capital to output. The blue and purple arrows capture the two parts of the second relation, from output to saving and investment, and from investment to the change in the capital stock. Let's look at each relation in turn.

The Effects of Capital on Output

We started discussing the first of these two relations, the effect of capital on output, in Section 10-4 in Chapter 10. There we introduced the aggregate production function and you saw that, under the assumption of constant returns to scale, we can write the following relation between output and capital per worker (equation (10.3)):

$$\frac{Y}{N} = F\left(\frac{K}{N}, 1\right)$$

Output per worker (Y/N) is an increasing function of capital per worker (K/N). Under the assumption of decreasing returns to capital, the effect of a given increase in capital per worker on output per worker decreases as the ratio of capital per worker gets larger. When capital per worker is already high, further increases in capital per worker have only a small effect on output per worker.

To simplify notation, we will rewrite this relation between output and capital per worker as

$$\frac{Y}{N} = f\left(\frac{K}{N}\right)$$

where the function f represents the same relation between output and capital per worker as the function F:

$$f\left(\frac{K}{N}\right) \equiv F\left(\frac{K}{N}, 1\right)$$

Suppose, for example, that the function F has the double square root form $F(K,N) = \sqrt{K}\sqrt{N}$, so

$Y = \sqrt{K}\sqrt{N}$

Divide both sides by N, so

$Y/N = \sqrt{K}\sqrt{N}/N$

Note that $\sqrt{N}/N = \sqrt{N}/(\sqrt{N}\sqrt{N})$. Using this result in the preceding equation leads to the following model of income per person:

$Y/N = \sqrt{K}/\sqrt{N} = \sqrt{K/N}$

So, in this case, the function f giving the relation between output per worker and capital per worker is simply the square root function

$f(K/N) = \sqrt{K/N}$

In this chapter, we shall make two further assumptions:

- The first is that the size of the population, the participation rate, and the unemployment rate are all constant. This implies that employment, N, is also constant. To see why, go back to the relations we saw in Chapter 2 and again in Chapter 7, between population, the labor force, unemployment, and employment:
  - The labor force is equal to population multiplied by the participation rate. So if population is constant and the participation rate is constant, the labor force is also constant.
  - Employment, in turn, is equal to the labor force multiplied by 1 minus the unemployment rate. If, for example, the size of the labor force is 100 million and the unemployment rate is 5%, then employment is equal to 95 million (100 million times $(1 - 0.05)$). So, if the labor force is constant and the unemployment rate is constant, employment is also constant.

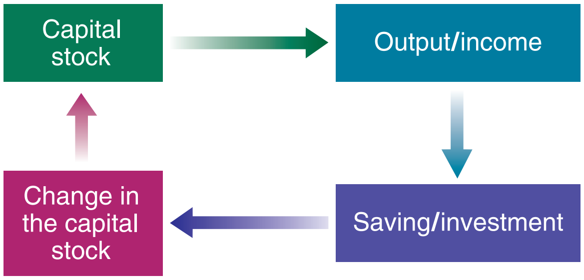


Figure 11-1

Capital, Output, and Saving/Investment

Under these assumptions, output per worker, output per person, and output itself all move proportionately. Although we will usually refer to movements in output or capital *per worker*, to lighten the text we shall sometimes just talk about movements in output or capital, leaving out the "per worker" or "per person" qualification.

The reason for assuming that N is constant is to make it easier to focus on how capital accumulation affects growth. If N is constant, the only factor of production that changes over time is capital. The assumption is not realistic, however, so we will relax it in the next two chapters. (There are two ways in which this is a simplification. The first is that population, and thus employment, typically increase over time. The other is that, in the short run, as we saw in the first nine chapters, employment may deviate from potential employment.)

In the United States in 2017, output per person (in 2011 PPP dollars) was $54,795; output per worker was much higher, at $115,120. (From these two numbers, can you derive the ratio of employment to population?)

- The second assumption is that there is no technological progress, so the production function f (or, equivalently, F) does not change over time.

  Again, the reason for making this assumption—which is obviously contrary to reality—is to focus just on the role of capital accumulation. In Chapter 12, we shall introduce technological progress and see that the basic conclusions we derive here about the role of capital in growth also hold when there is technological progress. Again, this step is better left to later.

With these two assumptions, our first relation between output and capital per worker, from the production side, can be written as

$$\frac{Y_t}{N} = f\left(\frac{K_t}{N}\right) \tag{11.1}$$

where we have introduced time indexes for output and capital—but not for labor, N, which we assume to be constant and so does not need a time index.

In words: Higher capital per worker leads to higher output per worker.

From the production side: The level of capital per worker determines the level of output per worker.

The Effects of Output on Capital Accumulation

To derive the second relation between output and capital accumulation, we proceed in two steps.

First, we derive the relation between output and investment.

Then we derive the relation between investment and capital accumulation.

Output and Investment

To derive the relation between output and investment, we make three assumptions:

- We continue to assume that the economy is closed. As we saw in Chapter 3 (equation (3.10)), this means that investment, I, is equal to saving—the sum of private saving, S, and public saving, $T - G$.

$$I = S + (T - G)$$

As we shall see in Chapter 17, saving and investment need not be equal in an open economy. A country can save less than it invests, and borrow the difference from the rest of the world. This is the case for the United States today.

- To focus on the behavior of private saving, we assume that public saving, $T - G$, is equal to zero. (We shall later relax this assumption when we focus on the effects of fiscal policy on growth.) With this assumption, the previous equation becomes

$$I = S$$

  Investment is equal to private saving.

You have now seen two specifications of saving behavior (equivalently consumption behavior): one for the short run in Chapter 3, and one for the long run in this chapter. You may wonder how the two specifications relate to each other and whether they are consistent. The answer is yes. A full discussion is given in Chapter 15.

- We assume that private saving is proportional to income, so

$$S = sY$$

  The parameter s is the saving rate. It has a value between zero and 1. This assumption captures two basic facts about saving. First, the saving rate does not appear to systematically increase or decrease as a country becomes richer. Second, richer countries do not appear to have systematically higher or lower saving rates than poorer ones.

Combining these two relations and introducing time indexes gives a simple relation between investment and output:

$$I_t = sY_t$$

Investment is proportional to output; the higher output is, the higher is saving and so the higher is investment.

Investment and Capital Accumulation

The second step relates investment, which is a flow (the new machines produced and new plants built during a given period), to capital, which is a stock (the existing machines and plants in the economy at a point in time).

Recall: Flows are variables that have a time dimension (that is, they are defined per unit of time); stocks are variables that do not have a time dimension (they are defined at a point in time). Output, saving, and investment are flows. Employment and capital are stocks.

Think of time as measured in years, so t denotes year t, $t + 1$ denotes year $t + 1$, and so on. Think of the capital stock as being measured at the beginning of each year, so K_t refers to the capital stock at the beginning of year t, K_{t+1} to the capital stock at the beginning of year $t + 1$, and so on.

Assume that capital depreciates at rate δ (the lowercase Greek letter delta) per year. That is, from one year to the next, a proportion δ of the capital stock breaks down and becomes useless. Equivalently, a proportion $(1 - \delta)$ of the capital stock remains intact from one year to the next.

The evolution of the capital stock is then given by

$$K_{t+1} = (1 - \delta)K_t + I_t$$

The capital stock at the beginning of year $t + 1$, K_{t+1}, is equal to the capital stock at the beginning of year t, which is still intact in year $t + 1$, $(1 - \delta)K_t$, plus the new capital stock put in place during year t (i.e., investment during year t, I_t).

We can now combine the relation between output and investment and the relation between investment and capital accumulation to obtain the second relation we need to think about growth: the relation between output and capital accumulation.

Replacing investment by its expression above and dividing both sides by N (the number of workers in the economy) gives

$$\frac{K_{t+1}}{N} = (1 - \delta)\frac{K_t}{N} + s\frac{Y_t}{N}$$

In words: Capital per worker at the beginning of year $t + 1$ is equal to capital per worker at the beginning of year t, adjusted for depreciation, plus investment per worker during year t, which is equal to the saving rate times output per worker during year t.

Expanding the term $(1 - \delta)K_t/N$ to $K_t/N - \delta K_t/N$, moving K_t/N to the left, and reorganizing the right side,

$$\frac{K_{t+1}}{N} - \frac{K_t}{N} = s\frac{Y_t}{N} - \delta\frac{K_t}{N} \tag{11.2}$$

In words: The change in the capital stock per worker, represented by the difference between the two terms on the left, is equal to saving per worker, represented by the first term on the right, minus depreciation, represented by the second term on the right. This equation gives us the second relation between output and capital per worker.

From the saving side: The level of output per worker determines the change in the level of capital per worker over time.

11-2 THE IMPLICATIONS OF ALTERNATIVE SAVING RATES

We have derived two relations:

- From the production side, we have seen in equation (11.1) how capital determines output.
- From the saving side, we have seen in equation (11.2) how output in turn determines capital accumulation.

We can now put the two relations together and see how they determine the behavior of output and capital over time.

Dynamics of Capital and Output

Replacing output per worker (Y_t/N) in equation (11.2) by its expression in terms of capital per worker (K_t/N) from equation (11.1) gives

$$\frac{K_{t+1}}{N} - \frac{K_t}{N} = sf\left(\frac{K_t}{N}\right) - \delta\left(\frac{K_t}{N}\right) \tag{11.3}$$

change in capital from year t to year $t+1$ = investment during year t − depreciation during year t

This relation describes what happens to capital per worker. The change in capital per worker from this year to next year depends on the difference between two terms:

- Investment per worker, the first term on the right: The level of capital per worker this year determines output per worker this year. Given the saving rate, output per worker determines the amount of saving per worker and thus the investment per worker this year.
- Depreciation per worker, the second term on the right: The capital stock per worker determines the amount of depreciation per worker this year.

$K_t/N \Rightarrow f(K_t/N) \Rightarrow sf(K_t/N)$

$K_t/N \Rightarrow \delta K_t/N$

If investment per worker exceeds depreciation per worker, the change in capital per worker is positive. Capital per worker increases.

If investment per worker is less than depreciation per worker, the change in capital per worker is negative. Capital per worker decreases.

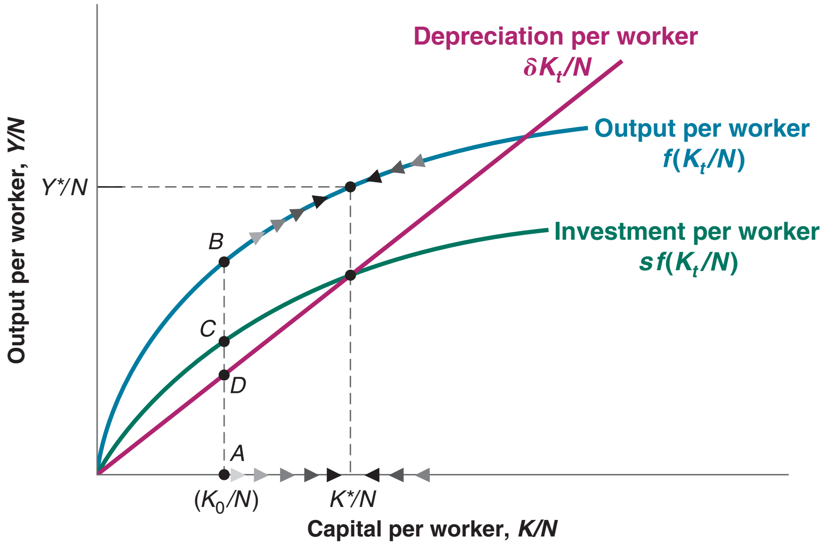
Given capital per worker, output per worker is then given by equation (11.1):

$$\frac{Y_t}{N} = f\left(\frac{K_t}{N}\right)$$

Figure 11-2

Capital and Output Dynamics

When capital and output are low, investment exceeds depreciation and capital increases. When capital and output are high, investment is less than depreciation and capital decreases.



The hard work is done. Equations (11.3) and (11.1) contain all the information we need to understand the dynamics of capital and output over time. The easiest way to interpret them is to use a graph. We do this in Figure 11-2: Output per worker is measured on the vertical axis, and capital per worker is measured on the horizontal axis.

In Figure 11-2, look first at the curve representing output per worker, $f(K_t/N)$, as a function of capital per worker. The relation is the same as in Figure 10-4: Output per worker increases with capital per worker, but, because of decreasing returns to capital, the effect is smaller the higher the level of capital per worker.

Now look at the two curves representing the two components on the right side of equation (11.3):

- The relation representing investment per worker, $sf(K_t/N)$, has the same shape as the production function except that it is lower by a factor s (the saving rate). Suppose the level of capital per worker is equal to K_0/N in Figure 11-2. Output per worker is then given by the vertical distance AB, and investment per worker is given by the vertical distance AC, which is equal to s times the vertical distance AB. Thus, just like output per worker, investment per worker increases with capital per worker, but by less and less as capital per worker increases. When capital per worker is already high, the effect of a further increase in capital per worker on output per worker, and by implication on investment per worker, is small.
- The relation representing depreciation per worker, $\delta K_t/N$, is represented by a straight line. Depreciation per worker increases in proportion to capital per worker so the relation is represented by a straight line with slope equal to δ. At the level of capital per worker K_0/N, depreciation per worker is given by the vertical distance AD.

To make the graph easier to read, I have assumed an unrealistically high saving rate. (Can you tell roughly what value I have assumed for s? What would be a plausible value for s?) ▶

The change in capital per worker is given by the difference between investment per worker and depreciation per worker. At K_0/N, the difference is positive; investment per worker exceeds depreciation per worker by an amount represented by the vertical distance $CD = AC - AD$, so capital per worker increases. As we move to the right along the horizontal axis and look at higher and higher levels of capital per worker, investment increases by less and less, while depreciation keeps increasing in proportion to capital. For some level of capital per worker, K^*/N in Figure 11-2, investment is just enough to cover depreciation, and capital per worker remains constant. To the left of K^*/N, investment exceeds depreciation and capital per worker increases. This is indicated by the arrows pointing to the right along the curve representing the production function. To the right of K^*/N, depreciation exceeds investment, and capital per worker decreases. This is indicated by the arrows pointing to the left along the curve representing the production function.

When capital per worker is low, capital per worker and output per worker increase over time. When capital per worker is high, capital per worker and output per worker decrease over time.

Characterizing the evolution of capital per worker and output per worker over time now is easy. Consider an economy that starts with a low level of capital per worker—say, K_0/N in Figure 11-2. Because investment exceeds depreciation at this point, capital per worker increases. And because output moves with capital, output per worker increases as well. Capital per worker eventually reaches K^*/N, the level at which investment is equal to depreciation. Once the economy has reached the level of capital per worker K^*/N, output per worker and capital per worker remain constant at Y^*/N and K^*/N, their long-run equilibrium levels.

Think, for example, of a country that loses part of its capital stock, say as a result of bombing during a war. The mechanism we have just seen suggests that, if the country has suffered larger capital losses than population losses, it will come out of the war with a low level of capital per worker; that is, at a point to the left of K^*/N. The country will then experience a large increase in both capital per worker and output per worker for some time. This describes well what happened after World War II to countries that had proportionately larger destructions of capital than losses of human lives (see the Focus Box "Capital Accumulation and Growth in France in the Aftermath of World War II").

What does the model predict for postwar growth if a country suffers proportional losses in population and in capital? Do you find this answer convincing? What elements may be missing from the model?

If a country starts instead from a high level of capital per worker—that is, from a point to the right of K^*/N—then depreciation will exceed investment, and capital per worker and output per worker will decrease. The initial level of capital per worker is too high to be sustained given the saving rate. This decrease in capital per worker will continue until the economy again reaches the point where investment is equal to depreciation and capital per worker is equal to K^*/N. From then on, capital per worker and output per worker will remain constant.

Let's look more closely at the levels of output per worker and capital per worker to which the economy converges in the long run. The state in which output per worker and capital per worker are no longer changing is called the **steady state** of the economy. Setting the left side of equation (11.3) equal to zero (in steady state, by definition, the change in capital per worker is zero), the steady-state value of capital per worker, K^*/N, is given by

$$s f\left(\frac{K^*}{N}\right) = \delta \frac{K^*}{N} \qquad (11.4)$$

The steady-state value of capital per worker is such that the amount of saving per worker (the left side) is just sufficient to cover depreciation of the capital stock per worker (the right side of the equation).

K^*/N is the long-run level of capital per worker.

Given steady-state capital per worker (K^*/N), the steady-state value of output per worker (Y^*/N) is given by the production function

$$\frac{Y^*}{N} = f\left(\frac{K^*}{N}\right) \qquad (11.5)$$

We now have all the elements we need to discuss the effects of the saving rate on output per worker, both over time and in steady state.

The Saving Rate and Output

Let's return to the question we posed at the beginning of the chapter: How does the saving rate affect the growth rate of output per worker? Our analysis leads to a three-part answer:

1. *The saving rate has no effect on the long-run growth rate of output per worker, which is equal to zero.*

    This conclusion is rather obvious; we have seen that, eventually, the economy converges to a constant level of output per worker. In other words, in the long run, the growth rate of output is equal to zero, no matter what the saving rate is.

FOCUS

Capital Accumulation and Growth in France in the Aftermath of World War II

When World War II ended in 1945, France had suffered some of the heaviest losses of all European countries. The losses in lives were large; out of a population of 42 million, more than 550,000 people died. Relatively speaking, though, the losses in capital were much larger. It is estimated that the French capital stock in 1945 was about 30% below its prewar value. A vivid picture of the destruction of capital is provided by the numbers in Table 1.

The model of growth we have just seen makes a clear prediction about what will happen to a country that loses a large part of its capital stock. The country will experience high capital accumulation and output growth for some time. In terms of Figure 11-2, a country with capital per worker initially far below K^*/N will grow rapidly as it converges to K^*/N and output per worker converges to Y^*/N.

This prediction fares well in the case of postwar France. There is plenty of anecdotal evidence that small increases in capital led to large increases in output. Minor repairs to a major bridge would lead to the reopening of the bridge. Reopening the bridge would significantly shorten the travel time between two cities, leading to much lower transport costs. The lower transport costs would then enable a plant to get much-needed inputs, increase its production, and so on.

More convincing evidence, however, comes directly from actual aggregate output numbers. From 1946 to 1950, the annual growth rate of French real GDP was a high 9.6% per year. This led to an increase in real GDP of about 60% over 5 years.

Was all of the increase in French GDP the result of capital accumulation? The answer is no. There were other forces at work in addition to the mechanism in our model. Much of the remaining capital stock in 1945 was old. Investment had been low in the 1930s (a decade dominated by the Great Depression) and nearly nonexistent during the war. A good portion of the postwar capital accumulation was associated with the introduction of more modern capital and the use of more modern production techniques. This was another reason for the high growth rates of the postwar period.[1]

Table 1 Proportion of the French Capital Stock Destroyed by the End of World War II

Railways	Tracks	6%	Rivers	Waterways	86%
	Stations	38%		Canal locks	11%
	Engines	21%		Barges	80%
	Hardware	60%	Buildings	(numbers)	
Roads	Cars	31%		Dwellings	1,229,000
	Trucks	40%		Industrial	246,000

Some economists argue that the high output growth achieved by the Soviet Union from 1950 to 1990 was the result of such a steady increase in the saving rate over time, which could not be sustained forever. Paul Krugman has used the term *Stalinist growth* to denote this type of growth, which is growth resulting from a higher and higher saving rate over time.

Note that the first proposition is a statement about the growth rate of output per worker. The second proposition is a statement about the level of output per worker.

There is, however, a way of thinking about this conclusion that will be useful when we allow for technological progress in Chapter 12. Think of what would be needed to sustain a constant positive growth rate of output per worker in the long run. Capital per worker would have to increase. Not only that, but because of decreasing returns to capital, it would have to increase faster than output per worker. This implies that each year the economy would have to save a larger and larger fraction of its output and dedicate it to capital accumulation. At some point, the fraction of output it would need to save would be greater than 1—something clearly impossible. This is why it is impossible, absent technological progress, to sustain a constant positive growth rate forever. In the long run, capital per worker must be constant, and so output per worker must also be constant.

2. Nonetheless, *the saving rate determines the level of output per worker in the long run*. Other things being equal, countries with a higher saving rate will achieve higher output per worker in the long run.

Figure 11-3 illustrates this point. Consider two countries with the same production function, the same level of employment, and the same depreciation rate, but with different saving rates, say s_0 and $s_1 > s_0$. Figure 11-3 draws

[1]Source: *Gilles Saint-Paul, "Economic Reconstruction in France, 1945–1958," in Rudiger Dornbusch, Willem Nolling, and Richard Layard, eds.,* Postwar Economic Reconstruction and Lessons for the East Today *(1993, MIT Press).*

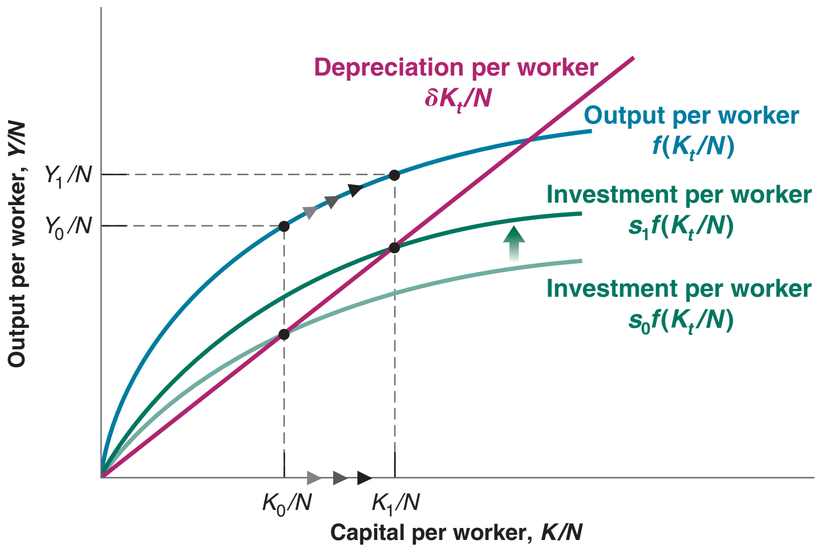


Figure 11-3

The Effects of Different Saving Rates

A country with a higher saving rate achieves a higher steady-state level of output per worker.

their common production function, $f(K_t/N)$, and the functions showing saving/investment per worker as a function of capital per worker for each of the two countries, $s_0f(K_t/N)$ and $s_1f(K_t/N)$. In the long run, the country with saving rate s_0 will reach the level of capital per worker K_0/N and output per worker Y_0/N. The country with saving rate s_1 will reach the higher levels K_1/N and Y_1/N.

3. *An increase in the saving rate will lead to higher growth of output per worker for some time, but not forever.*

This conclusion follows from the two propositions we just discussed. From the first, we know that an increase in the saving rate does not affect the long-run *growth rate of output per worker*, which remains equal to zero. From the second, we know that an increase in the saving rate leads to an increase in the long-run *level of output per worker*. It follows that, as output per worker increases to its new higher level in response to the increase in the saving rate, the economy will go through a period of positive growth. This period of growth will come to an end when the economy reaches its new steady state.

We can use Figure 11-3 again to illustrate this point. Consider a country that has an initial saving rate of s_0. Assume that capital per worker is initially equal to K_0/N, with associated output per worker Y_0/N. Now consider the effects of an increase in the saving rate from s_0 to s_1. The relation giving saving/investment per worker as a function of capital per worker shifts upward from $s_0f(K_t/N)$ to $s_1f(K_t/N)$.

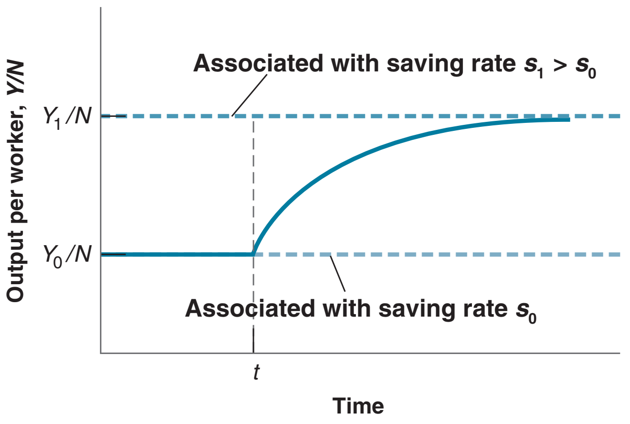
At the initial level of capital per worker, K_0/N, investment exceeds depreciation, so capital per worker increases. As capital per worker increases, so does output per worker, and the economy goes through a period of positive growth. When capital per worker eventually reaches K_1/N, however, investment is again equal to depreciation, and growth ends. From then on, the economy remains at K_1/N, with associated output per worker Y_1/N. The movement of output per worker is plotted against time in Figure 11-4. Output per worker is initially constant at level Y_0/N. After an increase in the saving rate at time t, output per worker increases for some time until it reaches Y_1/N and the growth rate returns to zero.

We have derived these three results under the assumption that there was no technological progress, and therefore, no growth of output per worker in the long run. But, as we will see in Chapter 12, the three results extend to an economy in which there is technological progress. Let us briefly see how.

Figure 11-4

The Effects of an Increase in the Saving Rate on Output per Worker in an Economy without Technological Progress

An increase in the saving rate leads to a period of higher growth until output reaches its new higher steady-state level.



An economy in which there is technological progress has a positive growth rate of output per worker, even in the long run. This long-run growth rate is independent of the saving rate—the extension of the first result just discussed. The saving rate affects the level of output per worker, however—the extension of the second result. An increase in the saving rate leads to growth greater than steady-state growth for some time until the economy reaches its new higher path—the extension of our third result.

These three results are illustrated in Figure 11-5, which extends Figure 11-4 by plotting the effect of an increase in the saving rate on an economy with positive technological progress. The figure uses a logarithmic scale to measure output per worker. It follows that an economy in which output per worker grows at a constant rate is represented by a line with slope equal to that growth rate. At the initial saving rate, s_0, the economy moves along AA. If, at time t, the saving rate increases to s_1, the economy experiences higher growth for some time until it reaches its new, higher path, BB. The growth rate then becomes the same as before the increase in the saving rate (that is, the slope of BB is the same as the slope of AA).

See the discussion of logarithmic scales in Appendix 2 at the end of the book. ▶

The Saving Rate and Consumption

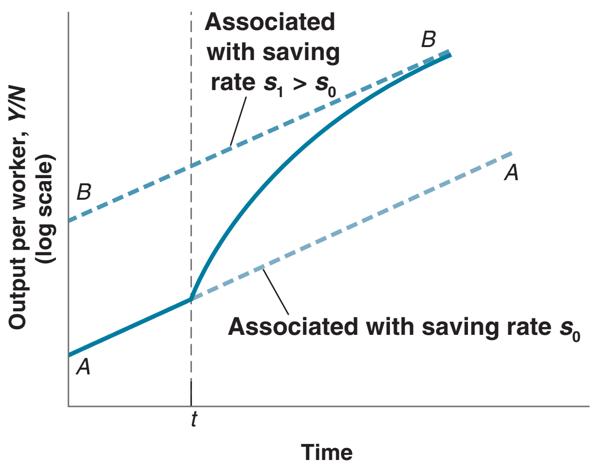
Governments can affect the saving rate in various ways. First, they can vary public saving. Given private saving, positive public saving—a budget surplus, in other words—leads to higher overall saving. Conversely, negative public saving—a budget deficit—leads to lower overall saving. Second, governments can use taxes to affect private saving. For example, they can give tax breaks to people who save, making it more attractive to save and thus increasing private saving.

Recall: Saving is the sum of private plus public saving. Recall also that Public saving ⇔ Budget surplus; Public dissaving ⇔ Budget deficit. ▶

Figure 11-5

The Effects of an Increase in the Saving Rate on Output per Worker in an Economy with Technological Progress

An increase in the saving rate leads to a period of higher growth until output reaches a new, higher path.



What saving rate should governments aim for? To think about the answer, we must shift our focus from the behavior of *output* to the behavior of *consumption*. The reason: What matters to people is not how much is produced, but how much they consume.

It is clear that an increase in saving must come initially at the expense of lower consumption (except when we think it helpful, we drop "per worker" in this subsection and just refer to consumption rather than consumption per worker, capital rather than capital per worker, and so on). A change in the saving rate this year has no effect on capital this year and consequently no effect on output and income *this year*. So an increase in saving comes initially with an equal decrease in consumption.

Because we assume that employment is constant, we are ignoring the short-run effect of an increase in the saving rate on output we focused on in Chapters 3, 5, and 9. In the short run, not only does an increase in the saving rate reduce consumption given income, but it may also create a recession and decrease income further. We shall return to a discussion of short-run and long-run effects of changes in saving in Chapters 16 and 22.

Does an increase in saving lead to an increase in consumption in the long run? Not necessarily. Consumption may decrease, not only initially but also in the long run. You may find this surprising. After all, we know from Figure 11-3 that an increase in the saving rate always leads to an increase in the level of *output* per worker. But output is not the same as consumption. To see why not, consider what happens for two extreme values of the saving rate.

- An economy in which the saving rate is (and has always been) zero is an economy in which capital is equal to zero. In this case, output is also equal to zero, and so is consumption. A saving rate equal to zero implies zero consumption in the long run.
- Now consider an economy in which the saving rate is equal to one. People save all their income. The level of capital, and thus output, in this economy will be high. But because people save all of their income, consumption is equal to zero. What happens is that the economy is carrying an excessive amount of capital. Simply maintaining that level of output requires that all output be devoted to replacing depreciation! A saving rate equal to one also implies zero consumption in the long run.

These two extreme cases mean that there must be some value of the saving rate between zero and one that maximizes the steady-state level of consumption. Increases in the saving rate below this value lead to a decrease in consumption initially, but they lead to an increase in consumption in the long run. Increases in the saving rate beyond this value decrease consumption not only initially but also in the long run. This happens because the increase in capital associated with the increase in the saving rate leads to only a small increase in output—an increase that is too small to cover the increased depreciation. In other words, the economy carries too much capital. The level of capital associated with the value of the saving rate that yields the highest level of consumption in steady state is known as the **golden-rule level of capital**. Increases in capital beyond the golden-rule level reduce steady-state consumption.

This argument is illustrated in Figure 11-6, which plots consumption per worker in steady state (on the vertical axis) against the saving rate (on the horizontal axis).

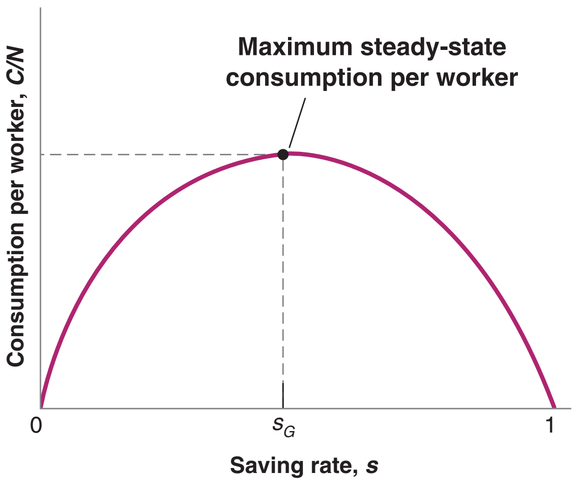


Figure 11-6

The Effects of the Saving Rate on Steady-State Consumption per Worker

An increase in the saving rate leads to an increase, then to a decrease in steady-state consumption per worker.

A saving rate equal to zero implies a capital stock per worker equal to zero, a level of output per worker equal to zero, and, by implication, a level of consumption per worker equal to zero. For s between zero and s_G (G for golden rule), a higher saving rate leads to higher capital per worker, higher output per worker, and higher consumption per worker. For s larger than s_G, increases in the saving rate still lead to higher values of capital per worker and output per worker; but they now lead to lower values of consumption per worker. This is because the increase in output is more than offset by the increase in depreciation as a result of the larger capital stock. For $s = 1$, consumption per worker is equal to zero. Capital per worker and output per worker are high, but all of the output is used just to replace depreciation, leaving nothing for consumption.

If an economy already has so much capital that it is operating beyond the golden rule, then increasing saving further will decrease consumption not only now, but also later. Is this a relevant worry? Do some countries actually have too much capital? The empirical evidence indicates that OECD countries are typically below their golden-rule level of capital. If they were to increase the saving rate, it would lead to higher consumption in the future—not lower consumption.

This means that, in practice, governments face a trade-off: An increase in the saving rate leads to lower consumption for some time but higher consumption later. So what should governments do? How close to the golden rule should they try to get? That depends on how much weight they put on the welfare of current generations—who are more likely to lose from policies aimed at increasing the saving rate—versus the welfare of future generations—who are more likely to gain. Enter politics: Future generations do not vote. This means that governments are unlikely to ask current generations to make large sacrifices, which, in turn, means that countries are more likely to underinvest than overinvest. These intergenerational issues are at the forefront of the debate on Social Security reform in the United States. The Focus Box "Social Security, Saving, and Capital Accumulation in the United States" explores this further.

11-3 GETTING A SENSE OF MAGNITUDES

How big an impact does a change in the saving rate have on output in the long run? For how long and by how much does an increase in the saving rate affect growth? How far is the United States from the golden-rule level of capital? To get a better sense of the answers to these questions, let's now make more specific assumptions, plug in some numbers, and see what we get.

Assume the production function is

$$Y = \sqrt{K}\sqrt{N} \tag{11.6}$$

Check that this production function exhibits both constant returns to scale and decreasing returns to either capital or labor. ▶

Output equals the product of the square root of capital and the square root of labor. (A more general specification of the production function known as the Cobb-Douglas production function, and its implications for growth, is given in the appendix to this chapter.)

Dividing both sides by N (because we are interested in output per worker),

$$\frac{Y}{N} = \frac{\sqrt{K}\sqrt{N}}{N} = \frac{\sqrt{K}}{\sqrt{N}} = \sqrt{\frac{K}{N}}$$

The second equality follows from $\sqrt{N}/N = \sqrt{N}/(\sqrt{N}\sqrt{N}) = 1/\sqrt{N}$. ▶

Output per worker equals the square root of capital per worker. Put another way, the production function f relating output per worker to capital per worker is given by

$$f\left(\frac{K_t}{N}\right) = \sqrt{\frac{K_t}{N}}$$

Social Security, Saving, and Capital Accumulation in the United States

FOCUS

Social Security was introduced in the United States in 1935. The goal of the program was to make sure the elderly would have enough to live on. Over time, Social Security has become the largest government program in the United States. Benefits paid to retirees now exceed 4% of GDP. For two-thirds of retirees, Social Security benefits account for more than 50% of their income. There is little question that, on its own terms, the Social Security system has been a great success and has decreased poverty among the elderly. There is also little question that it has also led to a lower US saving rate and therefore lower capital accumulation and lower output per person in the long run.

To understand why, we must take a theoretical detour. Think of an economy in which there is no social security system—one where workers must save to provide for their own retirement. Now, introduce a social security system that collects taxes from workers and distributes benefits to the retirees. It can do so in one of two ways:

- One way is by taxing workers, investing their contributions in financial assets, and paying back the principal plus the interest to the workers when they retire. Such a system is called a **fully funded social security system**: At any time, the system has funds equal to the accumulated contributions of workers, from which it will be able to pay benefits to these workers when they retire.
- The other way is by taxing workers and redistributing the tax contributions as benefits to the current retirees. Such a system is called a **pay-as-you-go social security system**. The system pays benefits out "as it goes," that is, as it collects them through contributions.

From the point of view of workers, the two systems may look broadly similar. In both cases, people pay contributions when they work and receive benefits when they retire. But there are two major differences.

First, what retirees receive is different in each case:

- What they receive in a fully funded system depends on the rate of return on the financial assets held by the fund.
- What they receive in a pay-as-you-go system depends on demographics—the ratio of retirees to workers—and on the evolution of the tax rate set by the system. When the population ages and the ratio of retirees to workers increases, then either retirees receive less or workers must contribute more. This is the case in the United States today. The ratio of retirees to workers, which is about 0.3 today, is forecast to increase to 0.4–0.5 in 20 years. Under current rules, benefits will increase from 5% of GDP today to 6% in 20 years. Thus, either benefits will have to be reduced, in which case the rate of return to workers who contributed in the past will be low, or contributions will have to be increased, in which case this will decrease the rate of return to workers who are contributing today, or more likely, some combination of both will have to be implemented.

Second, and leaving aside the aging issue, the two systems have different macroeconomic implications:

- In the fully funded system, workers save less because they anticipate receiving benefits when they are old. But the Social Security system saves on their behalf, by investing their contributions in financial assets. The presence of a social security system changes the composition of overall saving: Private saving goes down, and public saving goes up. But, to a first approximation, it has no effect on total saving and therefore no effect on capital accumulation.
- In the pay-as-you-go system, workers also save less because they again anticipate receiving benefits when they are old. But now, the Social Security system does not save on their behalf. The decrease in private saving is not compensated by an increase in public saving. Total saving goes down, and so does capital accumulation.

Most social security systems are somewhere between pay-as-you-go and fully funded systems. When the US system was set up in 1935, the intention was to partially fund it. But this did not happen. Rather than being invested, contributions from workers were used to pay benefits to the retirees, and this has been the case ever since. Today, because contributions have slightly exceeded benefits since the early 1980s, Social Security has built a **Social Security trust fund**. But this trust fund is far smaller than the value of benefits promised to current contributors when they retire. The US system is basically a pay-as-you-go system, and this has probably led to a lower US saving rate over the last 70 years.

In this context, some economists and politicians have suggested that the United States should shift back to a fully funded system. One of their arguments is that the US saving rate is indeed too low and that funding the Social Security system would increase it. Such a shift could be achieved by investing, from now on, tax contributions in financial assets rather than distributing them as benefits to retirees. Under such a shift, the Social Security system would steadily accumulate funds and would eventually become fully funded. Martin Feldstein, an economist at Harvard and an advocate of such a shift, has concluded that it could lead to a 34% increase of the capital stock in the long run.

How should we think about such a proposal? It would probably have been a good idea to fully fund the system at the start. The United States would have a higher saving rate. The US capital stock would be higher, and output and consumption would also be higher. But we cannot rewrite history. The existing system has promised benefits to retirees and these promises must be honored. This means that, under the proposal just described, current workers would, in effect, have to contribute twice; once to fund the system and finance their own retirement, and then again to finance the benefits owed to current retirees. This would impose a disproportionate cost on current workers. And this would come on top of the problems coming from the aging of population which will increase the ratio of retirees to workers, and will require, if retirement benefits are to be maintained, larger contributions from workers. The practical implication is that, if it is to happen, the move to a fully funded system will have to be slow, so that the burden of adjustment does not fall too much on one generation relative to the others.

The debate is likely to be with us for some time. In assessing proposals from the administration or from Congress, ask yourself how they deal with these issues. Take, for example, the proposal to allow workers, from now on, to make contributions to personal accounts instead of to the Social Security system, and to be able to draw from these accounts when they retire. By itself, this proposal would clearly increase private saving: Workers would be saving more. But its ultimate effect on saving depends on how the benefits already promised to current workers and retirees by the Social Security system are financed. If, as is the case under some proposals, these benefits are financed not through additional taxes but through debt finance, then the increase in private saving will be offset by an increase in deficits (i.e., a decrease in public saving). The shift to personal accounts will not increase the US saving rate. If, instead, these benefits are financed through higher taxes, then the US saving rate will increase, but current workers will have to both contribute to their personal accounts and pay the higher taxes. They will indeed pay twice.

To follow the debate on Social Security, look at the site run by the (nonpartisan) Concord Coalition (www.concordcoalition.org) and find the discussion related to Social Security, or look at the annual report of the Social Security Administration, www.ssa.gov/OACT/TR/2018/tr2018.pdf.

Replacing $f(K_t/N)$ by $\sqrt{K_t/N}$ in equation (11.3),

$$\frac{K_{t+1}}{N} - \frac{K_t}{N} = s\sqrt{\frac{K_t}{N}} - \delta\frac{K_t}{N} \tag{11.7}$$

This equation describes the evolution of capital per worker over time. Let's look at what it implies.

The Effects of the Saving Rate on Steady-State Output

How big an impact does an increase in the saving rate have on the steady-state level of output per worker?

Start with equation (11.7). In steady state the amount of capital per worker is constant, so the left side of the equation equals zero. This implies

$$s\sqrt{\frac{K^*}{N}} = \delta\frac{K^*}{N}$$

(We have dropped time indexes, which are no longer needed because in steady state K/N is constant. The star is to remind you that we are looking at the steady-state value of capital.) Square both sides:

$$s^2\frac{K^*}{N} = \delta^2\left(\frac{K^*}{N}\right)^2$$

Divide both sides by (K/N) and reorganize:

$$\frac{K^*}{N} = \left(\frac{s}{\delta}\right)^2 \tag{11.8}$$

Steady-state capital per worker is equal to the square of the ratio of the saving rate to the depreciation rate.

From equations (11.6) and (11.8), steady-state output per worker is given by

$$\frac{Y^*}{N} = \sqrt{\frac{K^*}{N}} = \sqrt{\left(\frac{s}{\delta}\right)^2} = \frac{s}{\delta} \tag{11.9}$$

Steady-state output per worker is equal to the ratio of the saving rate to the depreciation rate.

A higher saving rate and a lower depreciation rate both lead to higher steady-state capital per worker (equation (11.8)) and higher steady-state output per worker (equation (11.9)). To see what this means, let's take a numerical example. Suppose the depreciation rate is 10% per year and the saving rate is also 10%. Then, from equations (11.8) and (11.9), steady-state capital per worker and output per worker are both equal to 1. Now suppose that the saving rate doubles, from 10% to 20%. It follows from equation (11.8) that in the new steady state, capital per worker increases from 1 to 4. And, from equation (11.9), output per worker doubles, from 1 to 2. Thus, doubling the saving rate leads, in the long run, to doubling the output per worker. This is a large effect.

The Dynamic Effects of an Increase in the Saving Rate

We have just seen that an increase in the saving rate leads to an increase in the steady-state level of output. But how long does it take for output to reach its new steady-state level? Put another way, by how much and for how long does an increase in the saving rate affect the growth rate?

To answer these questions, we must use equation (11.7) and solve it for capital per worker in year 0, in year 1, and so on.

Suppose that the saving rate, which had always been equal to 10%, increases in year 0 from 10% to 20% and remains at this higher value forever. In year 0, nothing happens to the capital stock (recall that it takes one year for higher saving and higher investment to show up in higher capital). So capital per worker remains equal to the steady-state value associated with a saving rate of 0.1. From equation (11.8),

$$\frac{K_0}{N} = (0.1/0.1)^2 = 1^2 = 1$$

In year 1, equation (11.7) gives

$$\frac{K_1}{N} - \frac{K_0}{N} = s\sqrt{\frac{K_0}{N}} - \delta\frac{K_0}{N}$$

With a depreciation rate equal to 0.1 and a saving rate now equal to 0.2, this equation implies

$$\frac{K_1}{N} - 1 = [(0.2)(\sqrt{1})] - [(0.1)1]$$

so

$$\frac{K_1}{N} = 1.1$$

In the same way, we can solve for K_2/N, and so on. Once we have determined the values of capital per worker in year 0, year 1, and so on, we can then use equation (11.6) to solve for output per worker in year 0, year 1, and so on. The results of this computation are presented in Figure 11-7. Panel (a) plots the *level* of output per worker against time. (Y/N) increases from its initial value of 1 in year 0 to its steady-state value of 2 in the long run. Panel (b) gives the same information in a different

The difference between investment and depreciation is largest at the beginning. This is why capital accumulation, and in turn output growth, are highest at the beginning.

way, plotting instead the *growth rate* of output per worker against time. As panel (b) shows, growth of output per worker is highest at the beginning and then decreases over time. As the economy reaches its new steady state, growth of output per worker returns to zero.

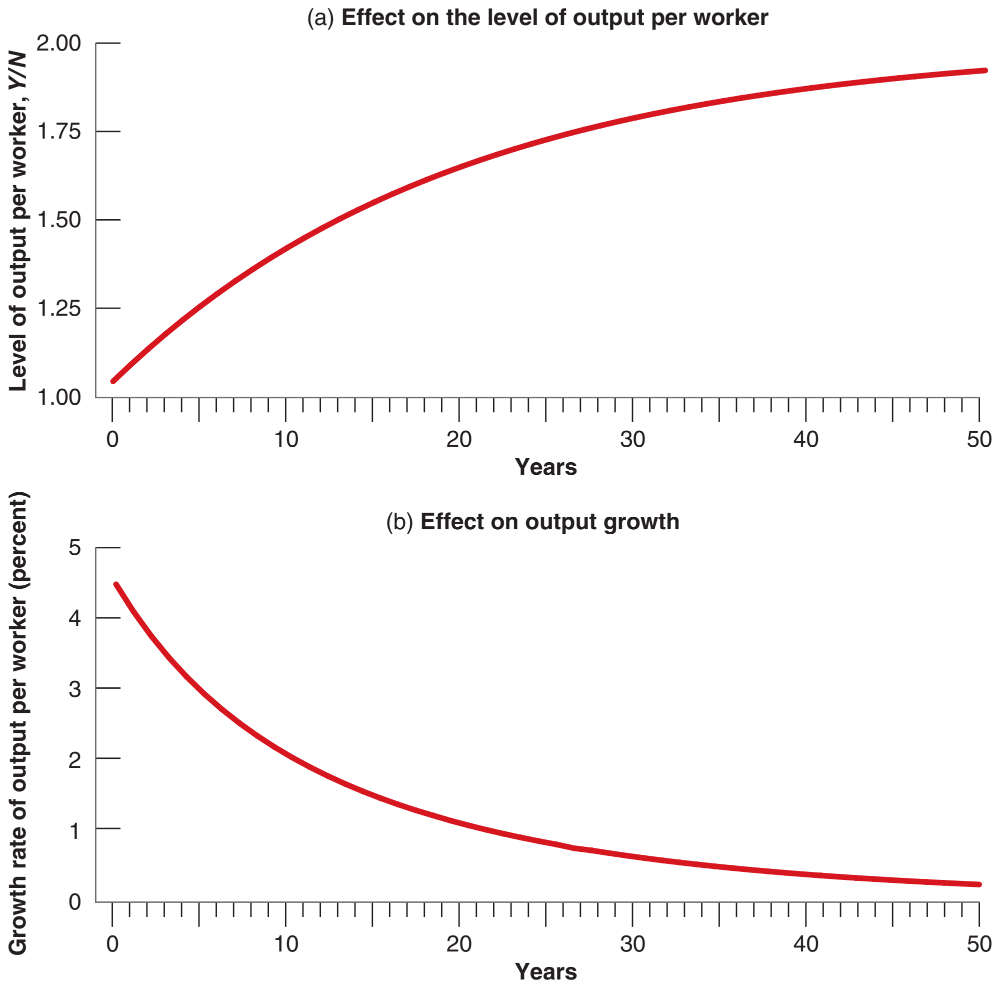
Figure 11-7 clearly shows that the adjustment to the new, higher, long-run equilibrium takes a long time. It is only 40% complete after 10 years, and 63% complete after 20 years. Put another way, the increase in the saving rate increases the growth rate of output per worker for a long time. The average annual growth rate is 3.1% for the first 10 years, and 1.5% for the next 10. Although the changes in the saving rate have no effect on growth in the long run, they do lead to higher growth for a long time.

To go back to the question raised at the beginning of the chapter, would US growth have been substantially higher if the saving rate had been higher since 1950? The answer would be yes if the United States had had a higher saving rate in the past, and *if this saving rate had fallen substantially in the last 70 years.* Then, this fall in the saving rate would have led, other things being equal, to lower growth in the United States in the last 70 years along the lines of the mechanism in Figure 11-7 (with the sign reversed, as we would be looking at a decrease,

Figure 11-7

The Dynamic Effects of an Increase in the Saving Rate from 10% to 20% on the Level and Growth Rate of Output per Worker

It takes a long time for output to adjust to its new higher level after an increase in the saving rate. Put another way, an increase in the saving rate leads to a long period of higher growth.



not an increase, in the saving rate). But this is not the case. The US saving rate has been low for a long time.

The US Saving Rate and the Golden Rule

What is the saving rate that would maximize steady-state consumption per worker? Recall that, in steady state, consumption is equal to what is left after enough is put aside to maintain a constant level of capital. More formally, in steady state, consumption per worker is equal to output per worker minus depreciation per worker:

$$\frac{C}{N} = \frac{Y}{N} - \delta\frac{K}{N}$$

Using equations (11.8) and (11.9) for the steady-state values of output per worker and capital per worker, consumption per worker is thus given by

$$\frac{C}{N} = \frac{s}{\delta} - \delta\left(\frac{s}{\delta}\right)^2 = \frac{s(1 - s)}{\delta}$$

Using this equation, together with equations (11.8) and (11.9), Table 11-1 gives the steady-state values of capital per worker, output per worker, and consumption per worker for different values of the saving rate (and for a depreciation rate equal to 10%).

Steady-state consumption per worker is largest when s equals 1/2. In other words, the golden-rule level of capital is associated with a saving rate of 50%. Below that level, increases in the saving rate lead to an increase in long-run consumption per worker. We saw previously that the average US saving rate since 1970 has been only 17%. So we can be quite confident that, at least in the United States, an increase in the saving rate would increase both output per worker and consumption per worker in the long run. What could be done to increase the saving rate is discussed in the Focus Box "Nudging US Households to Save More."

This may, however, not be true of all countries in the world. Some researchers argue that China, with a saving rate of close to 50%, may indeed be on the other side of the golden rule.

Check your understanding of the issues: Using the equations in this section, argue the pros and cons of policy measures aimed at increasing the US saving rate.

Table 11-1 The Saving Rate and the Steady-State Levels of Capital, Output, and Consumption per Worker

Saving Rate s	Capital per Worker K/N	Output per Worker Y/N	Consumption per Worker C/N
0.0	0.0	0.0	0.0
0.1	1.0	1.0	0.9
0.2	4.0	2.0	1.6
0.3	9.0	3.0	2.1
0.4	16.0	4.0	2.4
0.5	25.0	5.0	2.5
0.6	36.0	6.0	2.4
...	...	...	...
1.0	100.0	10.0	0.0

FOCUS

Nudging US Households to Save More

US households save little: The household saving rate, defined as the ratio of household saving to household disposable income, has averaged 6% since 2000, compared to, for example, 10% in Germany over the same period.

Successive governments have tried to increase the saving rate by giving tax breaks, which made saving more attractive, but with limited success. Considering the disappointing effects, Richard Thaler, an economist who is considered one of the fathers of the field of behavior economics and who received the Nobel Prize in 2017, suggested an alternative approach.

He argued that the reasons people saved little had little to do with interest rates or tax breaks, but much more with the way we actually behave. The evidence is that we do not make the right decisions, even judged from our own point of view.

First, we tend to ignore the future, or give it too little weight in making decisions. We suffer from something called hyperbolic discounting. Put simply, when presented with the choice of one apple today versus one apple next week, we strongly prefer the apple today. But when presented with the choice of one apple one week from now versus one apple in two weeks, we are nearly indifferent. This leads us to consume too much today. Second, we procrastinate, and leave decisions to tomorrow. A recent study finds that, as a result of these poor decisions, 43% of Americans run out of savings in retirement (and thus must rely on Social Security benefits only).

This led Richard Thaler and Shlomo Benartzi to design an approach called "Save More Tomorrow" aimed at dealing with these two issues. First, people were asked to make decisions long before implementing them, so people would not have to increase saving and decrease consumption today but only commit to do it in the future, something psychologically much less costly. Second, and even more importantly, the saving plan option, rather than the no-saving plan option, was the default in the choice of plans. As a result, procrastination would lead people to choose by default the saving option. They would thus be "nudged" into saving, for their own good.

Has this approach worked? Quoting from the web site of Benartzi: "This 'nudge' has gone on to achieve massive scale. In 2006, it was enshrined in law as part of the Pension Protection Act of 2006, which encouraged companies to adopt the core principles of the program. According to the latest calculations, it's estimated that 'Save More Tomorrow' has helped approximately 15 million Americans significantly boost their savings rate."[2]

11-4 PHYSICAL VERSUS HUMAN CAPITAL

We have concentrated so far on physical capital—machines, plants, office buildings, and so on. But economies have another type of capital: the skills of the workers in the economy, or what economists call **human capital**. An economy with many highly skilled workers is likely to be much more productive than an economy in which most workers cannot read or write.

The increase in human capital has been as large as the increase in physical capital over the last two centuries. At the beginning of the Industrial Revolution, only 30% of the population of the countries that now constitute the OECD knew how to read; today, the literacy rate in OECD countries is above 95%. Schooling was not compulsory prior to the Industrial Revolution; today it is compulsory, usually until the age of 16. Still, there are large differences across countries. Today, in OECD countries, nearly 100% of children get a primary education, 90% get a secondary education, and 38% get a higher education. The corresponding numbers in poor countries, countries with GDP per person below $400, are 95%, 32%, and 4%, respectively.

Even this comparison may be misleading because the quality of education can be quite different across countries.

How should we think about the effect of human capital on output? How does the introduction of human capital change our earlier conclusions? These are the questions we take up in this last section.

Extending the Production Function

The most natural way of extending our analysis to allow for human capital is to modify the production function relation (11.1) to read

$$\frac{Y}{N} = f\left(\frac{K}{N}, \frac{H}{N}\right) \qquad (11.10)$$

$$(+, +)$$

[2]Source: *Used with permission, Shlomo benartzi. Retreived from: www.shlomobenartzi.com/save-more-tomorrow*.

The level of output per worker depends on both the level of physical capital per worker, K/N, and the level of human capital per worker, H/N. As before, an increase in capital per worker K/N leads to an increase in output per worker. And an increase in the average level of skill H/N also leads to more output per worker. More skilled workers can do more complex tasks; they can deal more easily with unexpected complications. All of this leads to higher output per worker.

Note that we are using the same symbol, H, to denote central bank money in Chapter 4 and human capital in this chapter. Both uses are traditional. Do not be confused.

We assumed previously that increases in physical capital per worker increased output per worker, but that the effect became smaller as the level of capital per worker increased. We can make the same assumption for human capital per worker: Think of increases in H/N as coming from increases in the number of years of education. The evidence is that the returns to increasing the proportion of children acquiring a primary education are large. At the very least, the ability to read and write allows people to use equipment that is more complicated and more productive. For rich countries, however, primary education—and, for that matter, secondary education—are no longer the relevant margin. Most children now get both. The relevant margin is now higher education. We are sure it will come as good news to you that the evidence shows that higher education increases people's skills, at least as measured by the higher wages of those who acquire it. But to take an extreme example, it is not clear that forcing everyone to acquire an advanced college degree would increase aggregate output much. Many people would end up overqualified and probably more frustrated rather than more productive.

How should we construct the measure for human capital, H? The same way we construct the measure for physical capital, K. To construct K, we just add the values of the different pieces of capital, so that a machine that costs $2,000 gets twice the weight of a machine that costs $1,000. Similarly, we construct the measure of H such that workers who are paid twice as much get twice the weight. Take, for example, an economy with 100 workers, half of them unskilled and half of them skilled. Suppose the wage of the skilled workers is twice that of the unskilled workers. We can then construct H as $[(50 \times 1) + (50 \times 2)] = 150$. Human capital per worker, H/N, is then equal to $150/100 = 1.5$.

The rationale for using relative wages as weights is that they reflect relative marginal products. A worker who is paid three times as much as another is assumed to have a marginal product whose value is three times higher.

An issue, however, is whether relative wages accurately reflect relative marginal products. To take a controversial example, in the same job, with the same seniority, women still often earn less than men. Is it because their marginal product is lower? Should they be given a lower weight than men in the construction of human capital?

Human Capital, Physical Capital, and Output

How does the introduction of human capital change the analysis of the previous sections?

Our conclusions about *physical capital accumulation* remain valid. An increase in the saving rate increases steady-state physical capital per worker and therefore increases output per worker. But our conclusions now extend to *human capital accumulation* as well. An increase in how much society "saves" in the form of human capital—through education and on-the-job training—increases steady-state human capital per worker, which leads to an increase in output per worker. Our extended model gives us a richer picture of how output per worker is determined. In the long run, it tells us that output per worker depends on both how much society saves and how much it spends on education.

What is the relative importance of human capital and of physical capital in the determination of output per worker? A place to start is to compare how much is spent on formal education to how much is invested in physical capital. In the United States, spending on formal education is about 6.5% of GDP. This number includes both government expenditures on education and people's private expenditures on education. It is between one-third and one-half of the gross investment rate for physical capital (which is around 17%). But this comparison is only a first pass. Consider the following complications:

- Education, especially higher education, is partly consumption—done for its own sake—and partly investment. We should include only the investment part for our purposes, but the 6.5% number in the preceding paragraph includes both.

How large is your opportunity cost relative to your tuition?

- At least for postsecondary education, the opportunity cost of a person's education is his or her forgone wages while acquiring the education. Spending on education should include not only the actual cost of education but also this opportunity cost. The 6.5% number does not include this opportunity cost.
- Formal education is only a part of education. Much of what we learn comes from on-the-job training, formal or informal. Both the actual costs and the opportunity costs of on-the-job training should also be included. The 6.5% number does not include the costs associated with on-the-job training.
- We should compare investment rates net of depreciation. Depreciation of physical capital, especially of machines, is likely to be higher than depreciation of human capital. Skills deteriorate, but slowly. And unlike physical capital, they deteriorate less quickly (or indeed improve) the more they are used.

For all these reasons, it is difficult to come up with reliable numbers for investment in human capital. Recent studies conclude that investment in physical capital and in education play roughly similar roles in the determination of output. This implies that output per worker depends roughly equally on the amounts of physical capital and human capital in the economy. Countries that save more or spend more on education can achieve substantially higher steady-state levels of output per worker.

Endogenous Growth

Note what the conclusion we just reached did and did not say. It did say that a country that saves more or spends more on education will achieve a *higher level* of output per worker in steady state. It did not say that by saving or spending more on education a country can sustain permanently *higher growth* of output per worker.

Robert Lucas was awarded the Nobel Prize in 1995. He teaches at the University of Chicago. Paul Romer was awarded the Nobel Prize in 2018. He teaches at New York University.

However, following the lead of Robert Lucas and Paul Romer, researchers have explored the possibility that the joint accumulation of physical capital and human capital might actually be enough to sustain growth. Given human capital, increases in physical capital will run into decreasing returns. And given physical capital, increases in human capital will also run into decreasing returns. But these researchers have asked, what if both physical and human capital increase in tandem? Can't an economy grow forever just by steadily having more capital and more skilled workers?

Models that generate steady growth even without technological progress are called **models of endogenous growth**—in contrast to the model we saw in previous sections of this chapter—the growth rate depends, even in the long run, on variables such as the saving rate and the rate of spending on education. The jury on this class of models is still out, but the indications so far are that the conclusions we drew earlier need to be qualified and not abandoned. The current consensus is as follows:

- Output per worker depends on the level of both physical capital per worker and human capital per worker. Both forms of capital can be accumulated, one through physical investment, the other through education and training. Increasing either the saving rate or the fraction of output spent on education and training can lead to much higher levels of output per worker in the long run. However, given the rate of technological progress, such measures do not lead to a permanently higher growth rate.
- Note the qualifier in the last proposition: *given the rate of technological progress*. Is technological progress unrelated to the level of human capital in the economy? Can't a better educated labor force lead to a higher rate of technological progress? These questions take us to the topic of the next chapter, the sources and effects of technological progress.

SUMMARY

- In the long run, the evolution of output is determined by two relations. (To make the reading of this summary easier, we shall omit "per worker" in what follows.) First, the level of output depends on the amount of capital. Second, capital accumulation depends on the level of output, which determines saving and investment.
- These interactions between capital and output imply that, starting from any level of capital (and ignoring technological progress, the topic of Chapter 12), an economy converges in the long run to a *steady-state* (constant) level of capital. Associated with this level of capital is a steady-state level of output.
- The steady-state level of capital, and thus the steady-state level of output, depends positively on the saving rate. A higher saving rate leads to a higher steady-state level of output; during the transition to the new steady state, a higher saving rate leads to positive output growth. But (again ignoring technological progress) in the long run, the growth rate of output is equal to zero and so does not depend on the saving rate.
- An increase in the saving rate requires an initial decrease in consumption. In the long run, the increase in the saving rate may lead to an increase or a decrease in consumption, depending on whether the economy is below or above the *golden-rule level of capital*, which is the level of capital at which steady-state consumption is highest.
- Most countries have a level of capital below the golden-rule level. Thus, an increase in the saving rate leads to an initial decrease in consumption followed by an increase in consumption in the long run. When considering whether to adopt policy measures aimed at changing a country's saving rate, policymakers must decide how much weight to put on the welfare of current generations versus the welfare of future generations.
- Although most of the analysis of this chapter focuses on the effects of physical capital accumulation, output depends on the levels of both physical *and* human capital. Both forms of capital can be accumulated, one through investment, the other through education and training. Increasing the saving rate or the fraction of output spent on education and training can lead to large increases in output in the long run.

KEY TERMS

saving rate, 219
steady state, 225
golden-rule level of capital, 229
fully funded social security system, 231
pay-as-you-go social security system, 231
Social Security trust fund, 231
human capital, 236
models of endogenous growth, 238

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The saving rate is always equal to the investment rate.
b. A higher investment rate can sustain higher growth of output forever.
c. If capital never depreciated, growth could go on forever.
d. The higher the saving rate, the higher consumption in steady state.
e. We should transform Social Security from a pay-as-you-go system to a fully funded system. This would increase consumption both now and in the future.
f. The US government should give tax breaks for saving because the US capital stock is far below the golden-rule level.
g. Education increases human capital and thus output. It follows that governments should subsidize education.

2. Consider the following statement: "The Solow model shows that the saving rate does not affect the growth rate in the long run, so we should stop worrying about the low US saving rate. Increasing the saving rate wouldn't have any important effects on the economy." Explain why you agree or disagree with this statement?

3. In Chapter 3 we saw that an increase in the saving rate can lead to a recession in the short run (i.e., the paradox of saving). We examined the issue in the medium run in Problem 5 at the end of Chapter 7. We can now examine the long-run effects of an increase in saving.

Using the model presented in this chapter, what is the effect of an increase in the saving rate on output per worker likely to be after one decade? After five decades?

DIG DEEPER

4. Discuss how the level of output per person in the long run would likely be affected by each of the following changes:

a. The right to exclude saving from income when paying income taxes.
b. A higher rate of female participation in the labor market (but constant population).

5. Suppose the United States moved from the current pay-as-you-go Social Security system to a fully funded one and financed the transition without additional government borrowing. How would the shift to a fully funded system affect the level and the rate of growth of output per worker in the long run?

6. *Suppose that the production function is given by*

$$Y = 0.5\sqrt{K}\sqrt{N}$$

a. Derive the steady-state levels of output per worker and capital per worker in terms of the saving rate, s, and the depreciation rate, δ.
b. Derive the equation for steady-state output per worker and steady-state consumption per worker in terms of s and δ.
c. Suppose that $\delta = 0.05$. With your favorite spreadsheet software, compute steady-state output per worker and steady-state consumption per worker when $s = 0.2$; $s = 1$. Explain the intuition behind your results.
d. Use your favorite spreadsheet software to graph the steady-state level of output per worker and the steady-state level of consumption per worker as a function of the saving rate (i.e., measure the saving rate on the horizontal axis of your graph and the corresponding values of output per worker and consumption per worker on the vertical axis).
e. Does the graph show that there is a value of s that maximizes output per worker? Does the graph show that there is a value of s that maximizes consumption per worker? If so, what is this value?

7. *The Cobb-Douglas production function and the steady state.*

This problem is based on the material in the chapter appendix. Suppose that the economy's production function is given by

$$Y = K^{\alpha}N^{1-\alpha}$$

and assume that $\alpha = 1/3$

a. Is this production function characterized by constant returns to scale? Explain.
b. Are there decreasing returns to capital?
c. Are there decreasing returns to labor?
d. Transform the production function into a relation between output per worker and capital per worker.
e. For a given saving rate, s, and depreciation rate, δ, give an expression for capital per worker in the steady state.
f. Give an expression for output per worker in the steady state.
g. Solve for the steady-state level of output per worker when $s = 0.32$ and $\delta = 0.08$.
h. Suppose that the depreciation rate remains constant at $\delta = 0.08$, while the saving rate is reduced by half, to $s = 0.16$. What is the new steady-state output per worker?

8. *Continuing with the logic from Problem 7, suppose that the economy's production function is given by* $Y = K^{1/3}N^{2/3}$ *and that the saving rate, s, and the depreciation rate,* δ *are equal to 0.10.*

a. What is the steady-state level of capital per worker?
b. What is the steady-state level of output per worker?

Suppose that the economy is in steady state and that, in period t, the depreciation rate increases permanently from 0.10 to 0.20.

c. What will be the new steady-state levels of capital per worker and output per worker?
d. Compute the path of capital per worker and output per worker over the first three periods after the change in the depreciation rate.

9. *Deficits and the capital stock*

For the production function, $Y = \sqrt{K}\sqrt{N}$ *equation (11.9) gives the solution for the steady-state capital stock per worker.*

a. Retrace the steps in the text that derive equation (11.9).
b. Suppose that the saving rate, s, is initially 15% per year, and the depreciation rate, δ, is 7.5%. What is the steady-state capital stock per worker? What is steady-state output per worker?
c. Suppose that there is a government deficit of 5% of GDP and that the government eliminates this deficit. Assume that private saving is unchanged so that total saving increases to 20%. What is the new steady-state capital stock per worker? What is the new steady-state output per worker? How does this compare to your answer to part b?

EXPLORE FURTHER

10. *US saving and government deficits*

This question continues the logic of Problem 9 to explore the implications of the US government budget deficit for the long-run capital stock. The question assumes that the United States will have a budget deficit over the life of this edition of the text.

a. The World Bank reports gross domestic saving rate by country and year. The Web site is https://data.worldbank.org/indicator/NY.GNS.ICTR.ZS. Find the most recent number for the United States. What is the total saving rate in the United States as a percentage of GDP? Using the depreciation rate and the logic from Problem 9, what would be the steady-state capital stock per worker? What would be steady-state output per worker?
b. Go to the most recent *Economic Report of the President* (ERP) and find the most recent federal deficit as a percentage of GDP. In the 2018 ERP, this is found in Table B-18. Using the reasoning from Problem 9, suppose that the federal budget deficit was eliminated and there was no change in private saving. What would be the effect on the long-run capital stock per worker? What would be the effect on long-run output per worker?
c. Return to the World Bank table of gross domestic saving rates. How does the saving rate in China compare to the saving rate in the United States?

FURTHER READINGS

- The classic treatment of the relation between the saving rate and output is by Robert Solow, *Growth Theory: An Exposition* (1970).
- An easy-to-read discussion of whether and how to increase saving and improve education in the United States is given in Memoranda 23 to 27 in *Memos to the President: A Guide through Macroeconomics for the Busy Policymaker*, by Charles Schultze, who was the Chairman of the Council of Economic Advisers during the Carter administration (1992).

APPENDIX: The Cobb-Douglas Production Function and the Steady State

In 1928, Charles Cobb (a mathematician) and Paul Douglas (an economist, who went on to become a US senator) concluded that the following production function gave a good description of the relation between output, physical capital, and labor in the United States from 1899 to 1922:

$$Y = K^{\alpha}N^{1-\alpha} \tag{11.A1}$$

with α being a number between zero and one. Their findings proved surprisingly robust. Even today, the production function (11.A1), now known as the **Cobb-Douglas production function**, still gives a good description of the relation between output, capital, and labor in the United States, and it has become a standard tool in the economist's toolbox. (Verify for yourself that it satisfies the two properties we discussed in the text: constant returns to scale and decreasing returns to capital and to labor.)

The purpose of this appendix is to characterize the steady state of an economy when the production function is given by (11.A1). (All you need to follow the steps is knowledge of the properties of exponents.)

Recall that, in steady state, saving per worker must be equal to depreciation per worker. Let's see what this implies.

- To derive saving per worker, we must first derive the relation between output per worker and capital per worker implied by equation (11.A1). Divide both sides of equation (11.A1) by N:

  $$Y/N = K^{\alpha}N^{1-\alpha}/N$$

  Using the properties of exponents,

  $$N^{1-\alpha}/N = N^{1-\alpha}N^{-1} = N^{-\alpha}$$

  so replacing the terms in N in the preceding equation, we get:

  $$Y/N = K^{\alpha}N^{-\alpha} = (K/N)^{\alpha}$$

  Output per worker, Y/N, is equal to the ratio of capital per worker, K/N, raised to the power α.

  Saving per worker is equal to the saving rate times output per worker, so, using the previous equation, it is equal to

  $$s(K^*/N)^{\alpha}$$

- Depreciation per worker is equal to the depreciation rate times capital per worker:

  $$\delta(K^*/N)$$

- The steady-state level of capital, K^*, is determined by the condition that saving per worker be equal to depreciation per worker, so:

  $$s(K^*/N)^{\alpha} = \delta(K^*/N)$$

  To solve this expression for the steady-state level of capital per worker K^*/N, divide both sides by $(K^*/N)^{\alpha}$:

  $$s = \delta(K^*/N)^{1-\alpha}$$

  Divide both sides by δ, and change the order of the equality:

  $$(K^*/N)^{1-\alpha} = s/\delta$$

  Finally, raise both sides to the power $1/(1-\alpha)$:

  $$(K^*/N) = (s/\delta)^{1/(1-\alpha)}$$

  This gives us the steady-state level of capital per worker.

  From the production function, the steady-state level of output per worker is then equal to

  $$(Y^*/N) = (K/N)^{\alpha} = (s/\delta)^{\alpha/(1-\alpha)}$$

  Let's see what this last equation implies.

- In the text, we worked with a special case of an equation (11.A1), the case where $\alpha = 0.5$. (Taking a variable to the power 0.5 is the same as taking the square root of the variable.) If $\alpha = 0.5$, the preceding equation means

  $$Y^*/N = s/\delta$$

  Output per worker is equal to the ratio of the saving rate to the depreciation rate. This is the equation we discussed in the text. A doubling of the saving rate leads to a doubling in steady-state output per worker.

- The empirical evidence suggests, however, that, if we think of K as physical capital, α is closer to one-third than to one-half. Assuming $\alpha = 1/3$, then $\alpha(1-\alpha) = (1/3)/(1-(1/3)) = (1/3)/(2/3) = 1/2$, and the equation for output per worker yields

  $$Y^*/N = (s/\delta)^{1/2} = \sqrt{s/\delta}$$

  This implies smaller effects of the saving rate on output per worker than was suggested by the computations in the text. A doubling of the saving rate, for example, means that output per worker increases by a factor of $\sqrt{2}$ or only about 1.4.

- There is, however, an interpretation of our model in which the appropriate value of α is close to 1/2, so the computations in the text are applicable. If, along the lines of Section 11-4, we take human capital into account as well as physical capital, then a value of α around 1/2 for the contribution of this broader definition of capital to output is, indeed, roughly appropriate. Thus, one interpretation of the numerical results in Section 11-3 is that they show the effects of a given saving rate, but that saving must be interpreted to include saving in both physical capital and in human capital (more machines and more education).

Key Term

Cobb-Douglas production function, 241

This page is intentionally left blank

12

Technological Progress and Growth

The conclusion in Chapter 11 that capital accumulation cannot by itself sustain growth has a straightforward implication: Sustained growth *requires* technological progress. This chapter looks at the role of technological progress in growth.

Section 12-1 looks at the roles of technological progress and capital accumulation in growth.

Section 12-2 turns to the determinants of technological progress, the role of research and development (R&D), and the role of innovation versus imitation.

Section 12-3 discusses why some countries are able to achieve steady technological progress while others do not. In so doing, it looks at the role of institutions in sustaining growth.

> **If you remember one basic message from this chapter, it should be: Sustained growth requires sustained technological progress. Technological progress depends both on innovation and on institutions.** ▶▶▶

12-1 TECHNOLOGICAL PROGRESS AND THE RATE OF GROWTH

In an economy in which there is both capital accumulation and technological progress, at what rate will output grow? To answer this question, we need to extend the model developed in Chapter 11 to allow for technological progress. To do so, we must first revisit the aggregate production function.

Technological Progress and the Production Function

Technological progress has many dimensions:

The average number of items carried by a supermarket increased from 2,200 in 1950 to 38,700 in 2010. To get a sense of what this means, see Robin Williams (who plays an immigrant from the Soviet Union) in the supermarket scene in the movie *Moscow on the Hudson*. ▶

- It can lead to larger quantities of output for given quantities of capital and labor. Think of a new type of lubricant that allows a machine to run at a higher speed and to increase production.
- It can lead to better products. Think of the steady improvement over time in automobile safety and comfort.
- It can lead to new products. Think of the introduction of the iPad, wireless communication technology, flat screen monitors, and self-driving cars.
- It can lead to a larger variety of products. Think of the steady increase in the number of breakfast cereals available at your local supermarket.

As you saw in the Focus Box "Real GDP, Technological Progress, and the Price of Computers" in Chapter 2, thinking of products as providing a number of underlying services is the method used to construct the price index for computers. ▶

These dimensions are more similar than they appear. If we think of consumers as caring not about the goods themselves but about the services these goods provide, then they all have something in common: In each case, consumers receive more services. A better car provides more safety, a new product such as an iPad or faster communication technology provides more communication services, and so on. If we think of output as the underlying services provided by the goods produced in the economy, we can think of technological progress as leading to increases in output for given amounts of capital and labor. We can then think of the **state of technology** as a variable that tells us how much output can be produced from given amounts of capital and labor at any time. If we denote the state of technology by A, we can rewrite the production function as

$$Y = \underset{(+,+,+)}{F(K, N, A)}$$

For simplicity, we shall ignore human capital here. We return to it later in the chapter. ▶

This is our extended production function. Output depends on both capital and labor (K and N) and on the state of technology (A). Given capital and labor, an improvement in the state of technology, A, leads to an increase in output.

It will be convenient to use a more restrictive form of the preceding equation, namely

$$Y = F(K, AN) \tag{12.1}$$

This equation states that production depends on capital and on labor multiplied by the state of technology. Introducing the state of technology in this way makes it easier to think about the effect of technological progress on the relation between output, capital, and labor. Equation (12.1) implies that we can think of technological progress in two equivalent ways:

- Technological progress *reduces* the number of workers needed to produce a given amount of output. If A doubles, firms can produce the same quantity of output with only half the original number of workers, N.
- Technological progress *increases* the output that can be produced with a given number of workers. We can think of AN as the amount of **effective labor** in the

economy. If the state of technology A doubles, it is as if the economy had twice as many workers. In other words, we can think of output being produced by two factors: capital (K), and effective labor (AN).

AN is also sometimes called **labor in efficiency units**. The use of *efficiency* for "efficiency units" here and for "efficiency wages" in Chapter 6 is a coincidence; the two notions are unrelated.

What restrictions should we impose on the extended production function (12.1)? We can build directly here on our discussion in Chapter 11.

Again, it is reasonable to assume constant returns to scale. *For a given state of technology* (A), doubling both the amount of capital (K) and the amount of labor (N) is likely to lead to a doubling of output

$$F(2K, 2AN) = 2Y$$

More generally, for any number x,

$$F(x K, x AN) = xY$$

It is also reasonable to assume decreasing returns to each of the two factors—capital and effective labor. Given effective labor, an increase in capital is likely to increase output but at a decreasing rate. Symmetrically, given capital, an increase in effective labor is likely to increase output, but at a decreasing rate.

It was convenient in Chapter 11 to think in terms of output *per worker* and capital *per worker*. That was because the steady state of the economy was a state where output *per worker* and capital *per worker* were constant. It is convenient here to look at output *per effective worker* and capital *per effective worker*. The reason is the same; as we shall soon see, in steady state, output *per effective worker* and capital *per effective worker* are constant.

Per worker: divided by the number of workers (N). Per effective worker: divided by the number of effective workers (AN)—the number of workers, N, times the state of technology, A.

To get a relation between output per effective worker and capital per effective worker, take $x = 1/AN$ in the preceding equation. This gives

$$\frac{Y}{AN} = F\left(\frac{K}{AN}, 1\right)$$

Or, if we define the function f so that $f(K/AN) = F(K/AN, 1)$:

$$\frac{Y}{AN} = f\left(\frac{K}{AN}\right) \qquad (12.2)$$

In words: *Output per effective worker* (the left side) is a function of *capital per effective worker* (the expression in the function on the right side).

Suppose that F has the double square root form:

$$Y = F(K, AN) = \sqrt{K}\sqrt{AN}$$

Then

$$\frac{Y}{AN} = \frac{\sqrt{K}\sqrt{AN}}{AN} = \frac{\sqrt{K}}{\sqrt{AN}}$$

So, in this case, the function f is simply the square root function:

$$f\left(\frac{K}{AN}\right) = \sqrt{\frac{K}{AN}}$$

The relation between output per effective worker and capital per effective worker is drawn in Figure 12-1. It looks much the same as the relation we drew in Figure 11–2

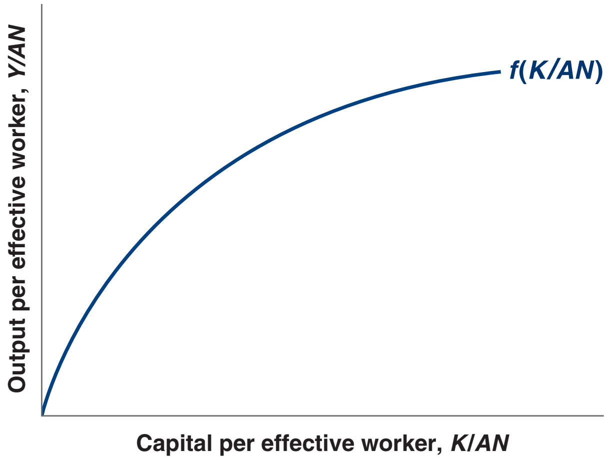


Figure 12-1

Output per Effective Worker versus Capital per Effective Worker

Because of decreasing returns to capital, increases in capital per effective worker lead to smaller and smaller increases in output per effective worker.

A simple key to understanding the results in this section: The results we derived for *output per worker* in Chapter 11 still hold in this chapter, but now for *output per effective worker*. For example, in Chapter 11, we saw that output per worker was constant in steady state. In this chapter, we shall see that output per effective worker is constant in steady state. And so on.

between output per worker and capital per worker in the absence of technological progress. There, increases in K/N led to increases in Y/N, but at a decreasing rate. Here, increases in K/AN lead to increases in Y/AN, but at a decreasing rate.

Interactions between Output and Capital

We now have the elements we need to think about the determinants of growth. Our analysis will parallel the analysis of Chapter 11. There we looked at the dynamics of *output per worker* and *capital per worker*. Here we look at the dynamics of *output per effective worker* and *capital per effective worker*.

In Chapter 11, we characterized the dynamics of output and capital per worker using Figure 11–2, where we drew three relations:

- The relation between output per worker and capital per worker.
- The relation between investment per worker and capital per worker.
- The relation between depreciation per worker—equivalently, the investment per worker needed to maintain a constant level of capital per worker—and capital per worker.

The dynamics of capital per worker and, by implication, output per worker, were determined by the relation between investment per worker and depreciation per worker. Depending on whether investment per worker was greater or smaller than depreciation per worker, capital per worker increased or decreased over time, as did output per worker.

We shall follow the same approach in building Figure 12-2. The difference is that we focus on output, capital, and investment *per effective worker* rather than per worker.

- The relation between output per effective worker and capital per effective worker was derived in Figure 12-1. This relation is repeated in Figure 12-2; output per effective worker increases with capital per effective worker, but at a decreasing rate.
- Under the same assumptions as in Chapter 11—that investment is equal to private saving, and the private saving rate is constant—investment is given by

$$I = S = s\,Y$$

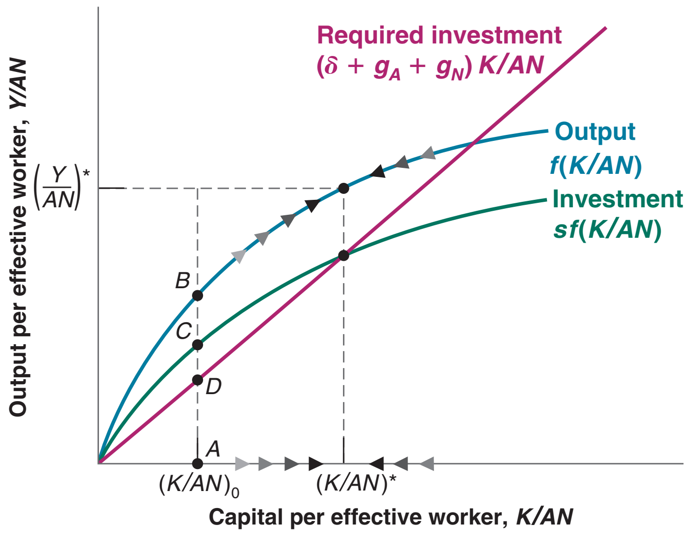
Divide both sides by the number of effective workers, AN, to get

$$\frac{I}{AN} = s\frac{Y}{AN}$$

Figure 12-2

The Dynamics of Capital per Effective Worker and Output per Effective Worker

Capital per effective worker and output per effective worker converge to constant values in the long run.



Replacing output per effective worker, Y/AN, by its expression from equation (12.2) gives

$$\frac{I}{AN} = sf\left(\frac{K}{AN}\right)$$

The relation between investment per effective worker and capital per effective worker is drawn in Figure 12-2. It is equal to the upper curve—the relation between output per effective worker and capital per effective worker—multiplied by the saving rate, s. This gives us the lower curve.

- Finally, we need to ask what level of investment per effective worker is needed to maintain a given level of capital per effective worker.

  In Chapter 11, the answer was: For capital to be constant, investment had to be equal to the depreciation of the existing capital stock. Here, the answer is slightly more complicated. The reason is as follows: Now that we allow for technological progress (so A increases over time), the number of effective workers (AN) increases over time. Thus, maintaining the same ratio of capital to effective workers (K/AN) requires an increase in the capital stock (K) proportional to the increase in the number of effective workers (AN). Let's look at this condition more closely.

In Chapter 11, we assumed $g_A = 0$ and $g_N = 0$. Our focus in this chapter is on the implications of technological progress, $g_A > 0$. Once we allow for technological progress, introducing population growth $g_N > 0$ is straightforward. Thus, we allow for both $g_A > 0$ and $g_N > 0$.

  Let δ be the depreciation rate of capital. Let the rate of technological progress be equal to g_A. Let the rate of population growth be equal to g_N. If we assume that the ratio of employment to the total population remains constant, the number of workers (N) also grows at annual rate g_N. Together, these assumptions imply that the growth rate of effective labor (AN) equals $g_A + g_N$. For example, if the number of workers is growing at 1% per year and the rate of technological progress is 2% per year, then the growth rate of effective labor is equal to 3% per year.

The growth rate of the product of two variables is the sum of the growth rates of the two variables. See Proposition 7 in Appendix 2 at the end of the book.

  These assumptions imply that the level of investment needed to maintain a given level of capital per effective worker is therefore given by

$$I = \delta K + (g_A + g_N)K$$

Or, equivalently,

$$I = (\delta + g_A + g_N)K \qquad (12.3)$$

  An amount δK is needed just to keep the capital stock constant. If the depreciation rate is 10%, then investment must be equal to 10% of the capital stock to maintain the same level of capital. And an additional amount $(g_A + g_N)K$ is needed to ensure that the capital stock increases at the same rate as effective labor. If effective labor increases at 3% per year, for example, then capital must increase by 3% per year to maintain the same level of capital per effective worker. Putting δK and $(g_A + g_N)K$ together in this example: If the depreciation rate is 10% and the growth rate of effective labor is 3%, then investment must equal 13% of the capital stock to maintain a constant level of capital per effective worker.

  Dividing the previous expression by the number of effective workers to get the amount of investment per effective worker needed to maintain a constant level of capital per effective worker gives

$$\frac{I}{AN} = (\delta + g_A + g_N)\frac{K}{AN}$$

  The level of investment per effective worker needed to maintain a given level of capital per effective worker is represented by the upward-sloping line "Required investment" in Figure 12-2. The slope of the line equals $(\delta + g_A + g_N)$.

Dynamics of Capital and Output

We can now give a graphical description of the dynamics of capital per effective worker and output per effective worker.

Consider a given level of capital per effective worker, say $(K/AN)_0$ in Figure 12-2. At that level, output per effective worker equals the vertical distance AB. Investment per effective worker is equal to AC. The amount of investment required to maintain that level of capital per effective worker is equal to AD. Because actual investment exceeds the investment level required to maintain the existing level of capital per effective worker, K/AN increases.

Hence, starting from $(K/AN)_0$, the economy moves to the right, with the level of capital per effective worker increasing over time. This goes on until investment per effective worker is just sufficient to maintain the existing level of capital per effective worker, until capital per effective worker equals $(K/AN)^*$.

In the long run, capital per effective worker reaches a constant level, and so does output per effective worker. Put another way, the steady state of this economy is such that *capital per effective worker and output per effective worker are constant and equal to* $(K/AN)^*$ *and* $(Y/AN)^*$*, respectively.*

If Y/AN is constant, Y must grow at the same rate as AN. So, it must grow at rate $g_A + g_N$.

This implies that, in steady state, output (Y) is growing at the same rate as effective labor (AN), so that the ratio of the two is constant. Because effective labor grows at rate $(g_A + g_N)$, output growth in steady state must also equal $(g_A + g_N)$. The same reasoning applies to capital. Because capital per effective worker is constant in steady state, capital is also growing at rate $(g_A + g_N)$.

Stated in terms of capital or output per effective worker, these results seem rather abstract. But it is straightforward to state them in a more intuitive way, and this gives us our first important conclusion:

In steady state, the growth rate of output equals the rate of population growth (g_N) *plus the rate of technological progress* (g_A)*. By implication, the growth rate of output is independent of the saving rate.*

To strengthen your intuition, let's go back to the argument we used in Chapter 11 to show that, in the absence of technological progress and population growth, the economy could not sustain positive growth forever.

- The argument went as follows: Suppose the economy tried to sustain positive output growth. Because of decreasing returns to capital, capital would have to grow faster than output. The economy would have to devote a larger and larger proportion of output to capital accumulation. At some point there would be no more output to devote to capital accumulation. Growth would come to an end.
- Exactly the same logic is at work here. Effective labor grows at rate $(g_A + g_N)$. Suppose the economy tried to sustain output growth in excess of $(g_A + g_N)$. Because of decreasing returns to capital, capital would have to increase faster than output. The economy would have to devote a larger and larger proportion of output to capital accumulation. At some point this would prove impossible. Thus, the economy cannot permanently grow faster than $(g_A + g_N)$.

The growth rate of Y/N is equal to the growth rate of Y minus the growth rate of N (see Proposition 8 in Appendix 2 at the end of the book). So the growth rate of Y/N is given by $(g_Y - g_N) = (g_A + g_N) - g_N = g_A$.

We have focused on the behavior of aggregate output. To get a sense of what happens not to aggregate output but rather to the standard of living over time, we must look instead at the behavior of output per worker (not output per *effective* worker). Because output grows at rate $(g_A + g_N)$ and the number of workers grows at rate g_N, output per worker grows at rate g_A. In other words, *when the economy is in steady state, output per worker grows at the rate of technological progress.*

Because output, capital, and effective labor all grow at the same rate $(g_A + g_N)$ in steady state, the steady state of this economy is also called a state of **balanced growth**.

Table 12-1	The Characteristics of Balanced Growth	
		Growth Rate
1	Capital per effective worker	0
2	Output per effective worker	0
3	Capital per worker	g_A
4	Output per worker	g_A
5	Labor	g_N
6	Capital	$g_A + g_N$
7	Output	$g_A + g_N$

In steady state, output and the two inputs, capital and effective labor, grow "in balance" at the same rate. The characteristics of balanced growth will be helpful later in the chapter and are summarized in Table 12-1.

On the balanced growth path (equivalently: in steady state; equivalently: in the long run):

- *Capital per effective worker* and *output per effective worker* are constant; this is the result we derived in Figure 12-2.
- Equivalently, *capital per worker* and *output per worker* are growing at the rate of technological progress, g_A.
- Or in terms of labor, capital, and output: *Labor* is growing at the rate of population growth, g_N; *capital* and *output* are growing at a rate equal to the sum of population growth and the rate of technological progress, $(g_A + g_N)$.

Back to the Effects of the Saving Rate

In steady state, the growth rate of output depends *only* on the rate of population growth and the rate of technological progress. Changes in the saving rate do not affect the steady-state growth rate. But changes in the saving rate do increase the steady-state level of output per effective worker.

This result is best seen in Figure 12-3, which shows the effect of an increase in the saving rate from s_0 to s_1. The increase in the saving rate shifts the investment relation up, from $s_0 f(K/AN)$ to $s_1 f(K/AN)$. It follows that the steady-state level of capital per

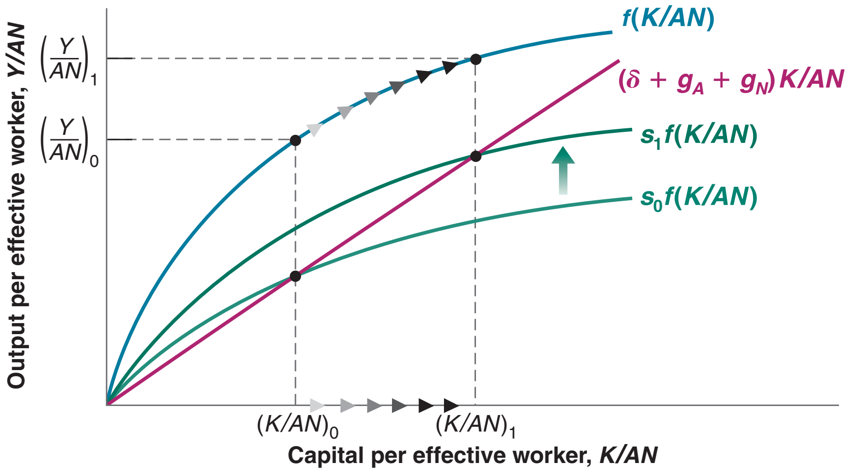


Figure 12-3

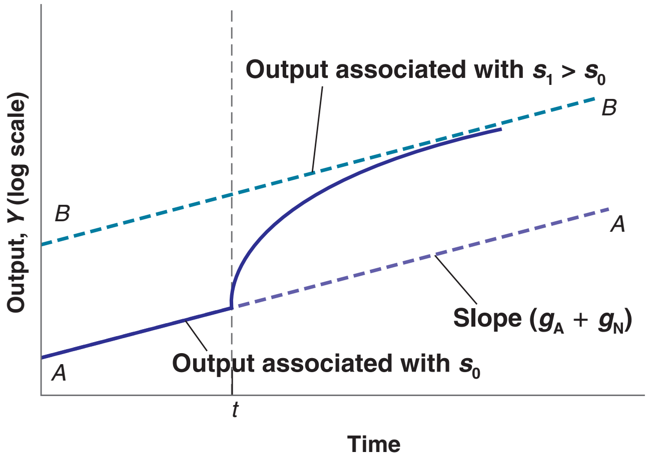
The Effects of an Increase in the Saving Rate: I

An increase in the saving rate leads to an increase in the steady-state levels of output per effective worker and capital per effective worker.

Figure 12-4

The Effects of an Increase in the Saving Rate: II

The increase in the saving rate leads to higher growth until the economy reaches its new, higher, balanced growth path.



effective worker increases from $(K/AN)_0$ to $(K/AN)_1$, with a corresponding increase in the level of output per effective worker from $(Y/AN)_0$ to $(Y/AN)_1$.

Figure 12-4 is the same as Figure 11-5, which anticipated the derivation presented here.

For a description of logarithmic scales, see Appendix 2 at the end of the book. When a logarithmic scale is used, a variable growing at a constant rate moves along a straight line. The slope of the line is equal to the rate of growth of the variable.

Following the increase in the saving rate, capital per effective worker and output per effective worker increase for some time as they converge to their new higher level. Figure 12-4 plots output against time. Output is measured on a logarithmic scale. The economy is initially on the balanced growth path AA: Output is growing at rate $(g_A + g_N)$—so the slope of AA is equal to $(g_A + g_N)$. After the increase in the saving rate at time t, output grows faster for some period of time. Eventually, output ends up at a higher level than it would have been without the increase in saving. But its growth rate returns to $g_A + g_N$. In the new steady state, the economy grows at the same rate, but on a higher growth path, BB, which is parallel to AA and also has a slope equal to $(g_A + g_N)$.

Let's summarize: In an economy with technological progress and population growth, output grows over time. In steady state, output *per effective worker* and capital *per effective worker* are constant. Put another way, output *per worker* and capital *per worker* grow at the rate of technological progress. Put yet another way, output and capital grow at the same rate as effective labor, and therefore at a rate equal to the growth rate of the number of workers plus the rate of technological progress. When the economy is in steady state, it is said to be on a balanced growth path.

The rate of output growth in steady state is independent of the saving rate. However, the saving rate affects the steady-state level of output per effective worker. And increases in the saving rate lead, for some time, to an increase in the growth rate above the steady-state growth rate.

12-2 THE DETERMINANTS OF TECHNOLOGICAL PROGRESS

We have just seen that the growth rate of output per worker is ultimately determined by the rate of technological progress. This leads naturally to the next question: What determines the rate of technological progress? This is the question we take up in this and the next sections.

The term *technological progress* brings to mind images of major discoveries—the invention of the microchip, the discovery of the structure of DNA, and so on. These discoveries suggest a process driven largely by scientific research and chance rather than by economic forces. But the truth is that most technological progress in modern advanced economies is the result of a humdrum process: the outcome of firms' **research and development (R&D)** activities. Industrial R&D expenditures account for between

2% and 3% of GDP in each of the four major rich countries we looked at in Chapter 10 (the United States, France, Japan, and the United Kingdom). About 75% of the roughly one million US scientists and researchers working in R&D are employed by firms. US firms' R&D spending equals more than 20% of their spending on gross investment, and more than 60% of their spending on net investment—gross investment less depreciation.

Firms spend on R&D for the same reason they buy new machines or build new plants: to increase profits. By increasing spending on R&D, a firm increases the probability that it will discover and develop a new product. (We shall use *product* as a generic term to denote new goods or new techniques of production.) If the new product is successful, the firm's profits will increase. There is, however, an important difference between purchasing a machine and spending more on R&D. The difference is that the outcome of R&D is fundamentally *ideas*. And unlike a machine, an idea can potentially be used by many firms at the same time. A firm that has just acquired a new machine does not have to worry that another firm will use that particular machine. A firm that has discovered and developed a new product can make no such assumption.

This last point implies that the level of R&D spending depends not only on the **fertility of research**—how spending on R&D translates into new ideas and new products—but also on the **appropriability** of research results, which is the extent to which firms can benefit from the results of their own R&D. Let's look at each aspect in turn.

The Fertility of the Research Process

If research is fertile—that is, if R&D spending leads to many new products—then, other things being equal, firms will have strong incentives to spend on R&D: R&D spending and, by implication, technological progress will be high. The determinants of the fertility of research lie largely outside the realm of economics. Many factors interact here.

The fertility of research depends on the successful interaction between basic research (the search for general principles and results) and applied research and development (the application of these results to specific uses, and the development of new products). Basic research does not by itself lead to technological progress. But the success of applied R&D depends ultimately on basic research. Much of the computer industry's development can be traced to a few breakthroughs, from the invention of the transistor to the invention of the microchip. On the software side, much of the progress comes from progress in mathematics. For example, progress in encryption comes from progress in the theory of prime numbers.

Some countries appear more successful at basic research; other countries are more successful at applied research and development. Studies point to differences in the education system as one of the reasons why. For example, it is often argued that the French higher education system, with its strong emphasis on abstract thinking, produces researchers who are better at basic research than at applied research. Studies also point to the importance of a "culture of entrepreneurship," in which a big part of technological progress comes from the ability of entrepreneurs to organize the successful development and marketing of new products—a dimension in which the United States appears better than most other countries.

In Chapter 11, we looked at the role of human capital as an input in production. People with more education can use more complex machines or handle more complex tasks. Here, we see a second role for human capital: better researchers and scientists and, by implication, a higher rate of technological progress.

It takes many years, and often many decades, for the full potential of major discoveries to be realized. The usual sequence is one in which a major discovery leads to the exploration of potential applications, then to the development of new products, and finally to the adoption of these new products. The Focus Box "The Diffusion of New Technology: Hybrid Corn" shows the results of one of the first studies of this process of the diffusion of ideas. Closer to us is the example of personal computers: 40 years after the commercial introduction of personal computers (Apple I was introduced in 1976), it often seems as if we have just begun discovering their uses.

FOCUS

The Diffusion of New Technology: Hybrid Corn

New technologies are not developed or adopted overnight. One of the first studies of the diffusion of new technologies was carried out in 1957 by Zvi Griliches, a Harvard economist, who looked at the diffusion of hybrid corn in different states in the United States.

Hybrid corn was, in the words of Griliches, "the invention of a method of inventing." Producing hybrid corn entails crossing different strains of corn to develop a type of corn adapted to local conditions. The introduction of hybrid corn can increase the corn yield by up to 20%.

Although the idea of hybridization was developed at the beginning of the 20th century, commercial application did not take place until the 1930s in the United States. Figure 1 shows the rate at which hybrid corn was adopted in selected states from 1932 to 1956.

The figure shows two dynamic processes at work. One is the process through which hybrid corns appropriate to each state were discovered. Hybrid corn became available in southern states (Texas and Alabama) more than 10 years after it had become available in northern states (Iowa, Wisconsin, and Kentucky). The other is the speed at which hybrid corn was adopted in each state. Within 8 years of its introduction, practically all corn in Iowa was hybrid corn. The process was much slower in the South. More than 10 years after its introduction, hybrid corn accounted for only 60% of total acreage in Alabama.

Why was the speed of adoption higher in the North than in the South? Griliches's article showed that the reason was economic: The speed of adoption in each state was a function of the profitability of introducing hybrid corn: profitability was higher in Iowa than in the southern states.[1]

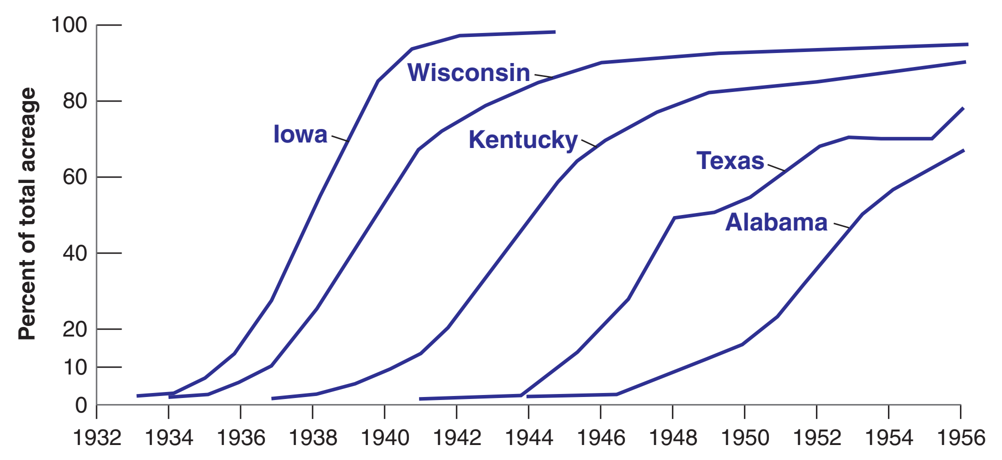


Figure 1

Percentage of Total Corn Acreage Planted with Hybrid Seed, Selected US States, 1932–1956

An age-old worry is that research will become less and less fertile, that most major discoveries have already taken place and that technological progress will begin to slow down. This fear may come from thinking about mining, where higher-grade mines were exploited first, and where we have had to exploit increasingly lower-grade mines. But this is only an analogy, and so far there is no evidence that it is correct.

The Appropriability of Research Results

The second determinant of the level of R&D and of technological progress is the degree of *appropriability* of research results. If firms cannot appropriate the profits from the development of new products, they will not engage in R&D and technological progress will be slow. Many factors are also at work here.

The nature of the research process itself is important. For example, if it is widely believed that the discovery of a new product by one firm will quickly lead to the discovery

[1]Source: *Zvi Griliches, "Hybrid Corn: An Exploration in the Economics of Technological Change,"* Econometrica, *1957, Vol 25, No. 4, pp. 501–522. A number of very nice graphs of other examples technology adoption are available at https://ourworldindata.org/technology-adoption. Used with Permission.*

of an even better product by another firm, there may be little advantage to being first. In other words, a highly fertile field of research may not generate high levels of R&D because no company will find the investment worthwhile. This example is extreme, but revealing.

Even more important is the legal protection given to new products. Without such legal protection, profits from the development of a new product are likely to be small. Except in rare cases where the product is based on a trade secret (such as Coca Cola), it will generally not take long for other firms to produce the same product, eliminating any advantage the innovating firm may have initially had. This is why countries have patent laws. **Patents** give a firm that has discovered a new product—usually a new technique or device—the right to exclude anyone else from the production or use of the new product for some time.

How should governments design patent laws? On the one hand, protection is needed to provide firms with the incentives to spend on R&D. On the other, once firms have discovered new products, it would be best for society if the knowledge embodied in those new products were made available without restriction to other firms and to people. Take, for example, biogenetic research. Only the prospect of large profits can lead bioengineering firms to embark on expensive research projects. Once a firm has found a new product, and the product can save many lives, it would clearly be best to make it available at cost to all potential users. But if such a policy were systematically followed, it would eliminate incentives for firms to do research in the first place. So patent law must strike a difficult balance. Too little protection will lead to little R&D. Too much protection will make it difficult for new R&D to build on the results of past R&D and may also lead to little R&D. (The difficulty of designing good patent or copyright laws is illustrated in the cartoon about cloning.)

This type of dilemma is known as *time inconsistency.* We shall see other examples and discuss it at length in Chapter 22.

These issues go beyond patent laws. To take two controversial examples: What is the role of open-source software? Should students download music, movies, and even textbooks (you see where I am going here) without making payments to the creators?

FOCUS

Management Practices: Another Dimension of Technological Progress

For a given technology and a given human capital of its workers, the way a firm is managed also affects its performance. Some researchers argue that management practices might be stronger than many of the other factors that determine a firm's performance, including technological innovation. In a project that examined the management practices and performance of more than 4,000 medium-sized manufacturing operations in Europe, the United States, and Asia, Nick Bloom of Stanford University and John Van Reenen of the London School of Economics found that firms across the globe that use the same technology but apply good management practices perform significantly better than those that do not. This suggests that improved management practices are one of the most effective ways for a firm to outperform its peers. ("Why Do Management Practices Differ across Firms and Countries?" by Nicholas Bloom and John Van Reenen, *Journal of Economic Perspectives*, 2010, 24(1): pp. 203–224).

A fascinating piece of evidence of the importance of management practices comes from an experimental study conducted by Nick Bloom and colleagues with 20 Indian textile plants. To investigate the role of good management practices, they provided free consulting on management practices to a randomly chosen group of the 20 plants. Then he compared the performance of the firms that received management advice with that of the control plants—those that did not receive advice. He found that adopting good management practices raised productivity by 18% through improved quality and efficiency and reduced inventory. ("Does Management Matter? Evidence from India," by Nicholas Bloom, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, *Quarterly Journal of Economics*, 2012, 128(1): pp. 1–51).

Innovation versus Imitation

Although R&D is clearly central to technological progress, other dimensions are also relevant. Existing technologies can be used more or less efficiently. Strong competition among firms forces them to be more efficient. And, as shown in the Focus Box "Management Practices: Another Dimension of Technological Progress," good management makes a substantial difference to the productivity of firms. Finally, for some countries, R&D may be less important than for others.

In this context, recent research on growth has emphasized the distinction between growth by innovation and growth by imitation. To sustain growth, advanced countries, which are at the **technology frontier**, must innovate. This requires substantial spending on R&D. Poorer countries, which are further from the technology frontier, can instead grow largely by imitating rather than innovating, by importing and adapting existing technologies instead of developing new ones. Importation and adaptation of existing technologies has clearly played a central role in generating high growth in China over the last three decades.

The difference between innovation and imitation also explains why countries that are less technologically advanced often have poorer patent protection (as is true of China). These countries are typically users rather than producers of new technologies. Much of their improvement in productivity comes not from inventions within the country but from the adaptation of foreign technologies. So the costs of weak patent protection are small because there would be few domestic inventions anyway. And the benefits of low patent protection are clear: They allow domestic firms to use and adapt foreign technology without having to pay high royalties to the foreign firms that developed the technology.

At this stage, you might have the following question: If in poor countries technological progress is more a process of imitation rather than a process of innovation, why are some countries, such as China and other Asian countries, good at doing this, whereas others, for example, many African countries, are not? This question takes us from macroeconomics to development economics, and it would take a text in development economics to do it justice. But it is too important a question to leave aside entirely; we discuss this issue in the next section.

12-3 INSTITUTIONS, TECHNOLOGICAL PROGRESS, AND GROWTH

To get a sense of why some countries are good at imitating existing technologies, whereas others are not, compare Kenya and the United States. PPP GDP per person in Kenya is about 1/18th of PPP GDP per person in the United States. The difference is partly due to a much lower level of capital per worker in Kenya. It is also due to a much lower technological level in Kenya, where the state of technology is estimated to be about 1/13th of the US level. Why is the state of technology in Kenya so low? The country potentially has access to most of the technological knowledge in the world. What prevents it from simply adopting much of the advanced countries' technology and quickly closing much of its technological gap with the United States?

One can think of a number of potential answers, ranging from Kenya's geography and climate to its culture. Most economists believe, however, that the main source of the problem, for poor countries in general and for Kenya in particular, lies in their poor institutions.

What institutions do economists have in mind? At a broad level, the protection of **property rights** may well be the most important. Few individuals are going to create firms, introduce new technologies, and invest in R&D if they expect that profits will be appropriated by the state, extracted in bribes by corrupt bureaucrats, or stolen by other people in the economy. Figure 12-5 plots PPP GDP per person in 1995 (using a logarithmic scale) for 90 countries against an index measuring the degree of protection from expropriation; the index was constructed for each of these countries by an international business organization. The positive correlation between the two is striking (the figure also plots the regression line). Low protection is associated with a low GDP per person (at the extreme left of the figure are the Democratic Republic of the Congo and Haiti); high protection is associated with a high GDP per person (at the extreme right are the United States, Luxembourg, Norway, Switzerland, and the Netherlands).

With an index of 6, Kenya is below the regression line, which means it has lower GDP per person than would be predicted based just on the index.

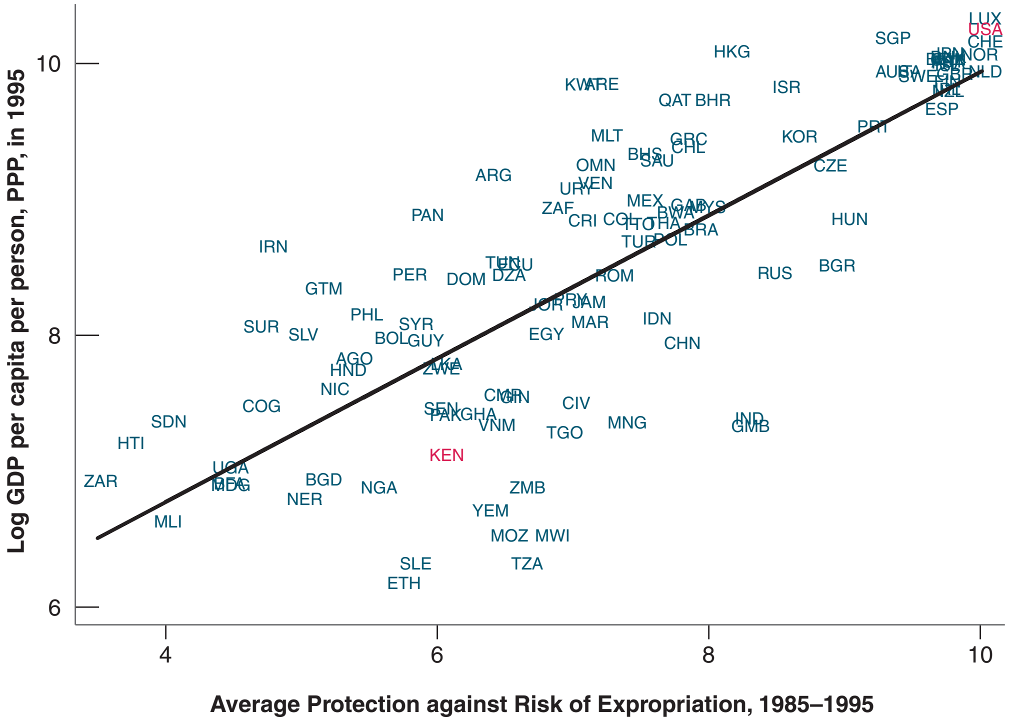


Figure 12-5

Protection from Expropriation and GDP per Person

There is a strong positive relation between the degree of protection from expropriation and the level of GDP per person.

Source: Daron Acemoglu, "Understanding Institutions," Lionel Robbins Lectures, 2004, London School of Economics. http://economics.mit.edu/files/1353.

FOCUS

The Importance of Institutions: North Korea and South Korea

Following the surrender of Japan in 1945, Korea formally acquired its independence but became divided at the 38th parallel into two zones of occupation, with Soviet armed forces occupying the North and US armed forces occupying the South. Attempts by both sides to claim jurisdiction over all of Korea triggered the Korean War (1950–1953). At the armistice in 1953, Korea became formally divided into two countries, the Democratic People's Republic of North Korea and, in the South, the Republic of Korea.

An interesting feature of Korea before separation was its ethnic and linguistic homogeneity. The North and the South were inhabited by essentially the same people, with the same culture and the same religion. Economically, the two regions were also highly similar at the time of separation. PPP GDP per person, in 1996 dollars, was roughly the same, about $700 in both the North and South.

But 50 years later, as shown in Figure 1, PPP GDP per person was 10 times higher in South Korea than in North Korea—$12,000 versus $1,100! South Korea had joined the OECD, the club of rich countries, whereas North Korea had seen its GDP per person decrease by nearly two-thirds from its peak of $3,000 in the mid-1970s and was facing famine on a large scale. (The figure, taken from the work of Daron Acemoglu, shows the numbers up to 1998. In 2017, the difference had further increased, with PPP GDP per person around $30,000 in South Korea versus $1,800 for North Korea.)

A striking visualization of the difference is shown in Figure 2, which shows a satellite picture of Korea at night. The dark space between South Korea (lower right quadrant) and China (upper left quadrant) is North Korea.

What happened? Institutions and the organization of the economy were dramatically different during that period in the South and in the North. South Korea relied on a capitalist organization of the economy, with strong state intervention but also private ownership and legal protection of private producers. North Korea relied on central planning: Industries were quickly nationalized. Small firms and farms were forced to join large cooperatives so they could be supervised by the state. There were no private property rights for individuals. The result was the decline of the industrial sector and the collapse of agriculture. The lesson is sad, but clear; institutions matter very much for growth.[2]

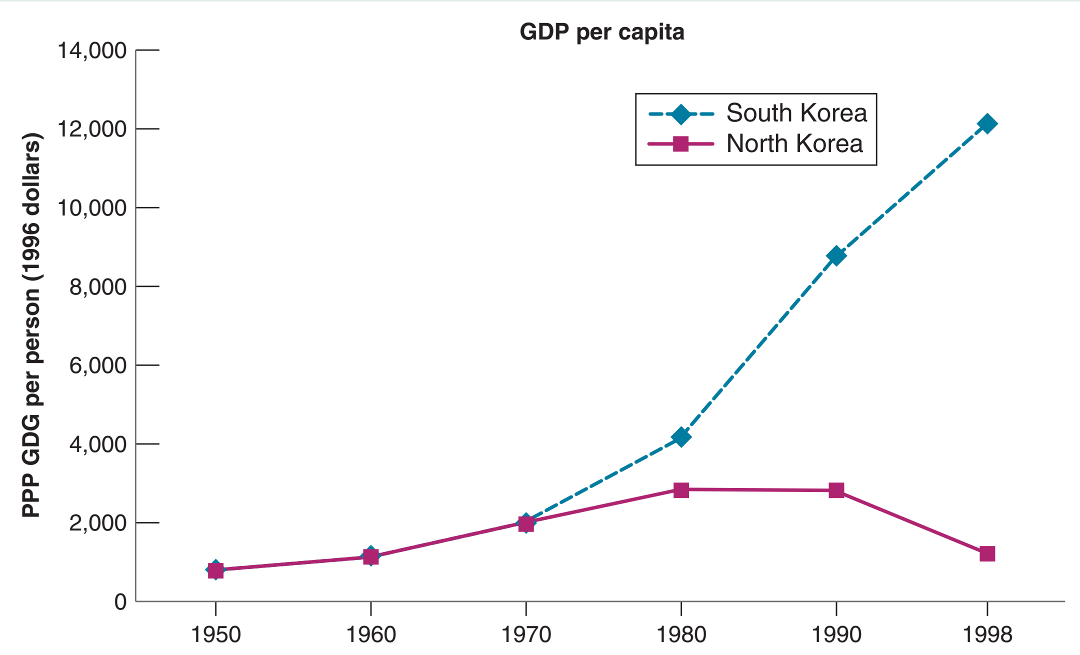


Figure 1

PPP GDP per Person: North and South Korea, 1950–1998

[2]Source: *Daron Acemoglu, "Understanding Institutions," Lionel Robbins Lectures, 2004, London School of Economics. http://economics.mit.edu/files/1353.*

NASA Earth Observatory

Figure 2

Korea at night

What does protection of property rights mean in practice? It means a good political system, in which those in charge cannot expropriate or seize the property of the citizens. It means a good judicial system, where disagreements can be resolved efficiently, rapidly, and fairly. Looking at a finer degree of detail, it means laws against insider trading in the stock market, so people are willing to buy stocks and thus provide financing to firms; clearly written and well-enforced patent laws, so firms have an incentive to do research and develop new products; good antitrust laws, so competitive markets do not turn into monopolies with few incentives to introduce new methods of production and new products—and the list obviously goes on. (A particularly dramatic example of the role of institutions is given in the Focus Box "The Importance of Institutions: North Korea and South Korea.")

This still leaves one essential question: Why don't poor countries adopt these good institutions? The answer is that it is hard! Good institutions are complex and difficult for poor countries to put in place. Certainly, causality runs both ways in Figure 12-5: Low protection against expropriation leads to low GDP per person. But it is also the case that low GDP per person leads to worse protection against expropriation. Poor countries are often too poor to afford a good judicial system and to maintain a good police force, for example. Thus, improving institutions and starting a virtuous cycle of higher GDP per person and better institutions is difficult. The fast-growing countries of Asia have succeeded. (The Focus Box "What Is behind Chinese Growth?" explores the case of China in more detail.) Some African countries appear also to be succeeding; others are still struggling.

A quote from Gordon Brown, a former UK prime minister: "In establishing the rule of law, the first five centuries are always the hardest!"

FOCUS

What Lies Behind Chinese Growth?

From 1949—the year the People's Republic of China was established—to the late 1970s, China's economic system was based on central planning. Two major politico-economic reforms, the Great Leap Forward in 1958 and the Cultural Revolution in 1966, ended up as human and economic catastrophes. Output decreased by 20% from 1959 to 1962, and it is estimated that 25 million people died of famine during the same period. Output again decreased by more than 10% from 1966 to 1968.

After Chairman Mao's death in 1976, the new leaders decided to progressively introduce market mechanisms in the economy. In 1978, an agricultural reform was put in place, allowing farmers, after satisfying a quota due to the state, to sell their production in rural markets. Over time, farmers obtained increasing rights to the land, and today state farms produce less than 1% of agricultural output. Outside of agriculture, and also starting in the late 1970s, state firms were given increasing autonomy over their production decisions, and market mechanisms and prices were introduced for an increasing number of goods. Private entrepreneurship was encouraged, often taking the form of "Township and Village Enterprises," collective ventures guided by a profit motive. Tax advantages and special agreements were used to attract foreign investors.

The economic effects of these cumulative reforms were dramatic. Average growth of output per worker increased from 2.5% between 1952 and 1977 to more than 9% since then. Even today, growth remains above 6% (see Chapter 1). This was the result of both very high rates of capital accumulation and of technological progress (see the further discussion in the appendix to this chapter).

Is such high growth surprising? One could argue that it is not. Looking at the more than 10-fold difference in productivity between North Korea and South Korea noted in the previous Focus Box, it is clear that central planning is a poor economic system. Thus, it would seem that, by moving from central planning to a market economy, countries could easily experience large increases in productivity. The answer is not so obvious, however, when one looks at the experience of the many countries that, since the late 1980s, have moved away from central planning. In most Central European countries, this transition was typically associated with an initial 10% to 20% drop in GDP, and it took five years or more for output to exceed its pre-transition level. In Russia and in the new countries carved out of the Soviet Union, the drop was even larger and longer lasting. (Many transition countries now have strong growth, although their growth rates are far below that of China.)

In Central and Eastern Europe, the initial effect of transition was a collapse of the state sector, only partially compensated by slow growth of the new private sector. In China, the state sector has declined more slowly, and its decline has been more than compensated by strong private sector growth. This gives a proximate explanation for the difference between China and the other transition countries. But it still begs the question: How was China able to achieve this smoother transition?

Some observers offer a cultural explanation. They point to the Confucian tradition, based on the teachings of Confucius, which still dominates Chinese values and emphasizes hard work, respect of one's commitments, and trustworthiness among friends. All these traits, they argue, are the foundations of institutions that allow a market economy to perform well.

Some observers offer a historical explanation. They point to the fact that, in contrast to Russia, central planning in China lasted only a few decades. Thus, when the shift back to a market economy took place, people still knew how such an economy functioned and adapted easily to the new economic environment.

Most observers point to the strong rule of the communist party in the process. They point out that, in contrast to Central and Eastern Europe, the political system did not change, and the government was able to control the pace of transition. It was able to experiment along the way, to allow state firms to continue production while the private sector grew, and to guarantee property rights to foreign investors (in Figure 12-5, China has an index of property rights of 7.7, not far from its typical value in rich countries). With foreign investors has come the technology from rich countries, and, in time, the transfer of this knowledge to domestic firms. For political reasons, such a strategy was simply not open to governments in Central and Eastern Europe.

The limits of the Chinese strategy are clear. Leaving aside the central issue of limits to democracy and focusing on the economic issues: The banking system is still inefficient. Recently, under President Xi JinPing, the shift from the state sector to the private sector has slowed down. So far, however, these problems have not stood in the way of high growth, which remains around 6%.

For more on China's economy, read Gregory Chow, *China's Economic Transformation*, 3rd ed. (2014, Wiley-Blackwell).

For a comparison between transition in Eastern Europe and China, read Jan Svejnar, "China in Light of the Performance of Central and East European Economies," IZA Discussion Paper 2791, May 2007.

SUMMARY

- When we think about the implications of technological progress for growth, it is useful to think of technological progress as increasing the amount of effective labor available in the economy (that is, labor multiplied by the state of technology). We can then think of output as being produced with capital and effective labor.
- In steady state, output *per effective worker* and capital *per effective worker* are constant. Put another way, output *per worker* and capital *per worker* grow at the rate of technological progress. Put yet another way, output and capital grow at the same rate as effective labor, thus at a rate equal to the growth rate of the number of workers plus the rate of technological progress.
- When the economy is in steady state, it is said to be on a balanced growth path. Output, capital, and effective labor are all growing "in balance," that is, at the same rate.
- The rate of output growth in steady state is independent of the saving rate. However, the saving rate affects the steady-state level of output per effective worker. And increases in the saving rate will lead, for some time, to an increase in the growth rate above the steady-state growth rate.
- Technological progress depends on both (1) the fertility of research and development, how spending on R&D translates into new ideas and new products, and (2) the appropriability of the results of R&D, which is the extent to which firms benefit from the results of their R&D.
- When designing patent laws, governments must balance their desire to protect future discoveries and provide incentives for firms to do R&D with their desire to make existing discoveries available without restriction to potential users.
- Sustained technological progress requires that the right institutions are in place. In particular, it requires well-established and well-protected property rights. Without good property rights, a country is likely to remain poor. But in turn, a poor country may find it difficult to put in place good property rights.

KEY TERMS

state of technology, 244
effective labor, 244
labor in efficiency units, 245
balanced growth, 248
research and development (R&D), 250
fertility of research, 251
appropriability, 251
patents, 253
technology frontier, 254
property rights, 255

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. Writing the production function in terms of capital and effective labor implies that as the level of technology increases by 10%, the number of workers required to achieve the same level of output decreases by 10%.
b. If the rate of technological progress increases, the investment rate (the ratio of investment to output) must increase to keep capital per effective worker constant.
c. In steady state, output per effective worker grows at the rate of population growth.
d. In steady state, output per worker grows at the rate of technological progress.
e. A higher saving rate implies a higher level of capital per effective worker in the steady state and thus a higher rate of growth of output per effective worker.
f. Even if the potential returns from research and development (R&D) spending are identical to the potential returns from investing in a new machine, R&D spending is much riskier for firms than investing in new machines.
g. The fact that one cannot patent a theorem implies that private firms will not engage in basic research.
h. Because eventually we will know everything, growth will have to come to an end.
i. Technology has not played an important part in Chinese economic growth.

2. R&D and growth

a. Why is the amount of R&D spending important for growth? How do the appropriability and fertility of research affect the amount of R&D spending?

How does each of the policy proposals listed in parts b through e affect the appropriability and fertility of research, R&D spending in the long run, and output in the long run?

b. An international treaty ensuring that each country's patents are legally protected all over the world. This may be a part of the proposed but now cancelled Trans-Pacific Partnership.
c. Tax credits for each dollar of R&D spending.
d. A decrease in funding of government-sponsored conferences between universities and corporations.
e. The elimination of patents on breakthrough drugs, so the drugs can be sold at low cost as soon as they become available.

3. Sources of technological progress: Leaders versus followers

a. Where does technological progress come from for the economic leaders of the world?
b. Do developing countries have alternatives to the sources of technological progress you mentioned in part a?
c. Do you see any reasons developing countries may choose to have poor patent protection? Are there any dangers in such a policy (for developing countries)?

DIG DEEPER

4. For each of the economic changes listed in parts a and b, assess the likely impact on the growth rate and the level of output over the next five years and over the next five decades.

a. A permanent reduction in the rate of technological progress.
b. A permanent reduction in the saving rate.

5. Measurement error, inflation, and productivity growth

Suppose that there are only two goods produced in an economy: haircuts and banking services. Prices, quantities, and the number of workers occupied in the production of each good for year 1 and for year 2 are given in the table:

Year	Year 1			Year 2		
Variables	P_1	Q_1	N_1	P_2	Q_2	N_2
Haircuts	10	100	50	12	100	50
Banking	10	200	50	12	230	60

a. What is nominal GDP in each year?
b. Using year 1 prices, what is real GDP in year 2? What is the growth rate of real GDP?
c. What is the rate of inflation using the GDP deflator?
d. Using year 1 prices, what is real GDP per worker in year 1 and year 2? What is labor productivity growth between year 1 and year 2 for the whole economy?

Now suppose that banking services in year 2 are not the same as banking services in year 1. Year 2 banking services include telebanking, which year 1 banking services did not include. The technology for telebanking was available in year 1, but the price of banking services with telebanking in year 1 was \$13, and no one chose to purchase this package. However, in year 2, the price of banking services with telebanking was \$12, and everyone chose to have this package (i.e., in year 2 no one chose to have the year 1 banking services package without telebanking). (Hint: Assume that there are now two types of banking services: those with telebanking and those without. Rewrite the preceding table but now with three goods: haircuts and the two types of banking services.)

e. Using year 1 prices, what is real GDP for year 2? What is the growth rate of real GDP?
f. What is the rate of inflation using the GDP deflator?
g. What is labor productivity growth between year 1 and year 2 for the whole economy?
h. Consider this statement: "If banking services are mismeasured—for example, by not taking into account the introduction of telebanking—we will overestimate inflation and underestimate productivity growth." Discuss this statement in light of your answers to parts a through g.

6. Suppose that the economy's production function is

$$Y = \sqrt{K}\sqrt{AN}$$

that the saving rate, s, is equal to 16%, and that the rate of depreciation, δ, is equal to 10%. Suppose further that the number of workers grows at 2% per year and that the rate of technological progress is 4% per year.

a. Find the steady-state values of the variables listed in (i) through (v).
   i. The capital stock per effective worker
   ii. Output per effective worker
   iii. The growth rate of output per effective worker
   iv. The growth rate of output per worker
   v. The growth rate of output
b. Suppose that the rate of technological progress doubles to 8% per year. Recompute the answers to part a. Explain.
c. Now suppose that the rate of technological progress is still equal to 4% per year, but the number of workers now grows at 6% per year. Recompute the answers to part a. Are people better off in part a or part c? Explain.

7. Discuss the potential role of each of the factors listed in parts a through g on the steady-state level of output per worker. In each case, indicate whether the effect is through A, through K, through H, or through some combination of A, K, and H. A is the level of technology, K is the level of capital stock, and H is the level of the human capital stock.

a. Geographic location
b. Education
c. Protection of property rights
d. Openness to trade
e. Low tax rates
f. Good public infrastructure
g. Low population growth

EXPLORE FURTHER

8. Growth accounting

The appendix to this chapter shows how data on output, capital, and labor can be used to construct estimates of the rate of growth of technological progress. We modify that approach in this problem to examine the growth of capital per worker.

$$Y = K^{1/3}(AN)^{2/3}$$

The function gives a good description of production in rich countries. Following the same steps as in the appendix, you can show that

$$(2/3)g_A = g_Y - (2/3)g_N - (1/3)g_K$$
$$= (g_Y - g_N) - (1/3)(g_K - g_N)$$

where g_Y denotes the growth rate of Y, g_K is the growth rate of capital, and g_N is the growth rate of the labor input.

a. What does the quantity $g_Y - g_N$ represent? What does the quantity $g_K - g_N$ represent?
b. Your download of the Penn World Tables in Chapter 10 includes the information needed to calculate the components of the growth accounting formula in the appendix. That formula is

$$\text{residual} \equiv g_Y - [\alpha g_N + (1 - \alpha)g_K]$$

From the Penn World Tables, you can fill in the chart below:

Use rgdpo for output (2011 US dollars); emp for number of employees labor input; ck for the capital stock (2011 US dollars)

The share of labor is labsh. These values are copied directly from the Penn World Tables version 9.

	Year 2000				Year 2014			
	Y	*N*	*K*	Share of labor	*Y*	*N*	*K*	Share of labor
China								
United States								

Calculate annual growth rates of these variables over the 14 years of available data and answer these questions:

Is the ratio of K/N rising or falling in China? In the United States? Which country has the highest labor share? Which country has the largest residual? (Use the average labor share over the two years to calculate the residual.) What does that residual represent? Are you surprised by any of your results?

FURTHER READINGS

- For more on growth, both theory and evidence, read Charles Jones, *Introduction to Economic Growth*, 3rd ed. (2013). Jones's Web page, http://web.stanford.edu/~chadj/, is a useful portal to the research on growth.
- For more on patents, see *The Economist*, Special Report: Patents and Technology, October 20th, 2005.
- For more on growth in two large, fast-growing countries, read Barry Bosworth and Susan M. Collins, "Accounting for Growth: Comparing China and India," *Journal of Economic Perspectives*, 2008, Vol. 22, No. 1: 45–66.
- For the role of institutions in growth, read sections 1 to 4 in "Growth Theory Through the Lens of Development Economics," by Abhijit Banerjee and Esther Duflo, Chapter 7, *Handbook of Economic Growth* (2005).
- For more on institutions and growth, you can read the slides from the 2004 Lionel Robbins lectures "Understanding Institutions" given by Daron Acemoglu. These are found at http://economics.mit.edu/files/1353.
- For the role of ideas in sustaining technological progress and growth, listen to Paul Romer's 2018 Nobel Prize lecture, www.youtube.com/watch?v=vZmgZGIZtiM.

APPENDIX: How to Measure Technological Progress, and the Application to China

In Section 12-1, we discussed the joint role of capital accumulation and technological progress in determining growth. We saw that, in steady state, growth of output per person was determined fully by the rate of technological progress. Out of steady state, however, a growth rate of the capital stock higher than the growth rate of output—due to, say, an increase in the saving rate—could lead to higher growth for some time.

This raises the question: How do we measure the rate of technological progress? The answer was given in 1957 by Robert Solow and is still in use today. It relies on one important assumption: that each factor of production is paid its marginal product.

Under this assumption, it is easy to compute the contribution of an increase in any factor of production to the increase in output. For example, if a worker is paid \$30,000 a year, the assumption implies that her contribution to output is equal to \$30,000. Now suppose that this worker increases the amount of hours she works by 10%. The increase in output from the increase in her hours will be equal to \$30,000 × 10%, or \$3,000.

Let us write this more formally. Denote output by Y, labor by N, and the real wage by W/P. The notation Δx means the change in variable x. Then, as we just established, the change in output is equal to the real wage multiplied by the change in labor.

$$\Delta Y = \frac{W}{P}\Delta N$$

Divide both sides of the equation by Y, divide and multiply the right side by N, and reorganize:

$$\frac{\Delta Y}{Y} = \frac{WN}{PY}\frac{\Delta N}{N}$$

Note that the first term on the right (WN/PY) is equal to the share of labor in output—the total wage bill in dollars divided by the value of output in dollars. Denote this share by α. $\Delta Y/Y$ is the rate of growth of output; denote it by g_Y. $\Delta N/N$ is the rate of change of the labor input; denote it by g_N. Then the previous relation can be written as

$$g_Y = \alpha\, g_N$$

More generally, this reasoning implies that the part of output growth attributable to growth of the labor input is equal to α times g_N. If, for example, employment grows by 1.7% (you will see below why I am using this and the other numbers) and the share of labor is 0.6, then the output growth due to the growth in employment is equal to 1.0% (0.6 times 1.7%).

Similarly, we can compute the part of output growth attributable to growth of the capital stock. Because there are only two factors of production, labor and capital, and because the share of labor is equal to α, the share of capital in income must be equal to $(1 - \alpha)$. If the growth rate of capital is equal to g_K, then the part of output growth attributable to growth of capital is equal to $(1 - \alpha)$ times g_K. If, for example, capital grows by 9.2%, and the share of capital is 0.4, then the output growth due to the growth of the capital stock is equal to 3.7% (0.4 times 9.2%).

Putting the contributions of labor and capital together, the growth in output attributable to growth in both labor and capital is equal to $(\alpha g_N + (1 - \alpha)g_K)$, so in our numerical example, 1.0% + 3.7% = 4.7%.

We can then measure the effects of technological progress by computing what Solow called the *residual*, the excess of actual growth of output g_Y over the growth attributable to growth of labor and the growth of capital $(\alpha g_N + (1 - \alpha)g_K)$.

$$\text{residual} \equiv g_Y - [\alpha g_N + (1 - \alpha)g_K]$$

This measure is called the **Solow residual**. It is easy to compute. All we need to know are the growth rate of output, g_Y, the growth rate of labor, g_N, and the growth rate of capital, g_K, together with the shares of labor, α, and capital, $(1 - \alpha)$. To continue with our previous numerical example: If output growth is equal to 7.2%, then the Solow residual is equal to 7.2% − 4.7% = 2.5%.

The Solow residual is sometimes called the **rate of growth of total factor productivity** (or the **rate of TFP growth**). The use of total factor productivity is to distinguish it from the *rate of growth of labor productivity*, which is defined as $(g_Y - g_N)$, the rate of output growth minus the rate of labor growth.

The Solow residual is related to the rate of technological progress in a simple way. The residual is equal to the share of labor times the rate of technological progress:

$$\text{residual} = \alpha g_A$$

We shall not derive this result here. But the intuition for this relation comes from the fact that what matters in the production function $Y = F(K, AN)$ (equation (12.1)) is the product of the state of technology and labor, AN. We saw that to get the contribution of labor growth to output growth, we must multiply the growth rate of labor by its share. Because N and A enter the production function in the same way, it is clear that to get the contribution of technological progress to output growth, we must also multiply it by the share of labor.

If the Solow residual is equal to zero, so is technological progress. To construct an estimate of g_A, we must construct the Solow residual and then divide it by the share of labor. This is how the estimates of g_A presented in the text are constructed. Going back to our numerical example, if the share of labor is 0.6, the rate of technological progress is equal to 2.5%/0.6 = 4.2%.

You may wonder why I used the numbers above. The reason is simple: These are the numbers for China for the period 1978 to 2017, the period that was described in the Focus Box as the period of high growth. What do the numbers say?

Output growth was very high, 7.2%. Growth of output per worker was also very high, at 5.5%. The rate of technological progress was also high, but not quite as much, namely 4.2%. The difference between the growth rate of output and the rate of technological progress reflected a steady increase in the ratio of capital to output, equal to 2% a year. In terms of Figure 12-2, the Chinese economy moved from $(K/AN)_0$ to $(K/AN)^*$.

Eventually, the ratio of capital to output will stabilize, and the growth of output per worker will reflect the rate of technological progress. Thus, if this rate remains the same, China's growth of output per worker will decrease to 4.2%. If growth of labor remains the same, namely 1.7%, then growth of output will converge to 5.9%. This does indeed roughly correspond to forecasts of future growth in China.

The original presentation of the ideas discussed in this appendix is found in Robert Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 1957, 39(3): pp. 312–320.

Key Terms

Solow residual, 262
rate of growth of total factor productivity, 262
rate of TFP growth, 262

This page is intentionally left blank

13

The Challenges of Growth

Much of the focus of Chapters 10 to 12 was on steady-state growth—the steady increase in the standard of living over time. This may have led you to form too optimistic a picture. Growth is in fact a complex process, facing many challenges. You read about them in newspapers every day: Can we be confident that technological progress will continue at the same rate? Should we be worried that robots will destroy jobs and lead to mass unemployment? Does everybody really benefit from the increased standard of living? And, in the process of growing, aren't we destroying the planet and mortgaging our future? These are the issues I take up in this chapter.

Section 13-1 discusses the future of technological progress. Some economists believe that the era of major inventions is largely over and we should expect technological progress to slow down considerably. Others believe instead that we are on the verge of another technological revolution and the future is bright. What are the arguments on both sides, and what should we expect?

Section 13-2 discusses whether technological progress will lead to mass unemployment. Such fears go back a long way but have so far proven unfounded. Will the future be different? If so, what policies should we adopt?

Section 13-3 discusses the relation between growth and inequality. Inequality has increased, and technological progress and globalization are blamed. What is the evidence? And, if they are responsible, what policies should be adopted to limit the increase in inequality?

Section 13-4 discusses climate change. Growth is blamed for various ecological disasters. It is believed to be one of the main factors behind climate change. What is the evidence? If it is responsible, what policies should be adopted? What is the chance that the policies will indeed be adopted?

I have started every chapter so far with "If you remember one basic message from this chapter, it should be...". I did not succeed in summarizing the conclusions of this chapter in one sentence. But the general theme is that growth is a complex process, and policies to make it fair and sustainable are of the essence. ▶▶▶

13-1 THE FUTURE OF TECHNOLOGICAL PROGRESS

There is a strange tension between, on the one hand, the perception of fast technological progress and, on the other, the productivity growth numbers. Looking around us, we have the impression of fast technological change, be it the introduction of robots, the development of artificial intelligence, the use of machine learning, or, more concretely, the number of applications that we can use on our smart phones. One might have expected this to be reflected in higher measured productivity growth. Yet, as we saw in Table 1-2 in Chapter 1, US measured productivity growth has slowed down since the mid-2000s, running at less than half the rate of earlier decades.

How can the two be reconciled? One possible answer is measurement error. As we discussed in the Focus Box "Real GDP, Technological Progress, and the Price of Computers" in Chapter 2, measuring technological progress is difficult. For example, how much better is a 2018 computer relative to a 2010 computer? Although statistical agencies try to adjust for changes in speed, hard disk size, and so on, they do not do a perfect job. And the challenge is even bigger with respect to new software. How much should we value the fact that we can now access instantaneously nearly any piece of music we want to listen to?

How much better is your smartphone than your previous smartphone? Better for sure, but how much better: 20%, 50%, 100%?

Could it be therefore that we understate the rate of productivity growth, and that the true rate is higher? Research on the topic suggests that the answer is yes, but it comes with a twist: It does not appear that, for the economy as a whole, the mismeasurement has substantially increased. Put another way, mismeasurement can explain only a small part of the decrease in measured productivity growth.

If we take as given that there is truly a decline in productivity growth, how can we explain it? This is the subject of intense debate among researchers.

We saw in Chapter 12 that slower productivity growth could come from a decrease in the ratio of capital to output or from a slowdown of technological progress. The ratio of capital to output has remained roughly constant, however, so the origin of slower productivity growth is slower technological progress.

Some argue that the current major innovations are less important than the major innovations of the past. Major innovations are innovations that have applications in many fields and many products. (For this reason, these innovations are called **general-purpose technologies.**) Robert Gordon, of Northwestern University, has studied the history of technological innovations and argues that the two major innovations of the last 150 years were electricity and the internal combustion engine. For many decades after their introduction, they transformed methods of production and life in general in often dramatic ways. To list a few examples: Electricity led to the use of air conditioning and the development of the South of the United States. It led to the development of refrigerators and thus the ability to keep food for much longer periods of time, changing the system of food production and distribution. The internal combustion engine led to automobiles, triggering the building of interstate highways and shaping the development of cities. The current major innovation, he argues, namely digitization, does not come close to having the same wide implications. Thus, we should not expect it to have the same impact as those previous major innovations. The period of high technological progress, he claims, is behind us.

There is no agreement about whether the innovation should be called digitization or digitalization.

Others, in particular Eric Brynjolfsson of MIT, argue that digitization, together with the increasing power of computers, will transform our lives as much as or more than electricity or the internal combustion engine did earlier. Artificial intelligence and machine learning may transform nearly any activity, from driverless cars to the mapping of the human genome and major progress in medicine, to the way lawyers access relevant jurisprudence. He argues that the current low productivity growth reflects the slow process of diffusion and the discovery of new applications. He concludes that the future is bright, and he sees no end to technological progress.

On the speed of diffusion: See the Focus Box in Chapter 12, "The Diffusion of New Technology: Hybrid Corn."

Who is right, who is wrong? The honest answer is that nobody really knows. There are a number of spectacularly wrong predictions, including that of Thomas Watson, former CEO of IBM, who in 1943 said, "I think there is a world market for maybe five computers." But the debate shows the importance of both the process of innovation and of the speed of diffusion in explaining technological progress at the frontier. (As I discussed in Chapter 12, technological progress in countries behind the frontier depends largely on other factors, the ability to adapt and use existing technologies, which in turn depend on institutions from property rights to education.)

For more wildly incorrect forecasts of technological progress, see www.pcworld.com/article/155984/worst_tech_predictions.html

13-2 ROBOTS AND UNEMPLOYMENT

Since the beginning of the Industrial Revolution, workers have worried that technological progress would eliminate jobs and increase unemployment. In early 19th-century England, groups of workers in the textile industry, known as the *Luddites*, destroyed the new machines that they saw as a direct threat to their jobs. Similar movements took place in other countries. The term *saboteur* comes from one of the ways French workers destroyed machines: by putting their sabots (heavy wooden shoes) into the machines.

The theme of **technological unemployment** typically resurfaces whenever unemployment is high. During the Great Depression, a movement called the *technocracy movement* argued that high unemployment came from the introduction of machinery, and that things would only get worse if technological progress were allowed to continue. In the late 1990s, France passed a law reducing the normal workweek from 39 to 35 hours. One of the reasons invoked was that, because of technological progress, there was no longer enough work for all workers to have full-time jobs; the proposed solution was to have each worker work fewer hours (at the same hourly wage) so that more of them could be employed. (This was not the only rationale behind the law. Another and better reason was that increases in productivity should be taken partly in the form of higher income and partly in the form of more leisure. The problem with that argument is why the choice between income and leisure should have been legislated rather than left to the workers and firms to decide.)

In its crudest form, the argument that technological progress must necessarily lead to higher unemployment is obviously false. The large improvements in the standard of living that advanced countries enjoy have come with large *increases* in employment and no increase in the unemployment rate. In the United States, real output per person has increased by a factor of 10 since 1890 and, far from declining, employment has increased by a factor of 6.5 (reflecting a parallel increase in the size of the US population). As we have seen, the unemployment rate is very low. Nor, looking across countries, is there any evidence of a systematic positive relation between the unemployment rate and the level of productivity.

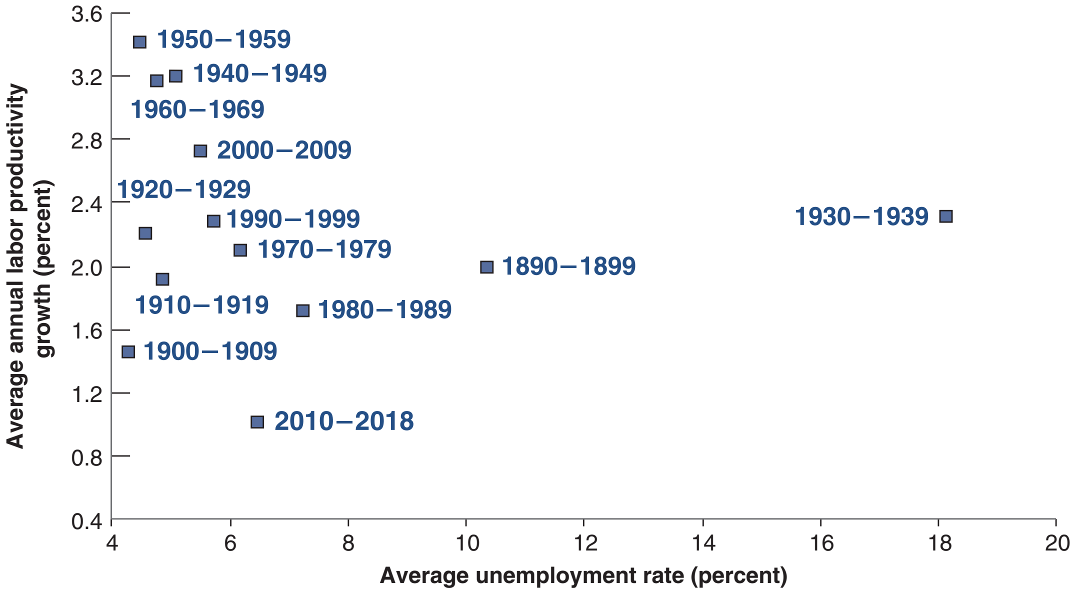
More sophisticated versions of the technological unemployment hypothesis, however, must be taken more seriously. First, one might expect that, as higher productivity leads some firms to reduce employment, it may take time for new jobs to replace the jobs that are destroyed, leading to higher unemployment for some time if not forever. The evidence, however, suggests that this is not the case.

Figure 13-1 plots average US labor productivity growth and the average unemployment rate during each decade since 1890. At first glance, there seems to be little relation between the two. But if we ignore the 1930s (the decade of the Great Depression), then a relation emerges between productivity growth and the unemployment rate. But it is the *opposite* of the relation predicted by those who believe in technological unemployment. Periods of *high productivity growth*, like the 1940s to the 1960s, have been associated

Figure 13-1

Productivity Growth and Unemployment, Averages by Decade, 1890–2018

Source: Data prior to 1960: Historical Statistics of the United States. Data after 1960: Bureau of Labor Statistics.



with *a lower unemployment rate*. Periods of *low productivity growth*, such as the United States saw during 2010–2018, have been associated with *a higher unemployment rate*.

What was true in the past may not, however, be true in the future. And robots seem threatening enough (the picture makes the case). By their very nature, they can replace the least skilled workers; then, as they become more sophisticated, the workers with higher skills, and so on. A recent study concludes that, over the next two decades, 47% of US workers are at risk of being replaced by robots.

A study by Daron Acemoglu and Pascual Restrepo of MIT, looking at the evidence on the effect of the introduction of robots on employment in local labor markets, concludes that, where they are introduced, robots do indeed destroy jobs. Even if the reduction in costs allows firms that use robots to sell their goods more cheaply and thus increase sales and output, the substitution of robots for workers dominates the outcome and leads to fewer jobs.

The question is whether robots lead to job creation elsewhere. For example, if firms that use robots produce cheaper intermediate inputs, other firms now will see their costs of production go down and are thus likely to increase employment. This is much harder to assess. What can be said is that, so far, the US unemployment rate is low by historical standards even for workers with low skills: The unemployment rates for workers with a high school degree but no college and for workers who have not finished high school are 3.8% and 5.7%, respectively, both at historical lows. Clearly, whatever jobs were eliminated because of robots have been replaced by jobs elsewhere. This might, however, be due to other factors, other favorable shocks to employment, and

one cannot be sure that it will hold in the future. And it also leaves open the issue of whether workers who lost their jobs to robots got new ones at the same wage or at a lower wage. This gets us to the issue of income inequality that we focus on in the next section.

In short, so far, robots have not led to anything like mass unemployment or, for that matter, to higher unemployment. Although this gets close to science fiction, it is interesting to think of what the economy would look like if robots were to eliminate all jobs. Put bluntly, it could be hell or heaven. Hell: All robots belong to, say, one extremely wealthy owner and everybody else is unemployed. Heaven: Everybody owns the rights to one robot apiece and lives on the income from the robot production while enjoying a life of leisure. Admittedly, these are beyond extreme scenarios, but they indicate the issues that policymakers may face in the future: how to fight inequality—whether, for example, to guarantee a minimum income for all, or how much to redistribute wealth and property rights.

13-3 GROWTH, CHURN, AND INEQUALITY

Robots are a particularly striking form of technological progress. But the issues they raise are more general. Technological progress is fundamentally a process of structural change, with profound implications for what happens in labor markets.

The theme was central to the work of Joseph Schumpeter, a Harvard economist who, in the 1930s, emphasized that the process of growth was fundamentally a process of **creative destruction**: New goods are developed, making old ones obsolete. New techniques of production are introduced, requiring new skills and making some old skills less useful. As this happens, old jobs are eliminated, new jobs are created.

The essence of this process is nicely captured in the following quote from Robert McTeer, a past president of the Federal Reserve Bank of Dallas, in his introduction to a report titled *The Churn*:

The Churn: The Paradox of Progress (1993)

> "My grandfather was a blacksmith, as was his father. My dad, however, was part of the evolutionary process of the churn. After quitting school in the seventh grade to work for the sawmill, he got the entrepreneurial itch. He rented a shed and opened a filling station to service the cars that had put his dad out of business. My dad was successful, so he bought some land on the top of a hill and built a truck stop. Our truck stop was extremely successful until a new interstate went through 20 miles to the west. The churn replaced US 411 with Interstate 75, and my visions of the good life faded."

Many professions, from those of blacksmiths to harness makers, have vanished forever. There were more than 11 million farm workers in the United States at the beginning of the 20th century; because of high productivity growth in agriculture, there are fewer than 1 million today. By contrast, there are now more than 3 million truck, bus, and professional car drivers; there were none in 1900. There are more than 1 million computer programmers today; there were practically none in 1960.

Even for those with the right skills, the firm where they work may be replaced by a more efficient firm, or the product their firm was selling may be replaced by another product. This tension between the benefits of technological progress for consumers (and, by implication, for firms and their shareholders) and the risks for workers is well captured in the cartoon below. The tension between the large gains for society from technological change and the potentially large costs for the workers who lose their jobs is explored in the Focus Box "Job Destruction, Churn, and Earnings Losses."

FOCUS

Job Destruction, Churn and Earnings Losses

Technological progress is good for consumers, but it can be tough on the workers who lose their jobs. This is documented in a study by Steve Davis and Till von Wachter, who used records from the Social Security Administration between 1974 and 2008 to look at what happens to workers who lose their job as a result of a mass layoff.

Davis and von Wachter first identified all the firms with more than 50 workers where at least 30% of the workforce was laid off during one quarter, an event they call a mass layoff. Then they identified the laid-off workers who had been employed at that firm for at least three years; they classified these as long-term employees. They compared the labor market experience of long-term employees who were laid off in a mass layoff to similar workers in the labor force who did not separate in the layoff year or in the next two years. Finally, they compared the workers who experience a mass layoff in a recession to those who experience a mass layoff in an expansion.

Figure 1 summarizes their results. The year 0 is the year of the mass layoff; years 1, 2, 3, and so on are the years after, and the years with negative numbers are those before.

Long-term employees saw their earnings rise relative to the rest of society prior to the mass layoff event. Having a long-term job at the same firm is good for an individual's wage growth. This is true in both recessions and expansions.

Look at what happened in the first year after the layoff. If the mass layoff happened during a recession, a laid-off worker's earnings fell by 40 percentage points relative to a worker who did not experience a mass layoff. For the less unfortunate worker who experienced a mass layoff during an expansion, the fall in relative earnings was smaller but was still 25 percentage points. The conclusion: A mass layoff causes enormous relative earnings declines whether it occurs in a recession or an expansion.

Figure 1 makes another important point. The decline in relative earnings of workers who were part of a mass layoff persisted for years after the layoff. Beyond 5 years or even up to 20 years after the mass layoff, workers who experienced it suffered a sustained decline in relative earnings of about 20 percentage points if the mass layoff took place in a recession and about 10 percentage points if it took place during an expansion. A mass layoff is thus associated with a substantial decline in lifetime earnings.

It is not hard to explain why such earnings losses are likely, even if the size of the loss is surprising. The workers who have spent a considerable part of their career at the same firm have specific skills—skills that are most useful in that firm or industry. These skills are likely to be less valuable elsewhere.

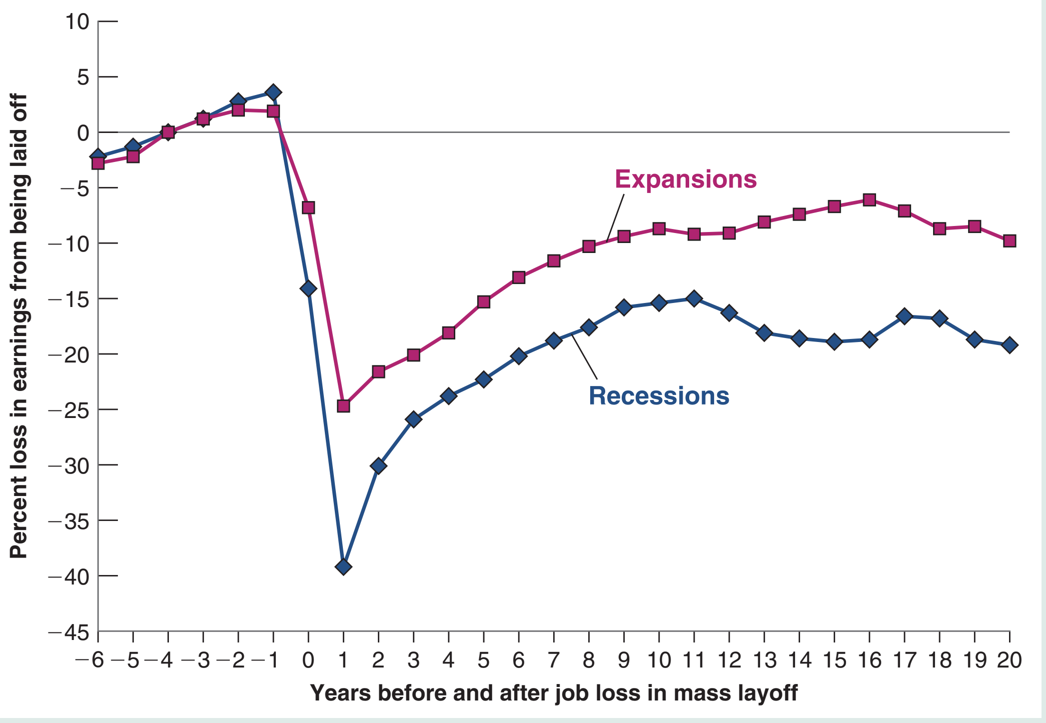
Other studies have found that in families that experience a mass layoff, the worker has a less stable employment path (more periods of unemployment), poorer health outcomes, and children who have a lower level of educational achievement and higher mortality when compared to the workers who have not experienced a mass layoff. These are additional costs associated with mass layoffs.

So, although technological change is the main source of growth in the long run and clearly enables a higher standard of living for the average person, workers who experience mass layoffs are clear losers. It is not surprising that technological change can and does generate anxiety.

Figure 1

Earnings Losses of Workers Who Experience a Mass Layoff

Source: Steven J. Davis and Till M. von Wachter, "Recessions and the Cost of Job Loss," National Bureau of Economics Working Paper No. 17638, 2011. Used with Permission.



The Increase in Wage Inequality

For those in growing sectors or those with the right skills, technological progress leads to new opportunities and higher wages. But for those in declining sectors or those with skills that are no longer in demand, technological progress can mean the loss of their job, a period of unemployment, and possibly much lower wages. The last 35 years in the United States have seen a large increase in wage inequality. Most economists believe that the main culprit behind this increase is technological change.

Figure 13-2 shows the evolution of relative wages for various groups of workers, by education level, since 1973. The figure is based on information about individual workers from the Current Population Survey. Each of the lines in the figure shows the evolution of the wage of workers with a given level of education—some high school, high school diploma, some college, college degree, advanced degree—*relative to* the wage of workers who only have high school diplomas. All relative wages are further divided by their value in 1973, so the resulting wage series are all equal to one in 1973. The figure yields a striking conclusion:

We described the Current Population Survey and some of its uses in Chapter 7.

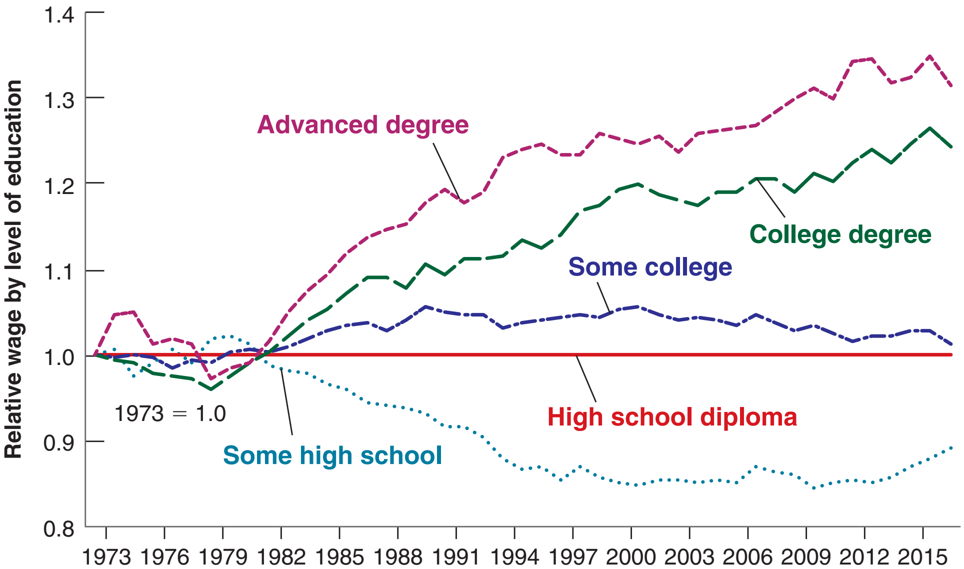


Figure 13-2

Evolution of Relative Wages by Education Level since 1973

Since the early 1980s, the relative wages of workers with a low education level have fallen; the relative wages of workers with a high education level have risen.

Source: Economic Policy Institute Data Zone. www.epi.org/types/data-zone/.

Starting around the early 1980s, workers with low levels of education have seen their relative wage steadily fall over time, whereas workers with high levels of education have seen their relative wage steadily rise. At the bottom end of the education ladder, the relative wage of workers who have not completed high school has declined by 11% since the early 1980s. This implies that, in many cases, these workers have seen a drop not only in their relative wage but in their absolute real wages as well. At the top end of the education ladder, the relative wage of those with an advanced degree has increased by 31%. In short, wage inequality has increased a lot in the United States over the last 30 years.

Note the small decrease in wage inequality in 2017 (the last year in the figure). Good news? Too early to tell.

The Causes of Increased Wage Inequality

What are the causes of this increase in wage inequality? The main force behind the increase in the wage of high-skill relative to low-skill workers is a steady increase in the demand for high-skill workers relative to the demand for low-skill workers. This trend in relative demand is not new, but it appears to have strengthened. Until the 1980s it was largely offset by a steady increase in the relative supply of high-skill workers: A steadily larger proportion of children finished high school, went to college, finished college, and so on. Since the early 1980s, however, relative supply has continued to increase, but not fast enough to match the continuing increase in relative demand. The result has been a steady increase in the relative wage of high-skill workers versus low-skill workers. The Focus Box "The Long View: Technology, Education, and Inequality" shows how not only the demand but also the supply of skills shaped the evolution of wage inequality in the United States during the 20th century.

This leads to the next question: What is behind this steady shift in relative demand?

- One line of argument has focused on the role of international trade. US firms that employ higher proportions of low-skill workers, the argument goes, are increasingly driven out of markets by imports from similar firms in low-wage countries. Alternatively, to remain competitive, firms must relocate some of their production to low-wage countries. In both cases, the result is a steady decrease in the relative demand for low-skill workers in the United States. There are clear similarities between the effects of trade and the effects of technological progress. Although both trade and technological progress are good for the economy as a whole, they lead nonetheless to structural change and make some workers worse off.

  Pursuing the effects of international trade would take us too far afield. For a more thorough discussion of who gains and who loses from trade, look at the text by Paul Krugman, Maurice Obstfeld, and Marc Melitz, *International Economics*, 10th ed. (2014).

  There is no question that trade is partly responsible for increased wage inequality. A closer examination, however, shows that trade accounts for only part of the shift in relative demand. The most telling fact countering explanations based solely on trade is that the shift in relative demand toward high-skill workers is present even in sectors that are not exposed to foreign competition.
- The other line of argument has focused on **skill-biased technological progress**. New machines and new methods of production, the argument goes, require more and more high-skill, high-education workers. The development of computers requires workers to be increasingly computer literate. New methods of production require workers to be more flexible and better able to adapt to new tasks. Greater flexibility in turn requires more skills and more education. Unlike explanations based on trade, skill-biased technological progress can explain why the shift in relative demand appears to be present in nearly all sectors of the economy. At this point, most economists believe it is the dominant factor in explaining the increase in wage inequality.

The Long View: Technology, Education, and Inequality

FOCUS

For the first three-quarters of the 20th century, wage inequality declined. Then it started to rise and has kept growing since. Claudia Goldin and Larry Katz, two economists at Harvard University, point to education as a major factor behind the two different trends in inequality.

US educational attainment, measured by the completed schooling levels of successive generations of students, was exceptionally rapid during the first three-quarters of the century. But educational advance slowed considerably for young adults beginning in the 1970s and for the overall labor force by the early 1980s. For generations born from the 1870s to about 1950, every decade was accompanied by an increase of about 0.8 year of education. During that 80-year period the vast majority of parents had children whose educational attainment greatly exceeded theirs. A child born in 1945 would have been in school 2.2 years more than his or her parents born in 1921. But a child born in 1975 would have been in school just half a year more than his or her parents born in 1951.

Underlying the decision to stay in school longer were clear economic incentives. As shown in Figure 1, the return to one more year of college education (meaning the increase in the average wage for a worker with one more year of college education) was high in the 1940s: 11% for young men and 10% for all men. This induced US families to keep their children in school longer and then send them to college. The increase in the supply of educated workers lowered both the returns to education and the wage differentials. By 1950, the return to one more year of college education had fallen back to 8% for young men, 9% for all men. But by 1990, rates of return were back to their 1930s levels. The return to a year of college today is higher than in the 1930s.

There are two lessons to be drawn from this evidence:

The first is that technological progress, even when skill-biased, that is accompanied by an increase in the demand for skilled and educated workers does not necessarily increase economic inequality. For the first three-quarters of the 20th century, the increase in demand for skills was more than met by an increase in the supply of skills, leading to decreasing inequality. Since then, demand growth has continued, whereas supply growth has decreased, leading once again to increasing inequality.

The second is that, although market forces provide incentives for demand to respond to wage differentials, institutions are also important. For most Americans in the early 20th century, access to schooling, at least through high school, was largely unlimited. Education was publicly provided and funded and thus free of direct charge, except at the highest levels. Even the most rural Americans could send their children to public secondary schools, although African Americans, especially in the South, were often excluded from various levels of schooling. This has made an essential difference.

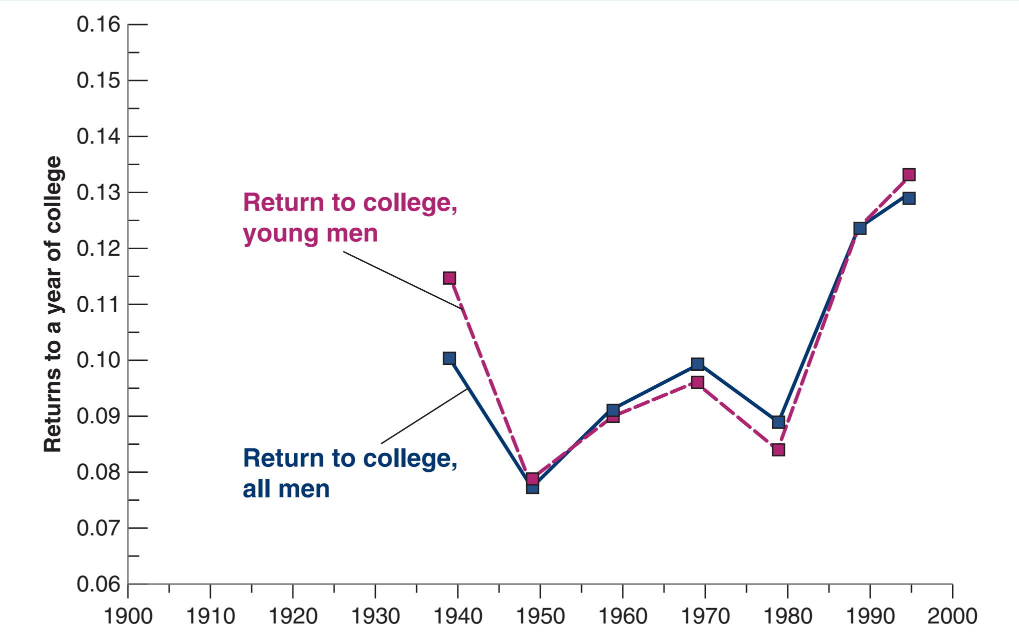


Figure 1

Wage Differentials and the Returns to Education, 1939 to 1995

Source: Claudia Goldin and Larry F. Katz, "Decreasing (and then Increasing) Inequality in America: A Tale of Two Half Centuries," In: Finis Welch *The Causes and Consequences of Increasing Inequality*. Chicago: University of Chicago Press; 2001.pp. 37–82. Used with Permission.

Inequality and the Top 1%

We have focused on wage inequality, the distribution of wages across wage earners. Another dimension of inequality, however, is the proportion of income that accrues to the richest households (e.g., those in the top 1% of the income distribution). When we consider inequality at very high levels of income, wages are not a good measure of income. Rich rentiers derive their income from the assets they own. Successful entrepreneurs derive a large fraction of their income (sometimes almost all of it) not from wages but from capital income and capital gains. This is because they are typically paid both through wages and through company shares that they can then sell (with some limitations) at a profit.

The evolution of the top 1% share, shown in Figure 13-3, is striking. The share of income going to households in the top 1% steadily decreased from the 1930s until the late 1970s. Since then, however, it has increased sharply, from about 9% to 22% today. By contrast, the share of the bottom 50% has sharply decreased, from about 21% in the early 1970s (data are not available before 1965) to less than 13% today.

To be in the top 1% in the United States in 2014, a household had to earn more than $408,000. To be in the top 0.1%, it had to earn more than $1.6 million. To be in the top 0.01%, it had to earn more than $7.8 million.

Looking within the top 1%, the numbers are even more striking: While the top 1% account for 22% of income, the top 0.1% alone account for 17%, and the top 0.01% account for 5.1%. Looking at wealth, inequality is even stronger. The top 0.01% account for 11% of total wealth. Inequality in the United States measured this way is "probably higher than in any other society at any time in the past, anywhere in the world," writes Thomas Piketty, whose book *Capital in the 21st Century* topped the list of best-selling books worldwide when it was published in 2014.

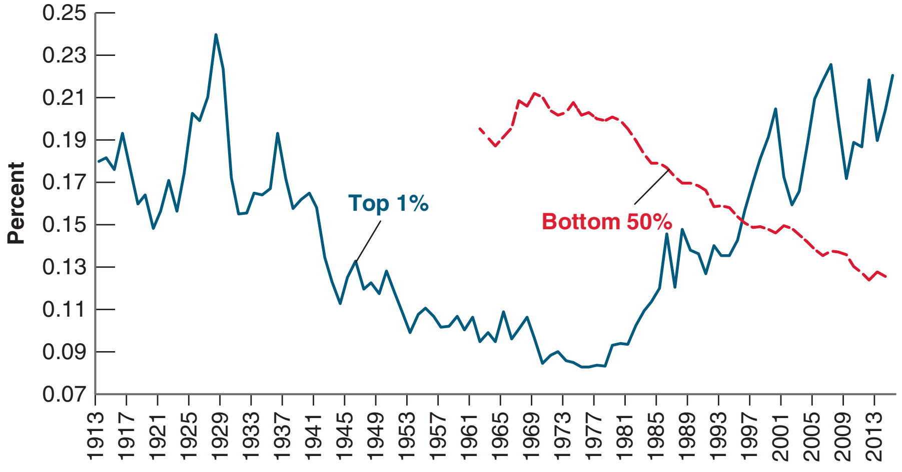
Who are the people in those high income groups? In the top 1%, the list is the one you would expect—doctors, lawyers, executives, owners of medium-size businesses. What about the 0.01%? One image is of heirs, living mostly on the income from their capital. Another is of successful entrepreneurs, founders of businesses like Amazon, Facebook, and Google. To get a sense of which image fits best, one can look at the yearly list put together by Forbes of the 400 wealthiest Americans (who account for 2.5% of the 16,000 households in the top 0.01% and have an average wealth of about $400 million). In that list, about one-third has inherited a substantial proportion of their wealth, and two-thirds are mostly self-made entrepreneurs and high level executives.

Why have these entrepreneurs and executives become so rich? Piketty points to bad corporate governance—company boards who grant their chief executives exorbitant pay packages. Above a certain level, he argues, it is hard to find in the data any link between

Figure 13-3

The Evolution of the Top 1% and the Bottom 50% Income Shares in the United States since 1913

Source: World Inequality Database. https://wid.world/



pay and performance. Although there is plenty of anecdotal evidence for such excesses, another factor is clearly at work, and it is again the nature of technological progress. Many new sectors exhibit strong increasing returns to scale: their costs are largely fixed costs and do not depend very much on the number of users. Thus, the more users there are, the smaller the cost per user. The result is the emergence of very large firms with a very large number of customers. At the end of 2018, the number of Facebook users was 2.3 billion, and the number of subscribers to Amazon Prime was 100 million. The same is true in finance, where the largest funds often manage hundreds of billions or even trillions of dollars. At the end of 2018, BlackRock, the largest asset management fund, managed $6 trillion of assets. Given the size of these firms and the size of their profits, it is not surprising that the founders and the top executives have such large income and large wealth.

Growth and Inequality

Is it possible for countries to sustain growth without having ever-increasing inequality? The trends we just saw are not reassuring. But there are two reasons to not be too pessimistic.

The first is that not all advanced economies are like the United States. Wage inequality and the share of the top 1% have been rising in most countries, but typically much less than in the United States. The second is that we must distinguish between **market income inequality**, which is income inequality before taxes and transfers, and **disposable income inequality**, which is income after taxes and transfers. So far, we have looked at market inequality, but what matters in the end is disposable income inequality. And here again, not all advanced economies are like the United States.

Figure 13-4, which compares the evolution of both types of inequality in the United States and in France, makes these points clearly. It plots the evolution of a standard measure of inequality, known as the **Gini coefficient**, or the Gini for short. The construction of the coefficient is explained in the Focus Box “Inequality and the Gini Coefficient,” but its general characteristics are simple: the coefficient ranges between 0 and 1. A coefficient of 0 means complete equality, with everybody having the same income; a coefficient of 1 means complete inequality, with one person receiving all the income, and everybody else receiving nothing. (In practice, the coefficient ranges from 0.2 for the most equal countries to 0.6 for the most unequal.)

The name comes from Corrado Gini, an Italian statistician who developed the coefficient in the early part of the 20th century.

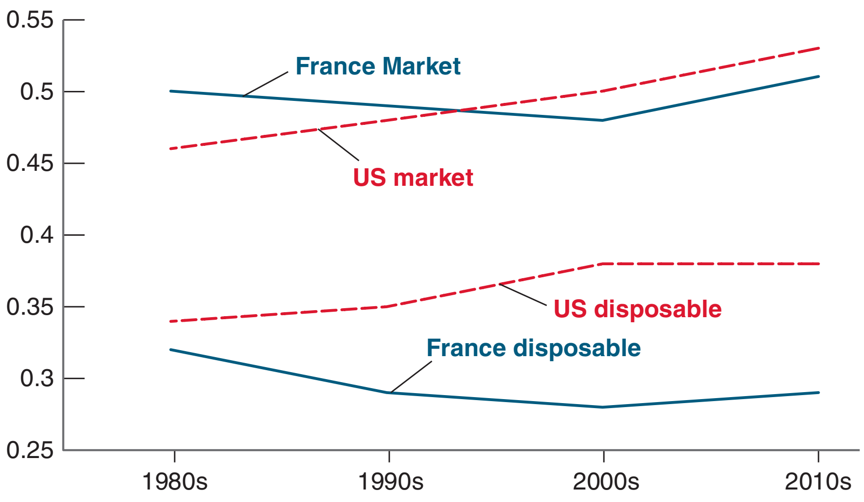


Figure 13-4

The Evolution of Gini Coefficients for Market Income and for Disposable Income in France and the United States since the 1980s

Source: Income Inequality by Max Roser and Esteban Ortiz-Ospina, First published in December 2013; updated October, 2016. (https://ourworldindata.org/income-inequality).

FOCUS

Inequality and the Gini Coefficient

The Gini coefficient is a measure of inequality. The best way to describe it is with the picture below. The share of population ranked by increasing income is plotted on the horizontal axis, from zero to one. The share of income is plotted on the vertical axis, from zero to one. The relation giving the share of income associated with a given population share is given by the convex curve in red. The Gini coefficient is given by the ratio of area A to area (A + B).

To see what it implies, consider an economy where everybody has the same income. The share of income is always equal to the share of population, so the relation between the two is given by the diagonal line. Area A is equal to zero, and thus the Gini coefficient is also equal to zero. Take the opposite extreme and consider an economy where all the income goes to the richest individual. Then the relation between the two is a flat line covering the horizontal axis, ending with a jump from zero to one at the end. Area B is equal to zero, and so the Gini coefficient is equal to one.

These are obviously extreme and counterfactual cases, and Gini coefficients typically range between 0.2 and 0.6. In general, the more inequality, the smaller the share of income going to a given share of population. If, for example, inequality is such that a large proportion of the population is very poor, the curve will be very flat starting from 0. If inequality is such that a small proportion of the population is very rich, the curve will be very steep close to 1.

Summarizing the complexity of the distribution of income in one number has obvious limits. Inequality can take different forms, which may all lead to the same Gini coefficient. But it is a simple and intuitive statistic and one that allows comparisons of income distributions across time or across countries.

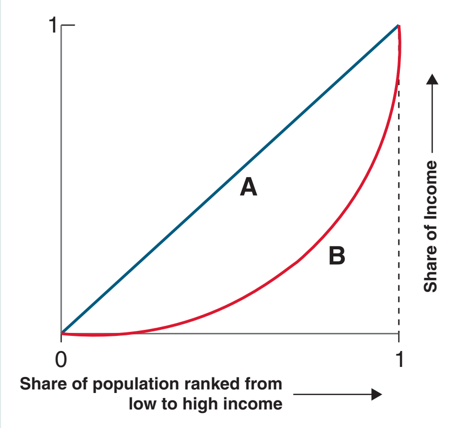


Figure 1

The two series at the top show the evolution of the Gini coefficients for *market income inequality* for the United States in red and for France in blue. The Gini for the United States goes up over time, reflecting the wage and top 1% income increasing inequalities we discussed earlier. The Gini for France, however, remains roughly constant. This is an important fact as both countries are affected in similar ways by globalization and have access to the same technologies. This suggests that, beyond trade and technological progress, other factors such as institutions and social norms about "acceptable pay" influence market inequality.

For example, looking at banks of the same size, French bankers are paid much less than their American counterparts. ▶

The two series at the bottom show the evolution of the Gini coefficients for *disposable income inequality* for the two countries. Again, there is a striking contrast between them. While the Gini coefficient for the United States goes up, by about half of the increase in the market Gini coefficient, the Gini coefficient for France is lower today than it was in the 1980s. In other words, redistribution through taxes and transfers has led to a (small) decrease in inequality. This is also an important fact. It suggests that, until now, redistribution has been enough to limit inequality. (The Gini coefficient is a simple but rough measure, and the fact that it has decreased does not imply that all groups have benefited equally from growth; some workers or groups of workers have seen a relative decrease in income.)

There are strong moral reasons to want to limit inequality, to achieve what is known as **inclusive growth**. There are also good political reasons for governments to do so. The evidence suggests that perceived inequality, be it about relative wages or about 1%, has played a role in the rise of populism in many countries. What can governments do?

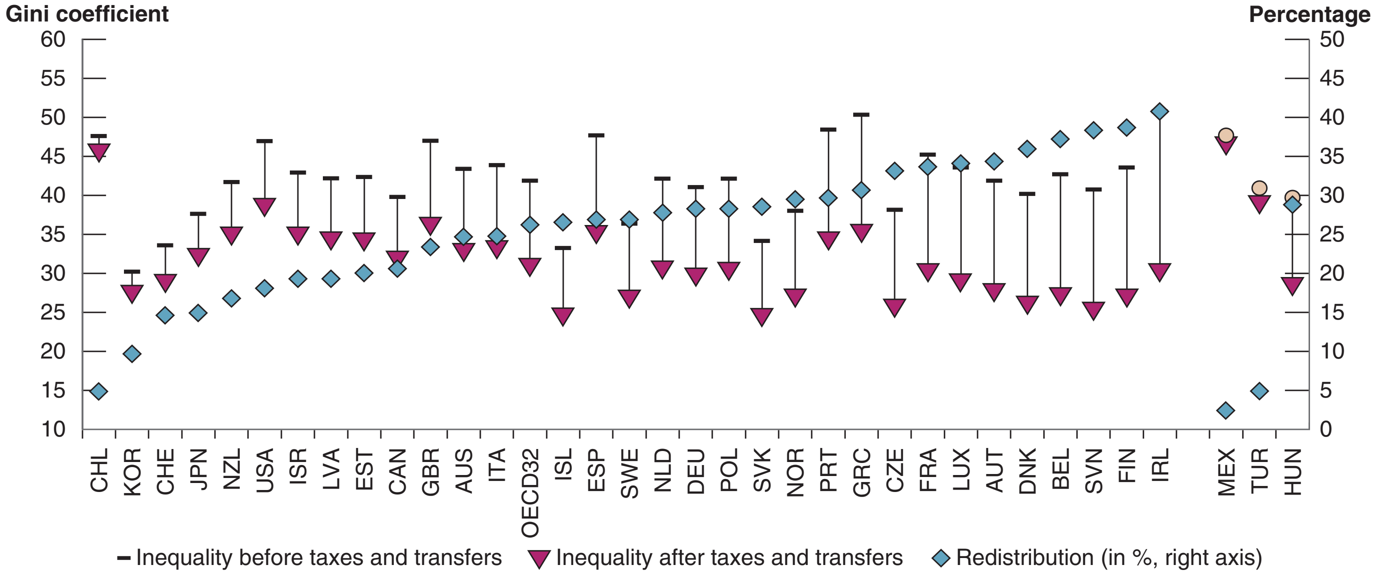


Figure 13-5

Market Income, Disposable Income, and Redistribution, across OECD Countries

Source: Market income, Disposable income, and redistribution, across OECD countries. Used with Permission. https://oecde-coscope.blog/2019/02/14/income-redistribution-across-oecd-countries-main-findings-andpolicy-implications/.

They can try to reduce market income inequality, through better education, through minimum wages, through rules about governance within firms. Or, given market income inequality, they can reduce disposable income inequality through progressive taxation, including the use of a negative income tax to increase revenue at the low end of the wage distribution, and through transfers such as food stamps and unemployment benefits. All governments use some of these tools, and redistribution can substantially reduce inequality. Evidence from a recent OECD study is given in Figure 13-5, which shows the Gini coefficients (multiplied by 100) associated with market income and with disposable income inequality, and by implication shows the degree of redistribution, for each OECD country.

The figure yields three conclusions.

- Countries vary substantially in their degree of market inequality, with the United States near the top. Gini coefficients range from lows of 30% in Korea and 34% in Iceland to 50% in Ireland and Greece, and 48% in the United States. As most of these countries are open to trade and have access to similar technologies, it shows the importance of other factors in determining the outcome.
- The degree of redistribution, defined as the difference in the market income Gini and the disposable income Gini, divided by the market income Gini, varies from lows of 5% in Chile and 10% in Korea to 38% in Finland and 40% in Ireland. The United States stands toward the low end, at 18%. This shows the importance of political choices in determining the final outcome.
- As a result of redistribution, disposable income Ginis are lower than market income Ginis, often by a large amount. For example, in Ireland market income inequality is high but, as a result of redistribution, disposable income inequality is low. The variation in disposable income inequality, from 25% to 45%, is smaller than in market income inequality, which ranges from 30% to 50%, but it remains substantial. In the United States, market income inequality is high and redistribution limited.

If inequality continues to increase, governments will have to do more. This is why, for example, it appears that discussions about the marginal income tax rate, or higher inheritance taxes, are likely to play an important role in the 2020 US elections, and why discussions about the provision of a universal income, which would give a minimum amount to people whether they work or not, are taking place in countries like Italy or Finland. Sustaining growth while limiting or reducing inequality is one of the most important challenges facing policymakers today.

13-4 CLIMATE CHANGE AND GLOBAL WARMING

It is well known that markets work poorly when there are externalities. In the context of growth, a major externality is the emission of greenhouse gases. These gases have a cost but it is not taken into account by firms and by people when they make decisions about, for example, whether to choose one technology or another, or whether to buy a car.

The most important greenhouse gas is carbon dioxide (CO_2). To understand its role, consider the light sent by the sun to the earth. Some of this sunlight is absorbed by the earth and some is reradiated back to space. The amount of CO_2 in the atmosphere determines the strength of the "greenhouse effect"—and how much is absorbed by the earth and how much is reradiated. How much is absorbed is in turn a major determinant of the global temperature. Were there no carbon dioxide or other greenhouse gases in the atmosphere, too little sunlight would be absorbed, too much would be reradiated, and it is estimated that the global temperature would be about—18 degrees Celsius (roughly zero Fahrenheit). If, however, there is too much carbon dioxide, the temperature of the earth will increase, leading to global warming. Substantial global warming would be catastrophic, leading to a large increase in sea levels, extreme weather events, and parts of the world becoming uninhabitable.

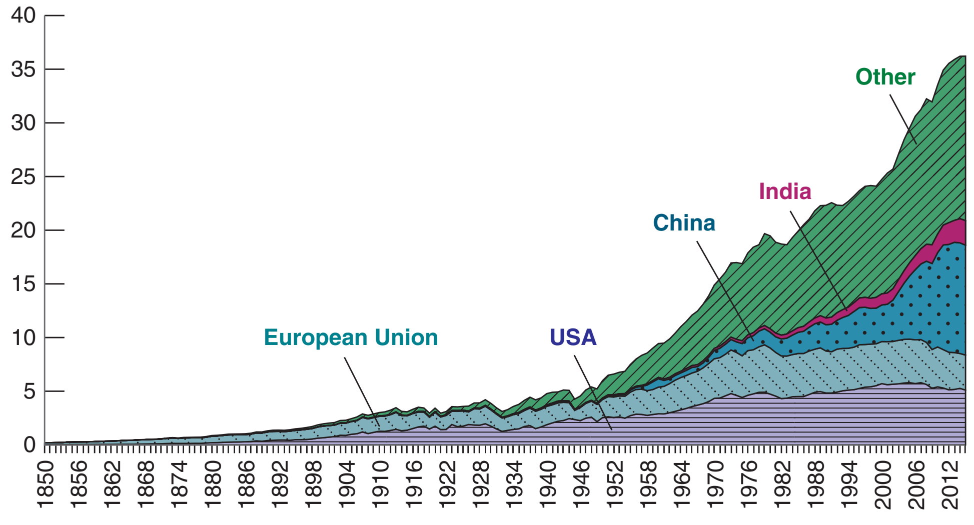
Since the Industrial Revolution the use of fossil fuels (primarily coal) has led to a large increase in CO_2 emissions. At the same time, there has been a steady increase in the average global temperature. The evidence for each one is shown in Figures 13-6 and 13-7.

Figure 13-6 shows the increase in CO_2 emissions since 1850, by region of the world. It shows a very large increase in emissions over time, starting with Europe during the Industrial Revolution and the United States taking the relay later. Since the 2000s however, emissions from China have exceeded the emissions of both Europe and the United States combined.

Figure 13-6

CO_2 Emissions since 1850, by Region

Source: Annual carbon dioxide (CO_2) emissions measured in billion tons (Gt) per year. Carbon Dioxide Information Analysis Center (CDIAC), http://cdiac.ornl.gov/CO2_Emission/.



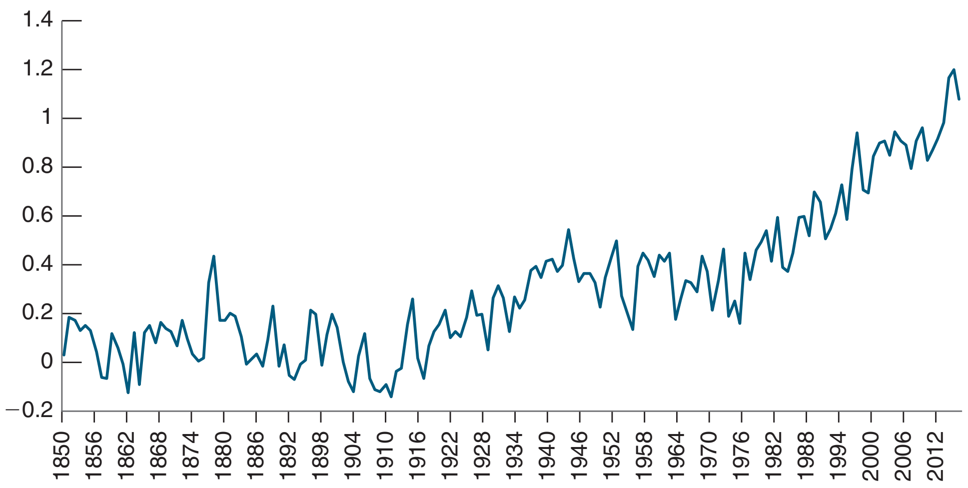


Figure 13-7

Global Average Temperature since 1850

Source: Met Office Hadley Center, www.metoffice.gov.uk/hadobs/hadcrut4/index.html.

Figure 13-7 shows the increase in the global average temperature since 1850, as a deviation from the temperature in 1850. The temperature has increased by roughly 1.2 degrees Celsius since 1850, with much of the increase taking place since the late 1970s.

The fact that both CO_2 levels and the global temperature have increased is indisputable; only cranks (who unfortunately exist) deny it. One may, however, question whether the increase in global temperature is due to the increase in CO_2 levels or to other factors. Nearly all scientists believe that the relation is indeed causal. And even it is not, global warming is a major concern, and limiting CO_2 can alleviate the problem.

Turning from the past to the future, the main question is how fast global warming is likely to happen. The consensus forecast is shown in Figure 13-8. If there were no climate policies at all, temperature would be forecast to increase, relative to its value before the pre–Industrial Revolution, by 4.1 to 4.8 degrees Celsius by 2100, making the planet largely uninhabitable. Under existing policies, the forecast is for an increase of 3.1 to 3.7 degrees, still a catastrophic outcome. Under current pledges, the forecast is for an increase of 2.6 to 3.2 degrees. Pledges, however, are not binding, and the evidence is that many countries are not delivering. Even if they were, the pledged changes would be insufficient to limit the increase to 1.5% to 2%, the level considered acceptable by scientists.

What policies should be adopted by countries to limit the increase? There is wide agreement among economists that the best policy is to put a price on carbon emissions, in effect to internalize the externality. The reason it has not happened is probably fourfold. First, until recently, global warming was not seen as a high priority. Indeed, some governments are still reluctant to accept the reality. Second, any policy that implies a cost today in exchange for difficult-to-assess benefits far in the future is politically difficult to sell. Third, because the relatively poor in each country tend to have cars that are older and have larger emissions, the policy, unless compensated by the right transfers, is regressive. Fourth, and probably most importantly, policy discussions lead to sharp tensions between emerging market and advanced economies. You can see why in Figure 13-6. China is now the main source of CO_2 emissions and thus would see the largest increase in costs. It argues that Europe and the United States were able to grow and emit earlier in time without paying this cost, and it is unfair to ask China now to pay the full cost.

As I write, a statement signed by more than 3,000 US economists, including 27 Nobel Laureates and 15 former chairs of the Council of Economic Advisers, argues for the introduction of a carbon tax.

Put more generally, it is difficult to achieve an agreement across all countries, and this has been reflected in the limited success of past climate conferences. Short of

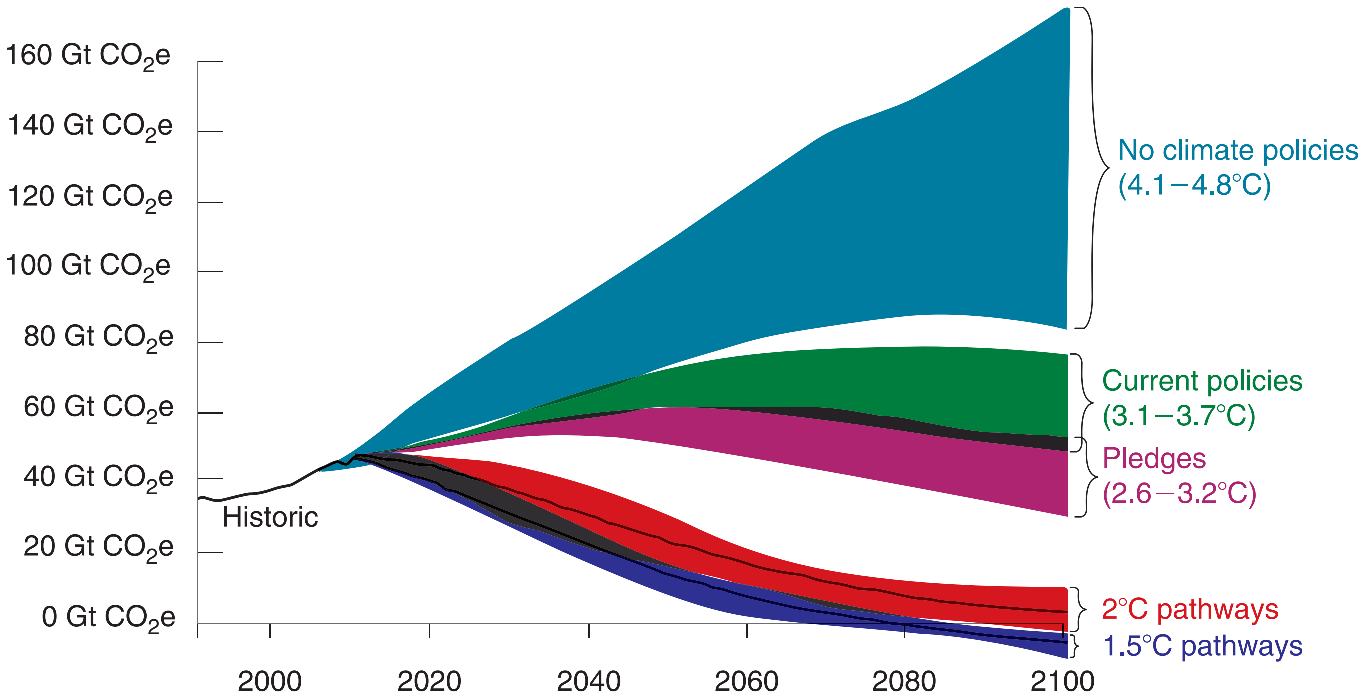


Figure 13-8

Global Warming Scenarios

Source: Max Roser, https://ourworldindata.org/uploads/2018/04/Greenhouse-gas-emission-scenarios-01.png.

William Nordhaus received the Nobel Prize for his work on climate change in 2018. ▶

an agreement, a solution has been proposed by William Nordhaus, of Yale University. Countries willing to adopt a carbon tax should do so and impose a carbon tariff on the goods imported from countries that do not have a carbon tax. This would in turn give incentives for those countries to adopt a carbon tax and no longer be subject to tariffs. Whether this or another solution is adopted, the problem will not go away.

SUMMARY

- There is a tension between the feeling of fast technological progress and the decrease in measured productivity growth. The decrease in measured productivity growth does not appear to be due to mismeasurement but to a true decline. In the past, technological progress came from the discovery of general-purpose technologies and their diffusion to many sectors in the economy. Nobody really knows whether digitization, the current general-purpose technology, will have the same effect.
- The worry that technological progress will lead to mass unemployment is an old worry that has not proven warranted in the past. Productivity is many times higher than it used to be and so is employment. Unemployment is low, even for low-skill workers. The future, however, could be different, and worries about robots are not necessarily wrong.
- Technological progress is a process of structural change. New products are introduced, others disappear. New firms are created, old firms disappear. While the process benefits people as consumers, it makes some of them worse off as workers. The evidence is that people who lose their jobs because their firm is not doing well end up with lower income for a long time. Skill-biased technological progress has led to increasing inequality between those with low skills or low education and those with more skills or more education.

- Increasing inequality has also taken the form of an increase in the relative income and wealth of the top 1% income group. Part of the reason is that some of the new technologies exhibit increasing returns to scale, leading to the presence of large firms and large incomes for the founders and executives.
- Market income inequality varies across countries, even across advanced countries exposed to similar globalization and technological progress trends, suggesting that it depends also on other factors, such as institutions or social norms. For given market income inequality, redistribution can also reduce disposable income inequality. There are again large differences in the degree of redistribution across countries.
- Climate change is perhaps growth's major challenge. The evidence is that the global average temperature has steadily increased and that this is most likely due to increased CO_2 emissions. So far, progress has been insufficient. The best policy would be the introduction of a carbon tax, but it faces difficult political and geopolitical obstacles.

KEY TERMS

general-purpose technology, 266
technological unemployment, 267
creative destruction, 269
skill-biased technological progress, 272
market income inequality, 275
disposable income inequality, 275
Gini coefficient, 275
inclusive growth, 276

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. It is clear that the rate of technological growth has declined in the last decade.
b. The change in employment and output per person in the United States since 1900 lends support to the argument that technological progress leads to a steady increase in employment.
c. Workers benefit equally from the process of creative destruction.
d. In the past two decades, the real wages of low-skill U.S. workers have declined relative to the real wages of high-skill workers.
e. Mismeasurement explains a large part of the decline in technological progress.
f. New technologies do generally displace workers with lower skills.
g. Innovations known as general-purpose technologies occur frequently.
h. The introduction of robots may actually increase total employment.
i. There is evidence that workers displaced by plant closures associated with technological change suffer a long-term wage loss.

2. How might the policy changes in parts a through e affect the wage gap between low-skill and high-skill workers in the United States?

a. Increased spending on computers in public schools.
b. Restrictions on the number of foreign temporary agricultural workers allowed to enter the United States.
c. An increase in the number of public colleges.
d. Tax credits in Central America for US firms.
e. Subsidies paid to business to introduce robots into the production process.

3. It is possible to extend the production function so that output is produced with labor input, N; capital input, K; and carbon-intensive energy input, E.

If we write a production function as $Y = N^{(1/3)}K^{(1/3)}E^{(1/3)}$

a. Using the input values below, does this production function exhibit constant returns to scale?

Units of N	Units of K	Units of E	Output
27	125	64	
54	250	128	
81	375	192	

b. Solve for the necessary amount of energy to produce the specified level of output in each row below using the production function $Y = N^{(1/3)}K^{(1/3)}E^{(1/3)}$

Units of N	Units of K	Units of E	Output
54	250		120
54	375		120
81	250		120

c. Comment on the following statement using your results from part b.
"Because of constant returns to scale, in order for output to rise, we must increase our consumption of energy."

d. Use words to describe how output can increase without an increase in the use of energy. The text points out that there is widespread agreement among economists that a tax on carbon emissions energy is the best way to reduce carbon emissions. How does this proposition fit into the production function above?

DIG DEEPER

4. Changes in the rate of technological growth

In Chapter 12 you computed a residual term representing the growth rate of technology using the expression

$$\text{residual} \equiv g_Y - [\alpha g_N + (1 - \alpha) g_K]$$

and the annual growth rates of output, g_Y; labor input, g_N; and capital input, g_K. What was the value of that residual for the period from 2000 to 2014?.

From the Penn World Tables, fill in the chart below:

Reminder: Use rgdpo for output (2011 US dollars); emp number of employees labor input; ck for the capital stock (2011 US dollars). The share of labor is labsh.

	Year 1985				Year 1999			
	Y	*N*	*K*	Share of labor	*Y*	*N*	*K*	Share of labor
United States								

Compute growth rates of *Y*, *N*, and *K* between 1985 and 1999. Do the same calculation for the period from 2000 to 2014 or use your results from Chapter 12.

a. Compare the growth rate of technology in the periods 1985–1999 and 2000–2014.

b. The FRED database provides the following variable as a measure of the unemployment rate of a low-skill group of labor force participants: High school graduates with no college aged 16–24 (series LNU04023068). This data begins in 1985. Find this series and describe the graph of this series from 1985 to 2018 (or the latest date available). What happened to the unemployment rate of this group from 2009 to 2013? Then download the series to a spreadsheet.

c. Now calculate the average unemployment rate of this group over these three periods: 1985–1999, 2000 to the end of the data, and 2017 to the end of your data. How do the three unemployment rates compare? Does your result support the proposition that more rapid technological change increases the unemployment rate of low-skill workers? Explain why you also looked at the years after 2017.

5. Technological progress, agriculture, and employment

Discuss the following statement: "Those who argue that technological progress does not reduce employment should look at agriculture. At the start of the last century, there were more than 11 million farm workers. Today, there are fewer than 1 million. If all sectors start having the productivity growth that took place in agriculture during the 20th century, no one will be employed a century from now."

6. Technology and the labor market

In the appendix to Chapter 7, we learned how the wage-setting and price-setting equations could be expressed in terms of labor demand and labor supply. In this problem, we extend the analysis to account for technological change.

Consider the wage-setting equation

$$W/P = F(u, z)$$

as the equation corresponding to labor supply. Recall that for a given labor force, L, the unemployment rate, u, can be written as

$$u = 1 - N/L$$

where N is employment.

a. Substitute the expression for u into the wage-setting equation.

b. Using the relation you derived in part a, graph the labor supply curve in a diagram with N on the horizontal axis and W/P, the real wage, on the vertical axis.

Now write the price setting equation as

$$P = (1 + m)MC$$

where MC is the marginal cost of production. To generalize somewhat our discussion in the text, we shall write

$$MC = W/MPL$$

where W is the wage and MPL is the marginal product of labor.

c. Substitute the expression for MC into the price-setting equation and solve for the real wage, W/P. The result is the labor demand relation, with W/P as a function of the MPL and the markup, m.

In the text, we assumed for simplicity that the MPL was constant for a given level of technology. Here, we assume that the MPL decreases with employment (again for a given level of technology), a more realistic assumption.

d. Assuming that the MPL decreases with employment, graph the labor demand relation you derived in part c. Use the same diagram you drew for part b.

e. What happens to the labor demand curve if the level of technology improves? (*Hint*: What happens to MPL when technology improves?) Explain. How is the real wage affected by an increase in the level of technology?

EXPLORE FURTHER

7. The churn

The Bureau of Labor Statistics presents a forecast of occupations with the largest job decline and the largest job growth. Examine the table at https://www.bls.gov/emp/tables/emp-by-detailed-occupation.htm.

a. Which occupations in decline can be linked to technological change? Which can be linked to foreign competition?

b. Which occupations that are forecast to grow can be linked to technological change? Which can be linked to demographic changes—in particular, the aging of the US population?

8. *Real wages*

The chapter has presented data on relative wages of high-skill and low-skill workers. In this question, we look at the evolution of real wages.

a. Go to the Web site of the most recent *Economic Report of the President* and find Table B-15. Look at the data on average hourly earnings for production and non-supervisory employees in 1982–1984 dollars (i.e., real hourly earnings) for the earliest and the latest data in this table. How do real hourly earnings in the earliest data compare to real hourly earnings in the latest year for which data are available? Are there periods in between where real hourly earnings in the private sector fell?
b. Given the data on *relative* wages presented in the chapter, what do your results from part b suggest about the evolution of *real* wages of low-skill workers since 1974? What do your answers suggest about the strength of the relative decline in demand for low-skill workers?
c. What might be missing from this analysis of worker compensation? Do workers receive compensation in forms other than wages?

9. *Income Inequality*

a. What evidence is presented in the text that income inequality has increased over time in the United States?
b. Use supply and demand of educated workers to explain the increase in income inequality.
c. Use supply and demand of less-educated workers to explain the increase in income inequality.
d. Do a Web search and contrast, if possible, the positions of the Democrats and the positions of the Republicans on whether increased income inequality is a problem in need of a policy solution.
e. There is some 2011 evidence on who married whom by level of education at www.theatlantic.com/sexes/archive/2013/04/college-graduates-marry-other-college-graduates-most-of-the-time/274654/. Explain how, if like-educated people are more likely to marry each other over time, this contributes to income inequality.

10. *Carbon emissions and growth*

There is a great deal of interest in carbon emissions because of global warming. The World Bank produces data on carbon emissions by country in a variety of forms. It is found at

https://data.worldbank.org/indicator/en.atm.co2e.pp.gd. One very interesting series is CO_2 emissions (kg per PPP $ of GDP). Download this data set to a spreadsheet. There are estimates of CO_2 emissions per dollar of real GDP, a variable sometimes called the carbon intensity of production. At the time of writing this data started in 1990 and ended in 2014.

a. Find the values for the United States. How has the carbon intensity of output in the United States changed from 1990 to the most recent data?
b. Sort the data in the spreadsheet to find the countries with the five largest values of emissions per dollar of GDP in 1990; in the latest year. Has the list of high emission countries changed?
c. Now reconsider the five countries with the highest intensity in 1990. Have they reduced carbon intensity over this period?
d. Go back to the Penn World Tables and use the data on GDP in PPP dollars to measure if carbon emissions are higher in the United States in 2014 than in 1990? Are carbon emissions in China higher in 1990 or in 2014?

FURTHER READINGS

- For more on the process of reallocation that characterizes modern economies, read *The Churn: The Paradox of Progress*, a report by the Federal Reserve Bank of Dallas (1993).
- For a fascinating account on how computers are transforming the labor market, read *The New Division of Labor: How Computers Are Creating the Next Job Market*, by Frank Levy and Richard Murnane (2004).
- For more statistics on various dimensions of inequality in the United States, a useful site is "The State of Working America," published by the Economic Policy Institute, at www.stateofworkingamerica.org/.
- For more on innovation and income inequality you can read, beyond Thomas Piketty's *Capital in the XXI Century* (2014), another piece by Thomas Piketty and Emmanuel Saez, "Income Inequality in the United States, 1913–1998." *The Quarterly Journal of Economics*, 118 (1): 1–41, and Emmanuel Saez (2013) "Striking it Richer: The Evolution of Top Incomes in the United States," mimeo UC Berkeley.
- For a more general view on technology and inequality, and one that comes from a slightly different perspective, you can also read "Technology and Inequality" by David Rotman, *MIT Technology Review*, October 21, 2014, available at www.technologyreview.com/featuredstory/531726/technology-and-inequality/.
- On the future of technological progress: The pessimistic view is presented by Robert Gordon, *The Rise and Fall of American Growth* (Princeton: Princeton University Press, 2016). A good video, summarizing his main arguments, is www.youtube.com/watch?v=PYHd7rpOTe8". For an optimistic view, read Erik Brynjolfsson and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (Norton and Company, 2014).

- On inequality: The World Inequality database is a major data collection with information on inequality over time and within and across countries in a very convenient form: https://wid.world/. On the 0.01% top income group, a nice summary is given by Howard Gold in http://review.chicagobooth.edu/economics/2017/article/never-mind-1-percent-lets-talk-about-001-percent.
- On climate change: The energy section of the site https://ourworldindata.org/has a large number of datasets and visually outstanding graphs on climate change. The Nobel lecture by William Nordhaus is also worth listening to: www.nobelprize.org/prizes/economic-sciences/2018/nordhaus/lecture/.

Expectations

The next three chapters cover the first extension of the core. They look at the role of expectations in output fluctuations.

Chapter 14

Chapter 14 focuses on the role of expectations in financial markets. It introduces the concept of expected present discounted value, which plays a central role in the determination of asset prices and in consumption and investment decisions. Using this concept, it studies the determination of bond prices and bond yields. It shows how we can learn about the course of expected future interest rates by looking at the yield curve. It then turns to stock prices and shows how they depend on expected future dividends and interest rates. Finally, it discusses whether stock prices always reflect fundamentals or may instead reflect bubbles or fads.

Chapter 15

Chapter 15 focuses on the role of expectations in consumption and investment decisions. It shows how consumption depends partly on current income, partly on human wealth, and partly on financial wealth. It shows how investment depends partly on current cash flow and partly on the expected present value of future profits.

Chapter 16

Chapter 16 looks at the role of expectations in output fluctuations. Starting from the IS-LM model, it extends the description of goods-market equilibrium (the IS relation) to reflect the effect of expectations on spending. It revisits the effects of monetary and fiscal policy on output, taking into account their effect through expectations.

This page is intentionally left blank

14

Financial Markets and Expectations

Our focus throughout this chapter will be on the role expectations play in the determination of asset prices, from bonds to stocks to houses. We discussed the role of expectations informally at various points in the core. It is now time to do it more formally. As you will see, not only are asset prices affected by current and expected future activity, but they in turn affect decisions that influence current economic activity. Understanding their determination is thus central to understanding fluctuations.

Section 14-1 introduces the concept of expected present discounted value, which plays a central role in the determination of asset prices and in consumption and investment decisions.

Section 14-2 looks at the determination of bond prices and bond yields. It shows how bond prices and yields depend on current and expected future short-term interest rates. It then shows how we can use the yield curve to learn about the expected course of future short-term interest rates.

Section 14-3 looks at the determination of stock prices. It shows how stock prices depend on current and expected future profits, as well as on current and expected future interest rates. It then discusses how movements in economic activity affect stock prices.

Section 14-4 looks more closely at the relevance of fads and bubbles—episodes in which asset prices (stock or house prices, in particular) appear to move for reasons unrelated to either current and expected future payments or interest rates.

If you remember one basic message from this chapter, it should be: Expectations determine bond and stock prices. ▶▶▶

14-1 EXPECTED PRESENT DISCOUNTED VALUES

To understand why present discounted values are important, consider the problem facing a manager who is deciding whether or not to buy a new machine. On the one hand, buying and installing the machine involves a cost today. On the other, the machine allows for higher production, higher sales, and higher profits in the future. The question facing the manager is whether the value of these expected profits is higher than the cost of buying and installing the machine. This is where the concept of expected present discounted value comes in handy. The **expected present discounted value** of a sequence of future payments is the value today of this expected sequence of payments. Once the manager has computed the expected present discounted value of the sequence of profits, her problem becomes simpler. She compares two numbers, the expected present discounted value and the initial cost. If the value exceeds the cost, she should go ahead and buy the machine. If it does not, she should not.

The practical problem is that expected present discounted values are not directly observable. They must be constructed from information on the sequence of expected payments and expected interest rates. Let's look at the mechanics of construction.

Computing Expected Present Discounted Values

In this section, to keep things simple, we ignore the issue of risk, which we discussed at length in Chapter 6. We return to it in the next section.

Denote the one-year nominal interest rate by i_t, so lending one dollar this year implies getting back $1 + i_t$ dollars next year. Equivalently, borrowing one dollar this year implies paying back $1 + i_t$ dollars next year. In this sense, one dollar this year is worth $1 + i_t$ dollars next year. This relation is represented graphically in the first line of Figure 14-1.

Turn the argument around and ask: How much is one dollar *next year* worth this year? The answer, shown in the second line of Figure 14-1, is $1/(1 + i_t)$ dollars. Think of it this way: If you lend $1/(1 + i_t)$ dollar this year, you will receive $1/(1 + i_t)$ times $(1 + i_t) = 1$ dollar next year. Equivalently, if you borrow $1/(1 + i_t)$ dollar this year, you will have to repay exactly one dollar next year. So, one dollar next year is worth $1/(1 + i_t)$ dollar this year.

i_t: discount rate. $1/(1 + i_t)$: discount factor. If the discount rate goes up, the discount factor goes down.

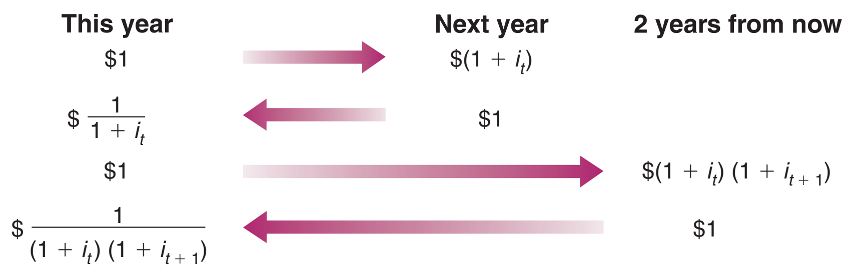
More formally, we say that $1/(1 + i_t)$ is the *present discounted value* of one dollar next year. The word *present* comes from the fact that we are looking at the value of a payment next year in terms of dollars *today*. The word *discounted* comes from the fact that the value next year is discounted, with $1/(1 + i_t)$ being the **discount factor**. (The rate at which you discount, in this case the nominal interest rate, i_t, is sometimes called the **discount rate**.)

The higher the nominal interest rate, the lower the value today of a dollar received next year. If $i = 5\%$, the value this year of a dollar next year is $1/1.05 \approx 95$ cents. If $i = 10\%$, the value today of a dollar next year is $1/1.10 \approx 91$ cents.

Now apply the same logic to the value today of a dollar received *two years from now*. For the moment, assume that current and future one-year nominal interest rates are known with certainty. Let i_t be the nominal interest rate for this year, and i_{t+1} be the one-year nominal interest rate next year.

Figure 14-1

Computing Present Discounted Values



If today you lend one dollar for two years, you will get $(1 + i_t)(1 + i_{t+1})$ dollars two years from now. Put another way, one dollar today is worth $(1 + i_t)(1 + i_{t+1})$ dollars two years from now. This relation is represented in the third line of Figure 14-1.

What is one dollar two years from now worth today? By the same logic as before, the answer is $1/(1 + i_t)(1 + i_{t+1})$ dollars. If you lend $1/(1 + i_t)(1 + i_{t+1})$ dollars this year, you will get exactly one dollar in two years. So, the *present discounted value of a dollar two years from now* is equal to $1/(1 + i_t)(1 + i_{t+1})$ dollars. This relation is shown in the last line of Figure 14-1. If, for example, the one-year nominal interest rate is the same this year and next and equal to 5%, so $i_t = i_{t+1} = 5\%$, then the present discounted value of a dollar in two years is equal to $1/(1.05)^2$ or about 91 cents today.

A General Formula

Having gone through these steps, it is easy to derive the present discounted value for the case where both payments and interest rates can change over time.

Consider a sequence of payments in dollars, starting today and continuing into the future. Assume for the moment that both future payments and future interest rates are known with certainty. Denote today's payment by $\$z_t$, the payment next year by $\$z_{t+1}$, the payment two years from today by $\$z_{t+2}$, and so on.

The present discounted value of this sequence of payments—that is, the value in today's dollars of the sequence of payments—which we shall call $\$V_t$ is given by

$$\$V_t = \$z_t + \frac{1}{(1 + i_t)}\$z_{t+1} + \frac{1}{(1 + i_t)(1 + i_{t+1})}\$z_{t+2} + \cdots$$

Each payment in the future is multiplied by its respective discount factor. The more distant the payment, the smaller the discount factor, and thus the smaller today's value of that distant payment. In other words, future payments are discounted more heavily, so their present discounted value is lower.

We have assumed that future payments and future interest rates were known with certainty. Actual decisions, however, must be based on expectations of future payments rather than on actual values for these payments. In our previous example, the manager cannot be sure how much profit the new machine will bring, nor does she know what interest rates will be in the future. The best she can do is get the most accurate forecasts she can and then compute the *expected present discounted value* of profits based on these forecasts.

How do we compute the expected present discounted value when future payments and interest rates are uncertain? Basically in the same way as before, but by replacing the *known* future payments and *known* interest rates with *expected* future payments and *expected* interest rates. Formally: Denote expected payments next year by $\$z^e_{t+1}$, expected payments two years from now by $\$z^e_{t+2}$, and so on. Similarly, denote the expected one-year nominal interest rate next year by i^e_{t+1}, and so on (the one-year nominal interest rate this year, i_t, is known today, so it does not need a superscript e). The expected present discounted value of this expected sequence of payments is given by

$$\$V_t = \$z_t + \frac{1}{(1 + i_t)}\$z^e_{t+1} + \frac{1}{(1 + i_t)(1 + i^e_{t+1})}\$z^e_{t+2} + \cdots \qquad (14.1)$$

"Expected present discounted value" is a heavy expression to carry; instead, we will often just use **present discounted value**, or even just **present value**. Also, it will be convenient to have a shorthand way of writing expressions like equation (14.1). To denote the present value of an expected sequence for $\$z$, we shall write $V(\$z_t)$, or just $V(\$z)$.

Using Present Values: Examples

Equation (14.1) has two important implications:

An increase in $\$z$ or an increase in future $\$z^e \Rightarrow$ an increase in V

- The present value depends positively on today's actual payment and expected future payments. An increase in either today's $\$z$ or any future $\$z^e$ leads to an increase in the present value.

An increase in i or an increase in future $i^e \Rightarrow$ a decrease in V

- The present value depends negatively on current and expected future interest rates. An increase in either current i or in any future i^e leads to a decrease in the present value.

Equation (14.1) is not simple, however, and so it will help to go through some examples.

Constant Interest Rates

To focus on the effects of the sequence of payments on the present value, assume that interest rates are expected to be constant over time, so that $i_t = i^e_{t+1} = \ldots$, and denote their common value by i. The present value formula—equation (14.1)—becomes:

$$\$V_t = \$z_t + \frac{1}{(1+i)}\$z^e_{t+1} + \frac{1}{(1+i)^2}\$z^e_{t+2} + \cdots \quad (14.2)$$

The weights correspond to the terms of a geometric series. See the discussion of geometric series in Appendix 2 at the end of the book.

In this case, the present value is a *weighted sum* of current and expected future payments, with weights that decline *geometrically* through time. The weight on a payment this year is 1, the weight on the payment n years from now is $(1/(1+i))^n$. With a positive interest rate, the weights get closer and closer to zero as we look further and further into the future. For example, with an interest rate equal to 10%, the weight on a payment 10 years from today is equal to $1/(1+0.10)^{10} = 0.386$, so that a payment of \$1,000 in 10 years is worth \$386 today. The weight on a payment in 30 years is $1/(1+0.10)^{30} = 0.057$, so that a payment of \$1,000 30 years from today is worth only \$57 today!

Constant Interest Rates and Payments

In some cases, the sequence of payments for which we want to compute the present value is simple. For example, a typical fixed-rate, 30-year mortgage requires constant dollar payments over 30 years. Consider a sequence of equal payments—call them $\$z$ without a time index—over n years, including this year. In this case, the present value formula in equation (14.2) simplifies to

$$\$V_t = \$z\left[1 + \frac{1}{(1+i)} + \cdots + \frac{1}{(1+i)^{n-1}}\right]$$

By now, geometric series should not hold any secret, and you should have no problem deriving this relation. But if you do, see Appendix 2 at the end of the book.

Because the terms in the expression in brackets represent a geometric series, we can compute the sum of the series and get

$$\$V_t = \$z\frac{1 - [1/(1+i)^n]}{1 - [1/(1+i)]}$$

Suppose you have just won \$1 million from your state lottery and have been presented with a 6-foot \$1,000,000 check on TV. Afterward, you are told that, to protect you from your worst spending instincts as well as from your many new "friends," the state will pay you the million dollars in equal yearly installments of \$50,000 over the next 20 years. What is the present value of your prize today? Taking, for example, an interest rate of 6% per year, the preceding equation gives

$V = \$50{,}000(0.688)/(0.057) =$ or about \$608,000. Not bad, but winning the prize did not make you a millionaire.

What is the present value if *i* equals 4%? 8%? (Answers: \$706,000, \$530,000)

Constant Interest Rates and Payments Forever

Let's go one step further and assume that payments are not only constant, but go on forever. Real-world examples are harder to come by for this case, but one example comes from 19th-century England, when the government issued *consols*, bonds that paid a fixed yearly amount forever. Let $\$z$ be the constant payment. Assume that payments start next year rather than right away as in the previous example (this makes for simpler algebra). From equation (14.2) we have

Some of these consols were still in circulation in 2015, when the British government decided to buy them back.

$$\$V_t = \frac{1}{(1+i)}\$z + \frac{1}{(1+i)^2}\$z + \cdots$$

$$= \frac{1}{(1+i)}\left[1 + \frac{1}{(1+i)} + \cdots\right]\$z$$

where the second line follows by factoring out $1/(1+i)$. The reason for factoring out $1/(1+i)$ should be clear from looking at the term in brackets. It is an infinite geometric sum, so we can use the property of geometric sums to rewrite the present value as

$$\$V_t = \frac{1}{1+i}\frac{1}{(1-(1/(1+i))}\$z$$

Or, simplifying (the steps are given in the application of Proposition 2 in Appendix 2 at the end of the book),

$$\$V_t = \frac{\$z}{i}$$

The present value of a constant sequence of payments $\$z$ is simply equal to the ratio of $\$z$ to the interest rate *i*. If, for example, the interest rate is expected to be 5% per year forever, the present value of a consol that promises \$10 per year forever equals $\$10/0.05 = \$200$. If the interest rate increases and is now expected to be 10% per year forever, the present value of the consol decreases to $\$10/0.10 = \$100$.

Zero Interest Rates

Because of discounting, computing present discounted values typically requires the use of a calculator. There is, however, a case where computations simplify. This is the case where the interest rate is equal to zero. If $i = 0$, then $1/(1+i)$ equals 1, and so does $(1/(1+i)^n)$ for any power *n*. For that reason, the present discounted value of a sequence of expected payments is just the *sum* of those expected payments. Because the interest rate is typically positive, assuming the interest rate is zero is only an approximation. But it can be a useful one for back-of-the-envelope computations.

Nominal versus Real Interest Rates and Present Values

So far, we have computed the present value of a sequence of dollar payments by using interest rates in terms of dollars—nominal interest rates. Specifically, we have written equation (14.1):

$$\$V_t = \$z_t + \frac{1}{(1+i_t)}\$z^e_{t+1} + \frac{1}{(1+i_t)(1+i^e_{t+1})}\$z^e_{t+2} + \cdots$$

where $i_t, i^e_{t+1}, \ldots$ is the sequence of current and expected future nominal interest rates and $\$z_t, \$z^e_{t+1}, \$z^e_{t+2}, \ldots$ is the sequence of current and expected future dollar payments.

Suppose we want to compute instead the present value of a sequence of *real* payments—that is, payments in terms of a basket of goods rather than in terms of dollars. Following the same logic as before, we need to use the right interest rates for this case, namely interest rates in terms of the basket of goods—*real interest rates*. Specifically, we can write the present value of a sequence of real payments as

$$V_t = z_t + \frac{1}{(1 + r_t)} z^e_{t+1} + \frac{1}{(1 + r_t)(1 + r^e_{t+1})} z^e_{t+2} + \cdots \qquad (14.3)$$

The proof is given in the appendix to this chapter. Although it may not be fun, go through it to test your understanding of the two concepts, real interest rate versus nominal interest rate and expected present value.

where $r_t, r^e_{t+1}, \ldots$ is the sequence of current and expected future real interest rates, $z_t, z^e_{t+1}, z^e_{t+2}, \ldots$ is the sequence of current and expected future real payments, and V_t is the real present value of future payments.

These two ways of writing the present value turn out to be equivalent. That is, the real value obtained by constructing $\$V_t$ using equation (14.1) and dividing by P_t, the price level, is equal to the real value V_t obtained from equation (14.3), so

$$\$V_t / P_t = V_t$$

In words: We can compute the present value of a sequence of payments in two ways. One way is to compute it as the present value of the sequence of payments expressed in dollars, discounted using nominal interest rates, and then divided by the price level today. The other way is to compute it as the present value of the sequence of payments expressed in real terms, discounted using real interest rates. The two ways give the same answer.

Do we need both formulas? Yes. Which one is more helpful depends on the context.

Take bonds, for example. Bonds typically are claims to a sequence of nominal payments over a period of years. For example, a 10-year bond might promise to pay $50 each year for 10 years, plus a final payment of $1,000 in the last year. So when we look at the pricing of bonds in the next section, we shall rely on equation (14.1) (which is expressed in terms of dollar payments) rather than on equation (14.3) (which is expressed in real terms).

But sometimes, we have a better sense of future expected real values than of future expected dollar values. You might not have a good idea of what your dollar income will be in 20 years. Its value depends very much on what happens to inflation between now and then. But you might be confident that your nominal income will increase by at least as much as inflation—in other words, that your real income will not decrease. In this case, using equation (14.1), which requires you to form expectations of future dollar income, will be difficult. However, using equation (14.3), which requires you to form expectations of future real income, may be easier. For this reason, when we discuss consumption and investment decisions in Chapter 15, we shall rely on equation (14.3) rather than equation (14.1).

14-2 BOND PRICES AND BOND YIELDS

Bonds differ in two basic dimensions:

- *Maturity:* The **maturity** of a bond is the length of time over which the bond promises to make payments to the holder of the bond. A bond that promises to make one payment of $1,000 in six months has a maturity of six months; a bond that promises to pay $100 per year for the next 20 years and a final payment of $1,000 at the end of those 20 years has a maturity of 20 years.

- *Risk:* This may be **default risk**, the risk that the issuer of the bond (it could be a government or a company) will not pay back the full amount promised by the bond. Or it may be **price risk**, the uncertainty about the price at which you can sell the bond in the future if you want to sell it before maturity.

Both maturity and risk matter in the determination of interest rates. As I want to focus here on the role of maturity and, by implication, the role of expectations, I shall ignore risk for now and reintroduce it later.

Some definitions first: Bonds of different maturities each have a price and an associated interest rate called the **yield to maturity**, or simply the **yield**. Yields on bonds with a short maturity, typically a year or less, are called **short-term interest rates**. Yields on bonds with a longer maturity are called **long-term interest rates**. On any given day, we observe the yields on bonds of different maturities, and so we can trace graphically how the yield depends on the maturity of a bond. This relation between maturity and yield is called the **yield curve**, or the **term structure of interest rates** (the word *term* is synonymous with maturity).

We introduced earlier two distinctions between different interest rates, real versus nominal interest rates, and policy rate versus borrowing rate (we are leaving this second distinction aside for the moment). We now introduce a third one, short versus long rates. Note that this makes for eight combinations.

Figure 14-2 gives, for example, the term structure of US government bonds on November 1, 2000, and the term structure of US government bonds on June 1, 2001. The choice of the two dates is not accidental; why I chose them will become clear later.

To find out what the yield curve for US bonds is at the time you read this chapter, go to yieldcurve.com and click on "yield curves." You will see the yield curves for both UK and US bonds.

Note that in Figure 14-2, on November 1, 2000, the yield curve was slightly downward sloping, declining from a three-month interest rate of 6.2% to a 30-year interest rate of 5.8%. In other words, long-term interest rates were slightly lower than short-term interest rates. Note that, seven months later, on June 1, 2001, the yield curve was sharply upward sloping, increasing from a three-month interest rate of 3.5% to a 30-year interest rate of 5.7%. In other words, long-term interest rates were much higher than short-term interest rates.

Why was the yield curve downward sloping in November 2000 but upward sloping in June 2001? Put another way, why were long-term interest rates slightly lower than short-term interest rates in November 2000, but substantially higher than short-term interest rates in June 2001? What were financial market participants thinking at each date? To answer these questions, and more generally to think about the determination of the yield curve and the relation between short- and long-term interest rates, we proceed in two steps:

1. First, we derive *bond prices* for bonds of different maturities.
2. Second, we go from bond prices to *bond yields* and examine the determinants of the yield curve and the relation between short- and long-term interest rates.

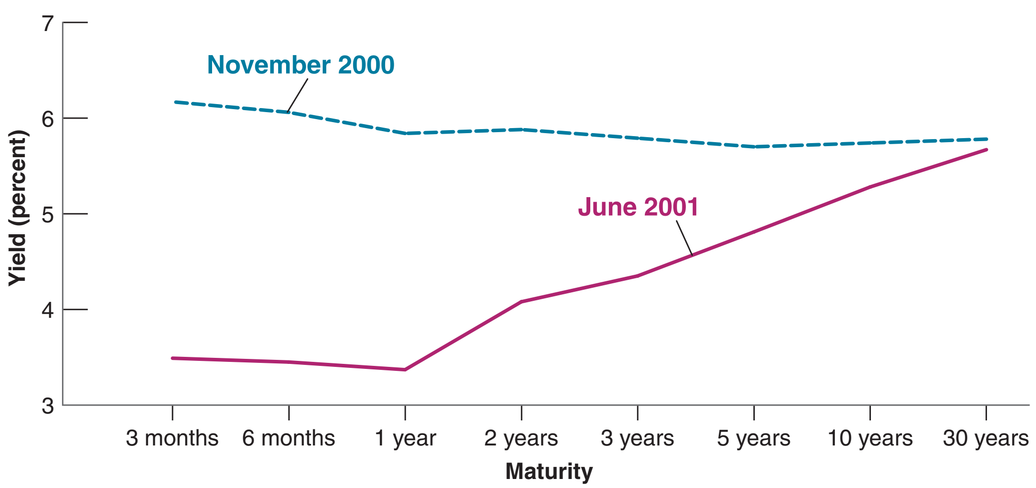


Figure 14-2

US Yield Curves: November 1, 2000 and June 1, 2001

The yield curve, which was slightly downward sloping on November 1, 2000, was sharply upward sloping seven months later.

Source: FRED. Series DGS1MO, DGS3MO, DGS6MO, DGS1, DGS2, DGS3, DGS5, DGS7, DGS10, DGS20, DGS30.

FOCUS

The Vocabulary of Bond Markets

Understanding the basic vocabulary of financial markets will help make them (a bit) less mysterious. Here is a basic vocabulary review.

- Bonds are issued by governments or by firms. If issued by the government or government agencies, the bonds are called **government bonds**. If issued by firms (corporations), they are called **corporate bonds**.
- Bonds are rated for their default risk (the risk that they will not be repaid) by rating agencies. The two major rating agencies are the Standard and Poor's Corporation (S&P) and Moody's Investors Service. Standard and Poor's **bond ratings** range from AAA to D. In August 2011, Standard and Poor's downgraded US government bonds from AAA to AA + (the next rating down), reflecting its worry about the large budget deficits. This downgrade created a strong controversy. A lower rating typically implies that the bond must pay a higher interest rate or investors will not buy it. The difference between the interest rate paid on a given bond and the interest rate paid on the bond with the highest (best) rating is called the **risk premium** associated with the given bond. Bonds with high default risk are sometimes called **junk bonds**.
- Bonds that promise a single payment at maturity, rather than annual payments along the way, are called **discount bonds**. The single payment is called the **face value** of the bond.
- Bonds that promise multiple payments before maturity and one payment at maturity are called **coupon bonds**. The payments before maturity are called **coupon payments.** The final payment is called the face value of the bond. The ratio of coupon payments to the face value is called the **coupon rate.** The **current yield** is the ratio of the coupon payment to the price of the bond.

  For example, a bond with coupon payments of $5 each year, a face value of $100, and a price of $80 has a coupon rate of 5% and a current yield of $5/80 = 0.0625 = 6.25\%$. From an economic viewpoint, neither the coupon rate nor the current yield are interesting measures. The correct measure of the interest rate on a bond is its yield to maturity, or simply yield; you can think of it as roughly the average interest rate paid by the bond over its **life** (the life of a bond is the amount of time left until the bond matures). We shall define the *yield to maturity* more precisely later in this section.
- US government bonds range in maturity from a few days to 30 years. Bonds with a maturity of up to a year when they are issued are called **Treasury bills (T-bills).** They are discount bonds, making only one payment at maturity. Bonds with a maturity of 1 to 10 years when they are issued are called **Treasury notes.** Bonds with a maturity of 10 or more years when they are issued are called **Treasury bonds**. Both Treasury notes and Treasury bonds are coupon bonds. Bonds with longer maturities are riskier, and thus typically carry a risk premium, also called the **term premium.**
- Bonds are typically nominal bonds. They promise a sequence of fixed nominal payments—payments in terms of domestic currency. There are, however, other types of bonds. Among them are **indexed bonds**, bonds that promise payments adjusted for inflation rather than fixed nominal payments. Instead of promising to pay, say, $100 in a year, a one-year indexed bond promises to pay $100\ (1 + \pi)$ dollars, whatever π, the rate of inflation that will take place over the coming year, turns out to be. Because they protect bondholders against the risk of inflation, indexed bonds are popular in many countries. They play a particularly important role in the United Kingdom, where, over the last 30 years, people have increasingly used them to save for retirement. By holding long-term indexed bonds, people can make sure that the payments they receive when they retire will be protected from inflation. Indexed bonds, called **Treasury Inflation Protected Securities (TIPS)**, were introduced in the United States in 1997.

Bond Prices as Present Values

In much of this section, we shall look at just two types of bonds: a bond that promises one payment of $100 in one year—a one-year bond—and a bond that promises one payment of $100 in two years—a two-year bond. Once you understand how their prices and yields are determined, it will be easy to generalize the results to bonds of any maturity. I shall do so later.

Note that both bonds are *discount bonds* (see the Focus Box "The Vocabulary of Bond Markets").

Let's start by deriving the prices of the two bonds.

- Given that the one-year bond promises to pay $100 next year, it follows, from the previous section, that its price, call it $\$P_{1t}$, must be equal to the present value of a payment of $100 next year. Let the current one-year nominal interest rate be i_{1t}. Note that we now denote the one-year interest rate in year t by i_{1t} rather than simply by i_t as we did in previous chapters. This is to make it easier for you to remember that it is the *one-year* interest rate. So,

$$\$P_{1t} = \frac{\$100}{1 + i_{1t}} \tag{14.4}$$

The price of the one-year bond varies inversely with the current one-year nominal interest rate.

We saw this relation in Chapter 4, Section 4–2.

- Given that the two-year bond promises to pay \$100 in two years, its price, call it $\$P_{2t}$, must be equal to the present value of \$100 two years from now:

$$\$P_{2t} = \frac{\$100}{(1 + i_{1t})(1 + i^e_{1t+1})} \tag{14.5}$$

where i_{1t} denotes the one-year interest rate this year and i^e_{1t+1} denotes the one-year rate expected by financial markets for next year. The price of the two-year bond depends inversely on both the current one-year rate and the one-year rate expected for next year.

Arbitrage and Bond Prices

Before further exploring the implications of equations (14.4) and (14.5), let us look at an alternative derivation of equation (14.5). This alternative derivation will introduce you to the important concept of **arbitrage**.

Suppose you have the choice between holding one-year bonds or two-year bonds and what you care about is how much you will have *one year from today*. Which bonds should you hold?

- Suppose you hold one-year bonds. For every dollar you put in one-year bonds, you will get $(1 + i_{1t})$ dollars next year. This relation is represented in the first line of Figure 14-3.
- Suppose you hold two-year bonds. Because the price of a two-year bond is $\$P_{2t}$, every dollar you put in two-year bonds buys you $\$1/\$P_{2t}$ bonds today. When next year comes, the bond will have one more year before maturity. Thus, one year from today, the two-year bond will now be a one-year bond. Therefore, the price at which you can expect to sell it next year is $\$P^e_{1t+1}$, the expected price of a one-year bond next year. So for every dollar you put in two-year bonds, you can expect to receive $\$1/\$P_{2t}$ multiplied by $\$P^e_{1t+1}$, or, equivalently, $\$P^e_{1t+1}/\$P_{2t}$ dollars next year. This is represented in the second line of Figure 14-3.

Which bonds should you hold? Suppose you and other financial investors care *only* about the expected return and do not care about risk. This assumption is known as the **expectations hypothesis**. It is a simplification. You and other investors are likely to care not only about the expected return but also about the risk associated with holding each bond. If you hold a one-year bond, you know with certainty what you will get next year. If you hold a two-year bond, the price at which you will sell it next year is uncertain; holding the two-year bond for one year is risky. As I indicated previously, I am disregarding this for now but shall come back to it later.

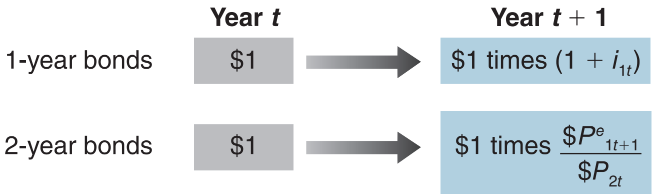


Figure 14-3

Returns from Holding One-Year and Two-Year Bonds for One Year

Under the assumption that you and other financial investors care only about expected return, it follows that the two bonds must offer the same expected one-year return. Suppose this condition was not satisfied. Suppose that, for example, the one-year return on one-year bonds was lower than the expected one-year return on two-year bonds. In this case, no one would want to hold the existing supply of one-year bonds, and the market for one-year bonds could not be in equilibrium. Only if the expected one-year return is the same on both bonds will you and other financial investors be willing to hold both one-year bonds and two-year bonds.

If the two bonds offer the same expected one-year return, it follows from Figure 14-3 that

$$1 + i_{1t} = \frac{\$P^e_{1t+1}}{\$P_{2t}} \tag{14.6}$$

I use *arbitrage* to denote the proposition that *expected* returns on two assets must be equal. Some finance economists reserve *arbitrage* for the narrower proposition that *riskless* profit opportunities do not go unexploited. ▶

The left side of the equation gives the return per dollar from holding a one-year bond for one year; the right side gives the expected return per dollar from holding a two-year bond for one year. We shall call equations such as (14.6)—equations that state that the expected returns on two assets must be equal—*arbitrage* relations. Rewrite equation (14.6) as

$$\$P_{2t} = \frac{\$P^e_{1t+1}}{1 + i_{1t}} \tag{14.7}$$

Arbitrage implies that the price of a two-year bond today is the present value of the expected price of the bond next year. This naturally raises the next question: What does the expected price of one-year bonds next year ($\$P^e_{1t+1}$) depend on?

The answer is straightforward. Just as the price of a one-year bond this year depends on this year's one-year interest rate, the price of a one-year bond next year will depend on the one-year interest rate next year. Writing equation (14.4) for next year (year $t + 1$) and denoting expectations in the usual way, we get

$$\$P^e_{1t+1} = \frac{\$100}{(1 + i^e_{1t+1})} \tag{14.8}$$

The price of the bond next year is expected to equal the final payment, \$100, discounted by the one-year interest rate expected for next year.

Replacing $\$P^e_{1t+1}$ from equation (14.8) in equation (14.7) gives

$$\$P_{2t} = \frac{\$100}{(1 + i_{1t})(1 + i^e_{1t+1})} \tag{14.9}$$

The relation between arbitrage and present values: Arbitrage between bonds of different maturities implies that bond prices are equal to the expected present values of payments on these bonds. ▶

This expression is the same as equation (14.5). This is the point. What we have shown is that *arbitrage* between one- and two-year bonds implies that the price of two-year bonds is the *present value* of the payment in two years, namely \$100, discounted using current and next year's expected one-year interest rates.

From Bond Prices to Bond Yields

Having looked at bond prices, we now go on to bond yields. The basic point: Bond yields contain the same information about future expected interest rates as bond prices. They just do so in a much clearer way.

To begin, we need a definition of the yield to maturity. The *yield to maturity* on an n-year bond, or equivalently, the ***n*-year interest rate**, is defined as the constant annual interest rate that makes the bond price today equal to the present value of future payments on the bond.

This definition is simpler than it sounds. Take, for example, the two-year bond we introduced previously. Denote its yield by i_{2t}, where the subscript 2 is there to remind us that this is the *yield to maturity* on a two-year bond, or, equivalently, the two-year interest rate. Following the definition of the yield to maturity, this yield is the constant annual interest rate that would make the present value of \$100 in two years equal to the price of the bond today. So, it satisfies the following relation:

$$\$P_{2t} = \frac{\$100}{(1 + i_{2t})^2} \tag{14.10}$$

Suppose the bond sells for \$90 today. Then the two-year interest rate i_{2t} is given by $\sqrt{100/90} - 1$, or 5.4%. In other words, holding the bond for two years—until maturity—yields an interest rate of 5.4% per year.

$\$90 = \$100/(1 + i_{2t})^2 \Rightarrow$
$(1 + i_{2t})^2 = \$100/\$90 \Rightarrow$
$(1 + i_{2t})\sqrt{\$100/\$90} \Rightarrow$
$i_{2t} = 5.4\%$

What is the relation of the two-year interest rate to the current one-year interest rate and the expected one-year interest rate? To answer that question, look at equation (14.10) and equation (14.9). Eliminating $\$P_{2t}$ between the two gives

$$\frac{\$100}{(1 + i_{2t})^2} = \frac{\$100}{(1 + i_{1t})(1 + i^e_{1t+1})}$$

Rearranging,

$$(1 + i_{2t})^2 = (1 + i_{1t})(1 + i^e_{1t+1})$$

This gives us the relation between the two-year interest rate i_{2t}, the current one-year interest rate i_{1t}, and next year's expected one-year interest rate i^e_{t+1}. A useful approximation to this relation is given by

We used a similar approximation when we looked at the relation between the nominal interest rate and the real interest rate in Chapter 6. See Proposition 3 in Appendix 2 at the end of the book.

$$i_{2t} \approx \frac{1}{2}\,(i_{1t} + i^e_{1t+1}) \tag{14.11}$$

Equation (14.11) simply says that *the two-year interest rate is (approximately) the average of the current one-year interest rate and next year's expected one-year interest rate.*

We have focused on the relation between the prices and yields of one- and two-year bonds. But our results generalize to bonds of any maturity. For instance, we could have looked at bonds with maturities of less than a year. To take an example, the yield on a bond with a maturity of six months is (approximately) equal to the average of the current three-month interest rate and next quarter's expected three-month interest rate. Or we could have looked instead at bonds with maturities longer than two years. For example, the yield on a 10-year bond is (approximately) equal to the average of the current one-year interest rate and the one-year interest rates expected for the next nine years.

The general principle is clear: Long-term interest rates reflect current and future expected short-term interest rates. Before we return to an interpretation of the yield curves in Figure 14-2, we need to reintroduce risk.

Reintroducing Risk

We have assumed so far that investors did not care about risk. But they do care. Go back to the choice between holding a one-year bond for one year or holding a two-year bond for one year. The first option is riskless. The second is risky as you do not know the price at which you will sell the bond in a year. You are thus likely to ask for a risk premium to hold the two-year bond, and the arbitrage equation takes the form:

$$1 + i_{1t} + x = \frac{\$P^e_{1t+1}}{\$P_{2t}}$$

The expected return on the two-year bond (the right-hand side) must exceed the return on the one-year bond by some risk premium x. Reorganizing gives:

$$\$P_{2t} = \frac{\$P^e_{1t+1}}{1 + i_{1t} + x}$$

The price of the two-year bond is the discounted value of the expected price of a one-year bond next year, with the discount rate now reflecting the risk premium. As one-year bonds have a known return and are therefore not risky, the expected price of a one-year bond next year is still given by equation (14.8). So replacing in the previous equation gives:

$$\$P_{2t} = \frac{\$100}{(1 + i_{1t})(1 + i^e_{1t+1} + x)} \quad (14.12)$$

Now, to go from prices to yields, let's go through the same steps as before. Using the two expressions for the price of the two-year bond, equation (14.10) and equation (14.12), gives:

$$\frac{\$100}{(1 + i_{2t})^2} = \frac{\$100}{(1 + i_{1t})(1 + i^e_{1t+1} + x)}$$

Manipulating the equation gives:

$$(1 + i_{2t})^2 = (1 + i_{1t})(1 + i^e_{1t+1} + x)$$

Finally, using the same approximation as before gives:

$$i_{2t} \approx \frac{1}{2}(i_{1t} + i^e_{1t+1} + x) \quad (14.13)$$

Recently, the term premium has decreased, due to the use of quantitative easing by the Fed (more on this in Chapter 23) ▶

The two-year rate is the average of the current and expected one-year rate plus a risk premium. Take the case where the one-year rate is expected to be the same next year as this year. Then the two-year rate will exceed the one-year rate by a term reflecting the risk in holding two-year bonds. As the price risk increases with the maturity of the bonds, the risk premium typically increases with maturity, typically reaching 1–2% for long-term bonds. This implies that, on average, the yield curve is slightly upward sloping, reflecting the higher risk involved in holding longer maturity bonds.

Interpreting the Yield Curve

We now have what we need to interpret Figure 14-2.

Consider the yield curve for November 1, 2000. Recall that when investors expect interest rates to be constant over time, the yield curve should be slightly upward sloping, reflecting the fact that the risk premium increases with maturity. The fact that the yield curve was downward sloping, something relatively rare, tells us that investors expected interest rates to go down slightly over time, with the expected decrease in rates more than compensating for a rising term premium. And if we look at the macroeconomic situation at the time, they had good reasons to hold this view. At the end of November 2000, the US economy was slowing down. Investors expected what they called a *smooth landing*. They thought that to maintain growth, the Fed would slowly decrease the policy rate, and these expectations were what lay behind the downward-sloping yield curve. By June 2001, however, growth had declined much more than was expected in November 2000, and by then the Fed had decreased the interest rate much more than investors had expected. Investors now expected that, as the economy recovered, the Fed would start increasing the policy rate. So the yield curve sloped upward. Note, however, that the yield curve was nearly flat for maturities up to one year. This tells us that financial markets did not expect interest rates to start rising until a year hence—that is, before June 2002. Did they turn out to be right? Not quite. In fact, the recovery was much weaker than expected, and the Fed did not increase the policy rate until June 2004—fully two years later than financial markets had anticipated.

Let's summarize what you have learned in this section. We have focused on bonds. You have seen how arbitrage determines the price of bonds. You have seen how bond prices and bond yields depend on current and future expected interest rates and risk premiums and what can be learned by looking at the yield curve.

You may want to read again the Focus Box on the 2001 recession in Chapter 5.

14-3 THE STOCK MARKET AND MOVEMENTS IN STOCK PRICES

While governments finance themselves by issuing bonds, the same is not true of firms. Firms finance themselves in four ways. First, they rely on **internal finance**—that is, they use some of their own profits; second—and this is the main channel of **external finance** for small firms—through bank loans (as we saw in Chapter 6, this channel played a central role in the crisis); third, through **debt finance**—bonds and loans; and fourth, through **equity finance**, issuing **stocks—or shares**, as stocks are also called. Instead of paying predetermined amounts as bonds do, stocks pay **dividends** in an amount decided by the firm. Dividends are paid from the firm's profits. Typically, dividends are less than profits because firms retain some of their profits to finance their investment. But dividends move with profits: When profits increase, so do dividends.

Our focus in this section is on the determination of stock prices. As a way of introducing the issues, let's look at the behavior of an index of US stock prices, the *Standard & Poor's 500 Composite Index* (the S&P index) since 1980. Movements in the S&P index measure movements in the average stock price of 500 large companies.

Another and better-known index is the *Dow Jones Industrial Average*, an index of stocks of primarily industrial firms and therefore less representative of the average price of stocks than is the S&P index. Similar indexes exist for other countries. The *Nikkei index* reflects movements of stock prices in Tokyo, and the *FTSE* and *CAC 40* indexes reflect stock price movements in London and Paris, respectively.

Figure 14-4 plots the real stock price index constructed by dividing the S&P index by the consumer price index (CPI) for each month and normalizing so the index is equal to 1 in 1970. The striking feature of the figure is obviously the sharp movements in the value of the index. Note that the index went up from 1.4 in 1995 to 3.5 in 2000, only to decline sharply to 2.1 in 2003. And in the Great Financial Crisis, the index declined from 3.4 in 2007 to 1.7 in 2009, only to recover since then, standing at 4.3 at the end

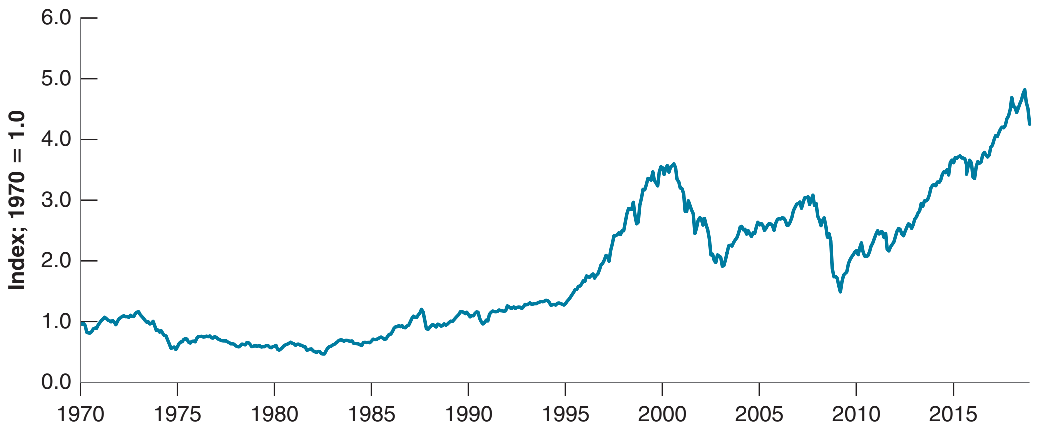


Figure 14-4

Standard and Poor's Stock Price Index in Real Terms since 1970

Note the sharp fluctuations in stock prices since the mid-1990s.

Source: FRED SP500, CPIAUCSL.

of 2018. What determines these sharp movements in stock prices? How do stock prices respond to changes in the economic environment and macroeconomic policy? These are the questions we take up in this section.

Stock Prices as Present Values

What determines the price of a stock that promises a sequence of dividends in the future? By now, I am sure the material in Section 14-1 has become second nature, and you already know the answer. The stock price must be equal to the present value of future expected dividends.

Just as we did for bonds, let's derive this result by looking at the implications of arbitrage between one-year bonds and stocks. Suppose you face the choice of investing either in one-year bonds or in stocks for a year. What should you choose?

- Suppose you decide to hold one-year bonds. Then for every dollar you put in one-year bonds, you will get $(1 + i_{1t})$ dollars next year. This payoff is represented in the upper line of Figure 14-5.
- Suppose you decide instead to hold stocks for a year. Let $\$Q_t$ be the price of the stock. Let $\$D_t$ denote the dividend this year, and $\$D^e_{t+1}$ the expected dividend next year. Suppose we look at the price of the stock after the dividend has been paid this year; this price is known as the **ex-dividend price**—so that the first dividend to be paid after the purchase of the stock is next year's dividend. (This is just a matter of convention; we could alternatively look at the price before this year's dividend has been paid. What term would we have to add?)

Holding the stock for a year implies buying a stock today, receiving a dividend next year, and then selling the stock. As the price of a stock is $\$Q_t$, every dollar you put in stocks buys you $\$1/\$Q_t$ stocks. And for each stock you buy, you expect to receive $(\$D^e_{t+1} + \$Q^e_{t+1})$, the sum of the expected dividend and the stock price next year. Therefore, for every dollar you put in stocks, you expect to receive $(\$D^e_{t+1} + \$Q^e_{t+1})/\$Q_t$. This payoff is represented in the lower line of Figure 14-5.

Let's use the same arbitrage argument we used for bonds. It is clear that holding a stock for one year is risky, much riskier than holding a one-year bond for a year (which is riskless). Rather than proceeding in two steps as we did for bonds (first leaving out risk considerations and then introducing a risk premium), let's take risk into account from the start and assume that financial investors require a risk premium to hold stocks.

In the case of stocks, the risk premium is called the **equity premium**. Equilibrium then requires that the expected rate of return from holding stocks for one year be the same as the rate of return on one-year bonds plus the equity premium:

$$\frac{\$D^e_{t+1} + \$Q^e_{t+1}}{\$Q_t} = 1 + i_{1t} + x$$

where x denotes the equity premium. Rewrite this equation as

$$\$Q_t = \frac{\$D^e_{t+1}}{(1 + i_{1t} + x)} + \frac{\$Q^e_{t+1}}{(1 + i_{1t} + x)} \qquad (14.14)$$

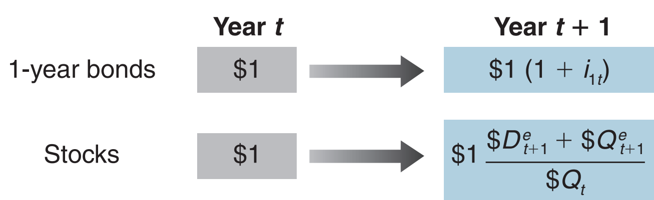


Figure 14-5

Returns from Holding One-Year Bonds or Stocks for One Year

Arbitrage implies that the price of the stock today must be equal to the present value of the expected dividend plus the present value of the expected stock price next year.

The next step is to think about what determines $\$Q^e_{t+1}$, the expected stock price next year. Next year, financial investors will again face the choice between stocks and one-year bonds. Thus, the same arbitrage relation will hold. Writing the previous equation, but now for time $t + 1$ and taking expectations into account, gives

$$\$Q^e_{t+1} = \frac{\$D^e_{t+2}}{(1 + i^e_{1t+1} + x)} + \frac{\$Q^e_{t+2}}{(1 + i^e_{1t+1} + x)}$$

The expected price next year is simply the present value next year of the sum of the expected dividend and price two years from now. Replacing the expected price $\$Q^e_{t+1}$ in equation (14.14) gives

$$\$Q_t = \frac{\$D^e_{t+1}}{(1 + i_{1t} + x)} + \frac{\$D^e_{t+2}}{(1 + i_{1t} + x)(1 + i^e_{1t+1} + x)} + \frac{\$Q^e_{t+2}}{(1 + i_{1t} + x)(1 + i^e_{1t+1} + x)}$$

The stock price is the present value of the expected dividend next year, plus the present value of the expected dividend two years from now, plus the expected price two years from now.

If we replace the expected price in two years with the present value of the expected price and dividends in three years, and so on for n years, we get

$$\$Q_t = \frac{\$D^e_{t+1}}{(1 + i_{1t} + x)} + \frac{\$D^e_{t+2}}{(1 + i_{1t} + x)(1 + i^e_{1t+1} + x)} + \cdots + \frac{\$D^e_{t+n}}{(1 + i_{1t} + x)\cdots(1 + i^e_{1t+n-1} + x)} + \frac{\$Q^e_{t+n}}{(1 + i_{1t} + x)\cdots(1 + i^e_{1t+n-1} + x)} \quad (14.15)$$

Look at the last term in equation (14.15), the present value of the expected price in n years. As long as people do not expect the stock price to explode in the future, then as we keep replacing Q^e_{t+n} and n increases, this term will go to zero. To see why, suppose the interest rate is constant and equal to i. The last term becomes

$$\frac{\$Q^e_{t+n}}{(1 + i_{1t} + x)\cdots(1 + i^e_{1t+n-1} + x)} = \frac{\$Q^e_{t+n}}{(1 + i + x)^n}$$

Suppose further that people expect the price of the stock to converge to some value, call it $\$\bar{Q}$, in the far future. Then, the last term becomes

$$\frac{\$Q^e_{t+n}}{(1 + i + x)^n} = \frac{\$\bar{Q}}{(1 + i + x)^n}$$

A subtle point: The condition that people expect the price of the stock to converge to some value over time seems reasonable. Indeed, most of the time it is likely to be satisfied. When, however, prices are subject to rational bubbles (Section 14–4), people expect large increases in the stock price in the future and this is when the condition that the expected stock price does not explode is not satisfied. This is why, when there are bubbles, this argument fails, and the stock price is no longer equal to the present value of expected dividends.

If the discount rate is positive, this expression goes to zero as n becomes large. Equation (14.15) reduces to

$$\$Q_t = \frac{\$D^e_{t+1}}{(1 + i_{1t} + x)} + \frac{\$D^e_{t+2}}{(1 + i_{1t} + x)(1 + i^e_{1t+1} + x)} + \cdots + \frac{\$D^e_{t+n}}{(1 + i_{1t} + x)\cdots(1 + i^e_{1t+n-1} + x)} + \ldots \quad (14.16)$$

Two equivalent ways of writing the stock price: The *nominal* stock price equals the expected present discounted value of future nominal dividends, discounted by current and future nominal interest rates.

The *real* stock price equals the expected present discounted value of future real dividends, discounted by current and future real interest rates.

The price of the stock is equal to the present value of the dividend next year, discounted using the current one-year interest rate plus the equity premium, plus the present value of the dividend two years from now, discounted using both this year's one-year interest rate and the next year's expected one-year interest rate, plus the equity premium, and so on.

Equation (14.16) gives the stock price as the present value of *nominal* dividends, discounted by *nominal* interest rates. From Section 14-1, we know we can rewrite this equation to express the *real* stock price as the present value of *real* dividends, discounted by *real* interest rates. So we can rewrite the real stock price as:

$$Q_t = \frac{D^e_{t+1}}{(1 + r_{1t} + x)} + \frac{D^e_{t+2}}{(1 + r_{1t} + x)(1 + r^e_{1t+1} + x)} + \cdots \qquad (14.17)$$

Q_t and D_t, without a dollar sign, denote the real price and real dividends at time t. The real stock price is the present value of future real dividends, discounted by the sequence of one-year real interest rates plus the equity premium.

This relation has three important implications:

- Higher expected future real dividends lead to a higher real stock price.
- Higher current and expected future one-year real interest rates lead to a lower real stock price.
- A higher equity premium leads to a lower stock price.

Let's now see what light this relation sheds on movements in the stock market.

The Stock Market and Economic Activity

Figure 14-4 showed the large movements in stock prices over the last two decades. It is not unusual for the index to go up or down by 15% within a year. In 1997, the stock market went up by 24% (in real terms); in 2008, it went down by 46%. Daily movements of 2% or more are not unusual. What causes these movements?

The first point to be made is that these movements should be, and are for the most part, unpredictable. The reason why is best understood by thinking in terms of the choice people have between stocks and bonds. If it were widely believed that, a year from now, the price of a stock was going to be 20% higher than today's price, holding the stock for a year would be unusually attractive, much more attractive than holding short-term bonds. There would be a very large demand for the stock. Its price would increase *today* to the point where the expected return from holding the stock was back in line with the expected return on other assets. In other words, the expectation of a high stock price next year would lead to a high stock price today.

You may have heard that stock prices follow a **random walk**. This is a technical term, but with a simple interpretation. Something—it can be a molecule, or the price of an asset—follows a random walk if each step it takes is as likely to be up as it is to be down. Its movements are therefore unpredictable.

There is a saying in economics that it is a sign of a *well-functioning stock market* that movements in stock prices are unpredictable. The saying is not quite right. At any moment, a few financial investors will have better information or simply be better at reading the future. If they are only a few, they may not buy enough of the stock to bid its price all the way up today. Thus, they may get large expected returns. But the basic idea is nevertheless correct. Financial market gurus who regularly predict large imminent movements in the stock market are quacks. Major movements in stock prices cannot be predicted.

If movements in the stock market cannot be predicted, if they are the result of news, where does this leave us? We can still do two things:

- We can do Monday-morning quarterbacking, looking back and identifying the news to which the market reacted.

- We can ask "what if" questions. For example: What would happen to the stock market if the Fed were to embark on a more expansionary policy, or if consumers were to become more optimistic and increase spending?

Let us look at two "what if" questions using the IS-LM model we developed earlier (we shall extend it in the next chapter to take explicit account of expectations; for the moment the old model will do). To simplify, let's assume, as we did earlier, that expected inflation equals zero, so that the real interest rate and the nominal interest rate are equal.

A Monetary Expansion and the Stock Market

Suppose the economy is in a recession and the Fed decides to decrease the policy rate. The LM curve shifts down to LM' in Figure 14-6, and equilibrium output moves from point A to point A'. How will the stock market react?

This assumes that the policy rate is positive to start with, so the economy is not in a liquidity trap.

The answer depends on what participants in the stock market expected monetary policy to be before the Fed's move:

Stock prices may go up. If the Fed's move is at least partly unexpected, stock prices are likely to increase, for two reasons: First, a more expansionary monetary policy implies lower interest rates for some time. Second, it also implies higher output for some time (until the economy returns to the natural level of output), and therefore higher dividends. As equation (14.17) tells us, both lower interest rates and higher dividends—current and expected—will lead to an increase in stock prices.

Stock prices may not change. If investors fully anticipated the expansionary policy, then the stock market will not react. Neither its expectations of future dividends nor its expectations of future interest rates are affected by a move it had already anticipated. Thus, in equation (14.17), nothing changes, and stock prices remain the same.

Stock prices may go down. If stock market participants believe that the Fed is acting because it knows something they don't, namely that the economy is much worse than they thought, they might conclude that, on net, lower interest rates will not be enough to offset the bad news. They might then lower their forecasts of output and of dividends, leading to a decrease in stock prices.

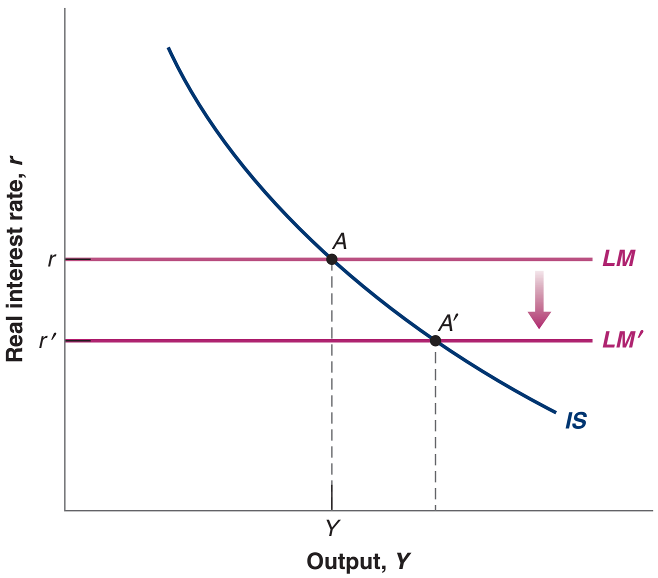


Figure 14-6

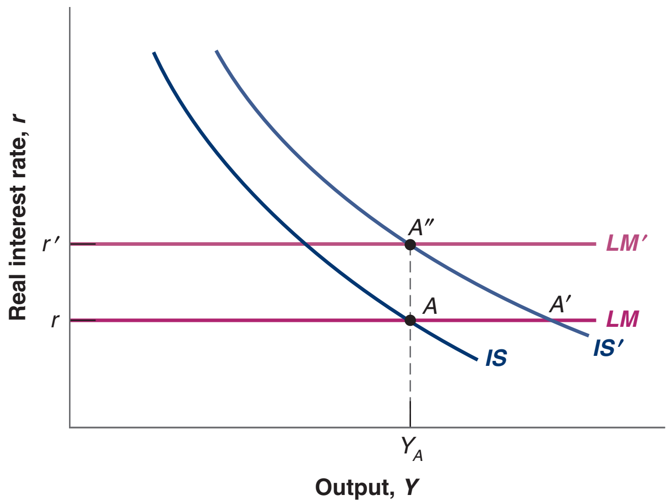
An Expansionary Monetary Policy and the Stock Market

A monetary expansion decreases the interest rate and increases output. What it does to the stock market depends on whether or not financial markets anticipated the monetary expansion and on the motives of the central bank.

Figure 14-7

An Increase in Consumption Spending and the Stock Market

An increase in consumption leads to a higher level of output. What happens to the stock market depends on the reaction of the Fed.



An Increase in Consumer Spending and the Stock Market

Now consider an unexpected shift of the IS curve to the right, resulting, for example, from stronger-than-expected consumer spending. As a result of the shift, output in Figure 14-7 increases from A to A'.

Will stock prices go up? You might be tempted to say yes. A stronger economy means higher profits and higher dividends for some time. But this answer is not necessarily right.

The reason is that it ignores the response of the Fed. If the market expects that the Fed will not respond and will keep the real policy rate unchanged at r, output will increase a lot, as the economy moves to A'. With unchanged interest rates and higher output, stock prices go up.

The Fed's behavior however is what financial investors often care about the most. After receiving the news of unexpectedly strong economic activity, the main question on Wall Street is: How will the Fed react?

What will happen if the market expects that the Fed might worry that an increase in output above Y_A may lead to an increase in inflation? This will be the case if Y_A was already close to the natural level of output. In this case, a further increase in output would lead to an increase in inflation, something that the Fed wants to avoid. A decision by the Fed to counteract the rightward shift of the IS curve with an increase in the policy rate causes the LM curve to shift up, from LM to LM', so the economy goes from A to A'' and output does not change. In that case, stock prices will surely go down: There is no change in expected profits, but the interest rate is now higher.

Let's summarize: Stock prices depend on current and future movements in activity. But this does not imply any simple relation between stock prices and output. How stock prices respond to a change in output depends on (1) what the market expected in the first place, (2) the source of the shocks behind the change in output, and (3) how the market expects the central bank to react to the output change. Test your newly acquired understanding by reading the Focus Box "Making (Some) Sense of (Apparent) Nonsense: Why the Stock Market Moved Yesterday, and Other Stories." Good luck!

Making (Some) Sense of (Apparent) Nonsense: Why the Stock Market Moved Yesterday, and Other Stories

FOCUS

Here are some quotes from *The Wall Street Journal* from April 1997 to August 2001. Try to make sense of them, using what you've just learned. (And if you have time, find your own quotes.)

- April 1997. Good news on the economy, leading to an increase in stock prices:

  "Bullish investors celebrated the release of market-friendly economic data by stampeding back into stock and bond markets, pushing the Dow Jones Industrial Average to its second-largest point gain ever and putting the blue-chip index within shooting distance of a record just weeks after it was reeling."

- December 1999. Good news on the economy, leading to a decrease in stock prices:

  "Good economic news was bad news for stocks and worse news for bonds ... The announcement of stronger-than-expected November retail-sales numbers wasn't welcome. Economic strength creates inflation fears and sharpens the risk that the Federal Reserve will raise interest rates again."

- September 1998. Bad news on the economy, leading to a decrease in stock prices:

  "Nasdaq stocks plummeted as worries about the strength of the US economy and the profitability of US corporations prompted widespread selling."

- August 2001. Bad news on the economy, leading to an increase in stock prices:

  "Investors shrugged off more gloomy economic news, and focused instead on their hope that the worst is now over for both the economy and the stock market. The optimism translated into another 2% gain for the Nasdaq Composite Index."

Dan Wasserman Editorial cartoon/Boston Globe/TNS

14-4 RISK, BUBBLES, FADS, AND ASSET PRICES

Do all movements in stock and other asset prices come from news about future dividends or interest rates? The answer is no, for two reasons: The first is that there is variation over time in perceptions of risk. The second is deviations of prices from their fundamental value, namely bubbles or fads. Let's look at each one in turn.

Stock Prices and Risk

In the previous section, I assumed that the equity premium x was constant. It is not. After the Great Depression, the equity premium was very high, perhaps reflecting the fact that investors, remembering the collapse of the stock market in 1929, were reluctant to hold stocks unless the premium was high enough. It started to decrease in the early 1950s, from around 7% to less than 4% today. And it can also change quickly. Part of the large stock market fall in 2008 was due not only to more pessimistic expectations of future dividends, but also to the large increase in uncertainty and the perception of higher risk by stock market participants. Thus, much of the movement in stock prices comes not just from expectations of future dividends and interest rates but also from shifts in the equity premium.

Asset Prices, Fundamentals, and Bubbles

In the previous section, we assumed that stock prices were always equal to their *fundamental value*, defined as the present value of expected dividends given in equation (14.17). Do stock prices always correspond to their fundamental value? Most economists doubt it. They point to Black October in 1929, when the US stock market fell

FOCUS

Famous Bubbles: From Tulipmania in 17th-Century Holland to Russia in 1994

Tulipmania in Holland

In the 17th century, tulips became increasingly popular in Western European gardens. A market developed in Holland for both rare and common forms of tulip bulbs.

An episode called the "tulip bubble" took place from 1634 to 1637. In 1634, the price of rare bulbs started increasing. The market went into a frenzy, with speculators buying tulip bulbs in anticipation of even higher prices. For example, the price of a bulb called "Admiral Van de Eyck" increased from 1,500 guineas in 1634 to 7,500 guineas in 1637, equal to the price of a house at the time. There are stories about a sailor mistakenly eating bulbs, only to realize the cost of his "meal" later. In early 1637, prices increased faster. Even the price of some common bulbs exploded, rising by a factor of up to 20 in January. But in February 1637, prices collapsed. A few years later, bulbs were trading for roughly 10% of their value at the peak of the bubble.

This account is taken from Peter Garber, "Tulipmania," *Journal of Political Economy*, 1989, 97(3): pp. 535–560.

The MMM Pyramid in Russia

In 1994 a Russian "financier," Sergei Mavrodi, created a company called MMM and proceeded to sell shares, promising shareholders a rate of return of at least 3,000% per year!

The company was an instant success. The price of MMM shares increased from 1,600 rubles (then worth $1) in February to 105,000 rubles (then worth $51) in July. And by July, according to company claims, the number of shareholders had grown to 10 million.

The trouble was that the company was not involved in any type of production and held no assets, except for its 140 offices in Russia. The shares were intrinsically worthless. The company's initial success was based on a standard pyramid scheme, with MMM using the funds from the sale of new shares to pay the promised returns on the old shares. Despite repeated warnings by government officials, including Boris Yeltsin, then president of the Russian Federation, that MMM was a scam and that the increase in the price of shares was a bubble, the promised returns were just too attractive to many Russians, especially in the midst of a deep economic recession.

The scheme could work only as long as the number of new shareholders—and thus new funds to be distributed to existing shareholders—increased fast enough. By the end of July 1994, the company could no longer make good on its promises and the scheme collapsed. The company closed. Mavrodi tried to blackmail the government into paying the shareholders, claiming that not doing so would trigger a revolution or a civil war. The government refused, leading many shareholders to be angry at the government rather than at Mavrodi. Later on in the year, Mavrodi actually ran for Parliament, as a self-appointed defender of the shareholders who had lost their savings. He won!

by 23% in two days, and to October 19, 1987, when the Dow Jones index fell by 22.6% in a single day. They point to the amazing rise in the Nikkei index (an index of Japanese stock prices) from around 13,000 in 1985 to around 35,000 in 1989, followed by a decline to 16,000 in 1992. In each of these cases, economists point to a lack of obvious news or at least of news important enough to cause such enormous movements.

Instead, they argue that stock prices are not always equal to their **fundamental value**, defined as the present value of expected dividends given in equation (14.17), and that stocks are sometimes underpriced or overpriced. Overpricing eventually comes to an end, sometimes with a crash, as in October 1929, or with a long slide, as in the case of the Nikkei index.

Under what conditions can such mispricing occur? The surprising answer is that it can occur even when investors are rational and when arbitrage holds. To see why, consider the case of a truly worthless stock (that is, the stock of a company that all financial investors know will never make profits and will never pay dividends). Putting D^e_{t+1}, D^e_{t+2}, and so on equal to zero in equation (14.17) yields a simple and unsurprising answer: The fundamental value of such a stock is equal to zero.

Might you nevertheless be willing to pay a positive price for this stock? Maybe. You might if you expect the price at which you can sell the stock next year to be higher than this year's price. And the same applies to a buyer next year. He may well be willing to buy at a high price if he expects to sell at an even higher price the following year. This process suggests that stock prices may increase just because investors expect them to. Such movements in stock prices are called **rational speculative bubbles**. Financial investors might well be behaving rationally as the bubble inflates. Even investors who hold the stock at the time of the crash, and therefore sustain a large loss, may have been rational. They may have realized there was a chance of a crash but also a chance that the bubble would continue and they could sell at an even higher price.

To make things simple, our example assumed the stock to be fundamentally worthless. But the argument is general and applies to stocks with a positive fundamental value as well. People might be willing to pay more than the fundamental value of a stock if they expect its price to increase in the future. And the same argument applies to other assets, such as housing, gold, and paintings. Two such bubbles are described in the Focus Box "Famous Bubbles: From Tulipmania in 17th-Century Holland to Russia in 1994."

Are all deviations from fundamental values in financial markets rational bubbles? Probably not. The fact is that many investors are not rational. An increase in stock prices in the past, say due to a succession of good news, often creates excessive optimism. If investors simply extrapolate from past returns to predict future returns, a stock may become "hot" (high priced) for no reason other than its price has increased in the past. This is true not only of stocks but also of houses. (See the Focus Box "The Increase in US Housing Prices in the United States in the 2000s: Fundamentals or Bubble?") Such deviations of stock prices from their fundamental value are sometimes called **fads**. We are all aware of fads outside of the stock market; there are good reasons to believe they exist in the stock market as well. The cartoon below makes the point nicely.

We have focused in this chapter on the determination of asset prices. The reason this belongs in a macroeconomic text is that asset prices are more than just a sideshow. They affect economic activity by influencing consumption and investment spending. There is little question, for example, that the decline in the stock market was one of the factors behind the 2001 recession. Most economists also believe that the stock market crash of 1929 was a major source of the Great Depression. And as we saw in Chapter 6, the decline in housing prices was the trigger for the Great Financial Crisis. These interactions among asset prices, expectations, and economic activity are the topics of the next two chapters.

FOCUS

The Increase in US Housing Prices During the First Half of the 2000s: Fundamentals or Bubble?

Recall from Chapter 6 that the trigger behind the current crisis was a decline in housing prices that started in 2006 (see Figure 6–7 for the evolution of the housing price index). In retrospect, the large increase from 2000 on that preceded the decline is now widely interpreted as a bubble. But in real time as prices went up, there was little agreement as to what laid behind this increase.

Economists belonged to three camps:

The pessimists argued that the price increases could not be justified by fundamentals. In 2005, Robert Shiller said: "The home-price bubble feels like the stock-market mania in the fall of 1999, just before the stock bubble burst in early 2000, with all the hype, herd investing and absolute confidence in the inevitability of continuing price appreciation."

To understand his position, go back to the derivation of stock prices in the text. We saw that, absent bubbles, we can think of stock prices as depending on current and expected future interest rates, current and expected future dividends, and a risk premium. The same applies to house prices. Absent bubbles, we can think of house prices as depending on current and expected future interest rates, current and expected rents, and a risk premium. In that context, pessimists pointed out that the increase in house prices was not matched by a parallel increase in rents. You can see this in Figure 1, which plots the price-to-rent ratio (i.e., the ratio of an index of house prices to an index of rents) since 1987 (the index is set so its value in January 1987 is 100). After remaining roughly constant from 1987 to the late 1990s, the ratio increased by nearly 60%, reaching a peak in 2006 before declining. (It remains substantially lower than the peak.) Furthermore, Shiller pointed out, surveys of house buyers suggested extremely high expectations of continuing large increases in housing prices—often in excess of 10% a year—and thus of large capital gains. As we saw previously, if assets are valued at their fundamental value, investors should not be expecting large capital gains in the future.

The optimists argued that there were good reasons for the price-to-rent ratio to go up. First, as we saw in Figure 6–2, the real interest rate was decreasing, increasing the present value of rents. Second, the mortgage market was changing. More people were able to borrow and buy a house. And those who borrowed were able to borrow a larger proportion of the value of the house. Both factors contributed to an increase in demand, and thus an increase in house prices. The optimists also pointed out that, every year since 2000, the pessimists had kept predicting the end of the bubble, and prices continued to increase. The pessimists were losing credibility.

The third group was by far the largest and remained agnostic. (Harry Truman is reported to have said: "Give me a one-handed economist! All my economists say, 'On the one hand... on the other.'") They concluded that the increase in house prices reflected both improved fundamentals and bubbles and that it was difficult to identify their relative importance.

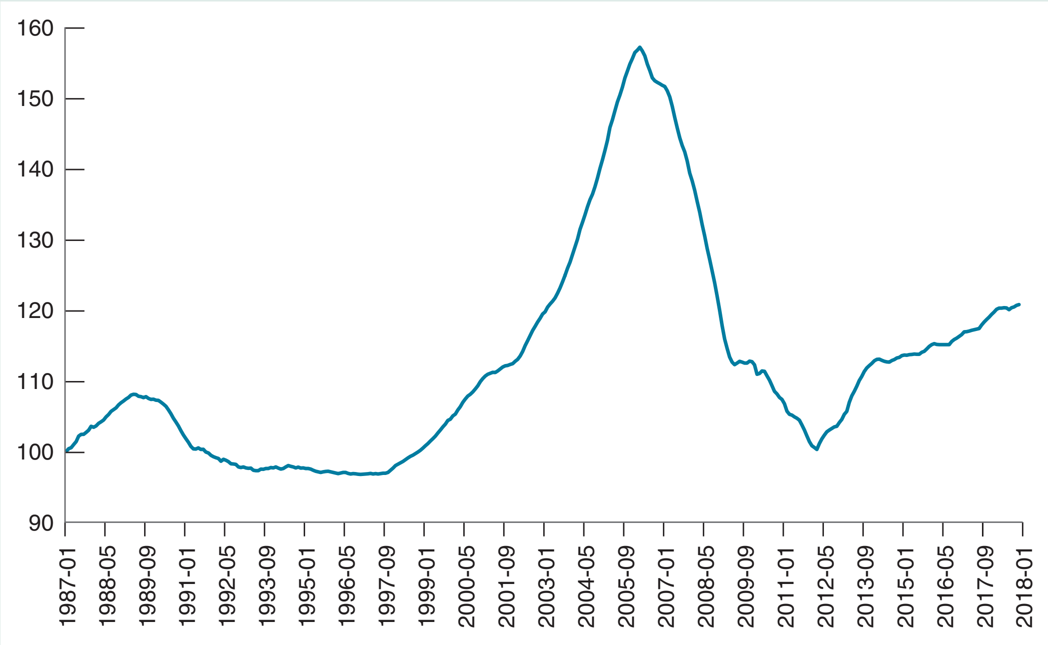


Figure 1

The US Housing Price-to-Rent Ratio since 1987

Source: FRED: CSUSHPISA: Case-Shiller Home Price Index. CUSR0000SEHA, Rent of Primary Residence.

What conclusions should you draw? The pessimists were clearly largely right. But bubbles and fads are clearer to see in retrospect than while they are taking place. This makes the task of policymakers much harder. If they were sure it was a bubble, they should try to stop it before it gets too large and then bursts. But they can rarely be sure until it is too late.[1]

Jason Zweig.com

SUMMARY

- The expected present discounted value of a sequence of payments equals the value this year of the expected sequence of payments. It depends positively on current and future expected payments and negatively on current and future expected interest rates.
- When discounting a sequence of current and expected future nominal payments, one should use current and expected future nominal interest rates. In discounting a sequence of current and expected future real payments, one should use current and expected future real interest rates.
- Arbitrage between bonds of different maturities implies that the price of a bond is the present value of the payments on the bond, discounted using current and expected short-term interest rates over the life of the bond, plus a risk premium. Higher current or expected short-term interest rates lead to lower bond prices.
- The yield to maturity on a bond is (approximately) equal to the average of current and expected short-term interest rates over the life of a bond, plus a risk premium.
- The slope of the yield curve—equivalently, the term structure—indicates what financial markets expect to happen to short-term interest rates in the future.
- The fundamental value of a stock is the present value of expected future real dividends, discounted using current and future expected one-year real interest rates plus the equity premium. In the absence of bubbles or fads, the price of a stock is equal to its fundamental value.
- An increase in expected dividends leads to an increase in the fundamental value of stocks; an increase in current and expected one-year interest rates leads to a decrease in their fundamental value.
- Changes in output may or may not be associated with changes in stock prices in the same direction. Whether they are or not depends on (1) what the market expected in the first place, (2) the source of the shocks, and (3) how the market expects the central bank to react to the output change.
- Asset prices can be subject to bubbles and fads that cause the price to differ from its fundamental value. Bubbles are episodes in which financial investors buy an asset for a price higher than its fundamental value, anticipating to resell it at an even higher price. Fads are episodes in which, because of excessive optimism, financial investors are willing to pay more for an asset than its fundamental value.

[1]Source: *"Reasonable People Did Disagree: Optimism and Pessimism about the US Housing Market before the Crash," Kristopher S. Gerardi, Christopher L. Foote, and Paul S. Willen, Federal Reserve Bank of Boston, Discussion Paper No. 10–5, September 10, 2010, available at www.bostonfed.org/economic/ppdp/2010/ppdp1005.pdf.*

KEY TERMS

expected present discounted value, 288
discount factor, 288
discount rate, 288
present discounted value, 289
present value, 289
maturity, 292
default risk, 293
price risk, 293
yield to maturity, 293
yield, 293
short-term interest rate, 293
long-term interest rate, 293
yield curve, 293
term structure of interest rates, 293
government bonds, 294
corporate bonds, 294
bond ratings, 294
risk premium, 294
junk bonds, 294
discount bonds, 294
face value, 294
coupon bonds, 294
coupon payments, 294
coupon rate, 294
current yield, 294
life (of a bond), 294
Treasury bills (T-bills), 294
Treasury notes, 294
Treasury bonds, 294
term premium, 294
indexed bonds, 294
Treasury Inflation Protected Securities (TIPS), 294
arbitrage, 295
expectations hypothesis, 295
n-year interest rate, 296
internal finance, 299
external finance, 299
debt finance, 299
equity finance, 299
stocks or shares, 299
dividends, 299
ex-dividend price, 300
equity premium, 300
random walk, 302
fundamental value, 307
rational speculative bubbles, 307
fads, 309

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The present discounted value of a stream of returns can be calculated in real or nominal terms.
b. The higher the one-year interest rate, the lower the present discounted value of a payment next year.
c. One-year interest rates are normally expected to be constant over time.
d. Bonds are a claim to a sequence of constant payments over a number of years.
e. Stocks are a claim to a sequence of dividend payments over a number of years.
f. House prices are a claim to a sequence of expected future rents over a number of years.
g. The yield curve normally slopes up.
h. All assets held for one year should have the same expected rate of return.
i. In a bubble, the value of the asset is the expected present value of its future returns.
j. The overall real value of the stock market does not fluctuate very much over a year.
k. Indexed bonds protect the holder against unexpected inflation.

2. For which of the problems listed in parts a through c would you want to use real payments and real interest rates, and for which would you want to use nominal payments and nominal interest rates to compute the expected present discounted value? In each case, explain why.

a. Estimating the present discounted value of the profits from an investment in a new machine.
b. Estimating the present value of a 20-year US Treasury bond.
c. Deciding whether to buy or lease a car.

3. Compute the two-year nominal interest rate using the exact formula and the approximation formula for each set of assumptions listed in parts a through c.

a. $i_t = 2\%$; $i^e_{t+1} = 3\%$
b. $i_t = 2\%$; $i^e_{t+1} = 10\%$
c. $i_t = 2\%$; $i^e_{t+1} = 3\%$. The term premium on a two-year bond is 1%.

4. The equity premium and the value of stocks

a. Explain why, in equation (14.14), it is important that the stock is ex-dividend, that is, it has just paid its dividend and is expected to pay its next dividend in one year.
b. Using of equation (14.14), explain the contribution of each component to today's stock price.

c. If the risk premium is larger, all else being equal, what happens to the price of the stock today?
d. If the one-period interest rate increases, what happens to the price of the stock today?
e. If the expected value of the stock at the beginning of period $t + 1$ increases, what happens to the value of the stock today?
f. Now look carefully at equation (14.15). Set $i_{1t} = i_{1t+n} = 0.05$ for all n. Set $x = 0.03$. Compute the coefficients on $\$D^e_{t+3}$ and $\$D^e_{t+10}$. Compare the effect of a \$1 expected increase in a dividend 3 years from now and 10 years from now.
g. Repeat the computation in part f with $i_{1t} = i_{1t+n} = 0.08$ for all n and $x = 0.05$.

5. *Approximating the price of long-term bonds*

The present value of an infinite stream of dollar payments of \$z (that starts next year) is \$z/i when the nominal interest rate, i, is constant. This formula gives the price of a consol—a bond paying a fixed nominal payment each year, forever. It is also a good approximation for the present discounted value of a stream of constant payments over long but not infinite periods, as long as i is constant. Let's examine how close the approximation is.

a. Suppose that $i = 10\%$. Let $\$z = 100$. What is the present value of the consol?
b. If $i = 10\%$, what is the expected present discounted value of a bond that pays $\$z$ over the next 10 years? 20 years? 30 years? 60 years? (*Hint*: Use the formula from the chapter but remember to adjust for the first payment.)
c. Repeat the calculations in parts a and b for $i = 2\%$ and $i = 5\%$.

6. *Monetary policy and the stock market*

Assume the short term real policy rate, current and expected, had been 2% until now. Suppose the Fed decides to tighten monetary policy and increase the short-term policy rate (r_{1t}) from 2% to 3%.

a. What happens to stock prices if the change in r_{1t} is expected to be temporary, that is, last for only one period? Assume that expected real dividends do not change. Use equation (14.17).
b. What happens to stock prices if the change in r_{1t} is expected to be permanent, that is, is expected to persist? Assume that expected real dividends do not change. Use equation (14.17).
c. What happens to stock prices today if the increase in the real interest rate, current and expected, reflects an increase in expected future output and expected future dividends?

DIG DEEPER

7. *Regular IRAs versus Roth IRAs*

You want to save \$2,000 today for retirement in 40 years. You have to choose between the two plans listed in (i) and (ii).

i. Pay no taxes today, put the money in an interest-yielding account, and pay taxes equal to 20% of the total amount withdrawn at retirement. (In the United States, such an account is known as a regular *individual retirement account* [IRA].)
ii. Pay taxes equivalent to 30% of the investment amount today, put the remainder in an interest-yielding account, and pay no taxes when you withdraw your funds at retirement. (In the United States, this is known as a *Roth IRA*.)
   a. What is the expected present discounted value of each of these plans if the real interest rate is 1%? 10%?
   b. Which plan would you choose in each case?

8. *House prices and bubbles*

Houses can be thought of as assets with a fundamental value equal to the expected present discounted value of their future real rents.

a. Would you prefer to use real payments and real interest rates to value a house or nominal payments and nominal interest rates?
b. The rent on a house, whether you live in the house yourself and thus save paying the rent to an owner, or whether you own the house and rent it, is like the dividend on a stock. Write the equivalent of equation (14.17) for a house.
c. Why would low interest rates help explain an increase in the price-to-rent ratio?
d. If housing is perceived as a safer investment, what will happen to the price-to-rent ratio?
e. The Focus Box "The Increase in US Housing Prices: Fundamental or Bubble?" has a graph of the price-to-rent ratio. You should be able to find the value of the Case-Shiller home price index and the rental component of the consumer price index in the FRED economic database maintained at the Federal Reserve Bank of St. Louis (variables *SPCS20RSA* and *CUSR0000SEHA*, respectively). The graph in Figure 1 in this Focus Box ends with data from November 2018. Calculate the percentage increase in the home price index between November 2018 and the latest date available. Calculate the percentage increase in the rent price index from November 2018 to the latest date available. Has the price-to-rent ratio increased or decreased since June 2018?

EXPLORE FURTHER

9. *House prices around the world*

The Economist annually publishes The Economist House Price Index. It attempts to assess which housing markets, by country, are the most overvalued or undervalued relative to fundamentals. Find the most recent version of this data on the Web.

a. One index of overvaluation is the ratio of house prices to rents. Why might this index help detect a housing price bubble? Using the data you are studying, in which country are house prices most overvalued by the ratio of prices to rents? Would this measure have helped predict the US housing market crash?
b. A second index is the ratio of house prices to income. Why might this index help detect a housing price bubble? Using this data, in which country are houses most overvalued by the ratio of prices to rents? Would this measure have helped predict the US housing market crash?

10. *Inflation-indexed bonds*

Some bonds issued by the US Treasury make payments indexed to inflation. These inflation-indexed bonds compensate *investors for inflation. Therefore, the current interest rates on these bonds are real interest rates—interest rates in terms of goods. These interest rates can be used, together with nominal interest rates, to provide a measure of expected inflation. Let's see how.*

Go to the Web site of the Federal Reserve Board and get the most recent statistical release listing interest rates (www.federalreserve.gov/releases/h15/Current). Find the current nominal interest rate on Treasury securities with a five-year maturity. Now find the current interest rate on "inflation-indexed" Treasury securities with a five-year maturity. What do you think participants in financial markets believe the average inflation rate will be over the next five years?

FURTHER READINGS

- There are many bad books written about the stock market. A good one, and one that is fun to read, is Burton Malkiel, *A Random Walk Down Wall Street*, 12th ed. (2019, W. W. Norton).
- A description of some historical bubbles is given by Peter Garber in "Famous First Bubbles," *Journal of Economic Perspectives*, Spring 1990, 4(2): pp. 35–54.

APPENDIX: Deriving the Expected Present Discounted Value Using Real or Nominal Interest Rates

This appendix shows that the two ways of expressing present discounted values, equations (14.1) and (14.3), are equivalent.

Equation (14.1) gives the present value as the sum of current and future expected *nominal payments*, discounted using current and future expected *nominal interest rates*:

$$\$V_t = \$z_t + \frac{1}{1+i_t}\$z^e_{t+1} + \frac{1}{(1+i_t)(1+i^e_{t+1})}\$z^e_{t+2} + \cdots \qquad (14.1)$$

Equation (14.3) gives the present value as the sum of current and future expected *real payments*, discounted using current and future expected *real interest rates*:

$$V_t = z_t + \frac{1}{1+r_t}z^e_{t+1} + \frac{1}{(1+r_t)(1+r^e_{t+1})}z^e_{t+2} + \cdots \qquad (14.3)$$

Divide both sides of equation (14.1) by the current price level, P_t, so:

$$\frac{\$V_t}{P_t} = \frac{\$z_t}{P_t} + \frac{1}{1+i_t}\frac{\$z^e_{t+1}}{P_t} + \frac{1}{(1+i_t)(1+i^e_{t+1})}\frac{\$z^e_{t+2}}{P_t} + \cdots \qquad (14.A1)$$

Let's look at each term on the right side of equation (14.A1) and show that it is equal to the corresponding term in equation (14.3):

- Take the first term, $\$z_t/P_t = z_t$, the real value of the current payment. So $\$z_t/P_t$ is the same as z_t, the first term on the right of equation (14.3).
- Take the second term:

$$\frac{1}{1+i_t}\frac{\$z^e_{t+1}}{P_t}$$

Multiply the numerator and the denominator by P^e_{t+1}, the price level expected for next year, to get:

$$\frac{1}{1+i_t}\frac{P^e_{t+1}}{P_t}\frac{\$z^e_{t+1}}{P^e_{t+1}}$$

The fraction on the right, $\$z^e_{t+1}/P^e_{t+1}$, is equal to z^e_{t+1}, the expected real payment at time $t+1$. The fraction in the middle, P^e_{t+1}/P_t, can be rewritten as $1 + [(P^e_{t+1} - P_t)/P_t]$. Using the definition of expected inflation as $(1+\pi^e_{t+1})$ and rewriting the middle term, we arrive at:

$$\frac{(1+\pi^e_{t+1})}{(1+i_t)}z^e_{t+1}$$

Recall the relation among the real interest rate, the nominal interest rate, and expected inflation in equation (14.3): $(1+r_t) = (1+i_t)/(1+\pi^e_{t+1})$. Using this relation in the previous equation gives:

$$\frac{1}{(1+r_t)}z^e_{t+1}$$

This term is the same as the second term on the right side of equation (14.3).

- The same method can be used to rewrite the other terms; make sure that you can derive the next one.

We have shown that the right sides of equations (14.A1) and (14.3) are equal to each other. It follows that the terms on the left side are equal, so:

$$V_t = \frac{\$V_t}{P_t}$$

This says that the present value of current and future expected *real payments*, discounted using current and future expected *real interest rates* (the term on the left side), is equal to the present value of current and future expected *nominal payments*, discounted using current and future expected *nominal interest rates*, divided by the current price level (the term on the left side).

15

Expectations, Consumption, and Investment

Having looked at the role of expectations in financial markets, we now turn to the role expectations play in determining the two main components of spending—consumption and investment. This description of consumption and investment will be the main building block of the expanded IS-LM model we develop in Chapter 16.

Section 15-1 looks at consumption and shows that consumption decisions depend not only on a person's current income but also on his or her expected future income and on financial wealth.

Section 15-2 turns to investment and shows that investment decisions depend on current and expected profits and on current and expected real interest rates.

Section 15-3 looks at movements in consumption and investment over time and shows how to interpret those movements in light of what you learned in this chapter.

If you remember one basic message from this chapter, it should be: Both consumption and investment decisions depend very much on expectations about the future. ▶▶▶

15-1 CONSUMPTION

How do people decide how much to consume and how much to save? In Chapter 3, we assumed that consumption depended only on current income. But even then, it was clear that consumption depended on much more, particularly on expectations about the future. We now explore how those expectations affect consumption decisions.

The modern theory of consumption, on which this section is based, was developed independently in the 1950s by Milton Friedman of the University of Chicago, who called it the **permanent income theory of consumption**, and by Franco Modigliani of MIT, who called it the **life-cycle theory of consumption**. Each chose his label carefully. Friedman's "permanent income" emphasized that consumers look beyond current income. Modigliani's "life-cycle" emphasized that consumers' natural planning horizon is their entire lifetime.

Friedman received the Nobel Prize in economics in 1976; Modigliani received the Nobel Prize in economics in 1985.

The behavior of aggregate consumption has remained a hot area of research ever since, for two reasons: One is simply the sheer size of consumption as a component of GDP and therefore the need to understand movements in consumption. The other is the increasing availability of large surveys of individual consumers, such as the Panel Study of Income Dynamics (PSID), described in the Focus Box "Up Close and Personal: Learning from Panel Datasets." These surveys, which were not available when Friedman and Modigliani developed their theories, have allowed economists to steadily improve their understanding of how consumers actually behave. This section summarizes what we know today.

From Chapter 3: Consumption spending accounts for 68% of total spending in the United States.

That's not to mention the information about consumer behavior increasingly available from the internet. How to best use this information, with machine learning techniques, is one of the frontiers of empirical research today.

The Very Foresighted Consumer

Let's start with an assumption that will surely—and rightly—strike you as extreme but will serve as a convenient benchmark. We'll call it the theory of the *very foresighted consumer*. How would a foresighted consumer decide how much to consume? She would proceed in two steps:

- First, she would add up the value of the stocks and bonds she owns, the value of her checking and savings accounts, the value of the house she owns minus the mortgage still due, and so on. This would give her an idea of her **financial wealth** and her **housing wealth**.

  She would also estimate what her after-tax labor income was likely to be over her working life and compute the present value of expected after-tax labor income. This would give her an estimate of what economists call her **human wealth**—to contrast it with her **nonhuman wealth**, defined as the sum of financial wealth and housing wealth.
- Adding her human wealth and nonhuman wealth, she would have an estimate of her **total wealth**. She would then decide how much to spend out of this total wealth. A reasonable assumption is that she would decide to spend a proportion of her total wealth such as to maintain roughly the same level of consumption each year throughout her life. If that level of consumption were higher than her current income, she would borrow the difference. If it were lower than her current income, she would save the difference.

With a slight abuse of language, we shall use *housing wealth* to refer not only to housing but also to the other goods that the consumer may own, from cars to paintings, and so on.

Human wealth + Nonhuman wealth = Total wealth

Let's write this formally. What we have described is a consumption decision of the form

$$C_t = C(\text{total wealth}_t) \qquad (15.1)$$

where C_t is consumption at time t, and (total wealth_t) is the sum of nonhuman wealth (financial plus housing wealth) and human wealth at time t (the expected present value, as of time t, of current and future after-tax labor income).

Up Close and Personal: Learning from Panel Datasets

FOCUS

Panel datasets show the value of one or more variables for many individuals or many firms over time. We described one such survey, the Current Population Survey (CPS), in Chapter 7. Another is the *Panel Study of Income Dynamics (PSID)*.

The PSID was started in 1968 with approximately 4,800 families. Interviews of these families have been conducted every year since and continue today. The survey has grown as new individuals have joined the original families surveyed, either by marriage or by birth. Each year, the survey asks people about their income, wage rate, number of hours worked, health, and consumption.

By providing five decades of information about individuals and their extended families, the survey has allowed economists to ask and answer questions for which there was previously only anecdotal evidence. Among the many questions for which the PSID has been used are:

- How much does consumption respond to transitory movements in income—for example, to the loss of income from becoming unemployed?
- How much risk sharing exists within families? For example, when a family member becomes sick or unemployed, how much help does he or she get from other family members?
- How much do people care about staying geographically close to their families? When someone becomes unemployed, for example, how does the probability that he will migrate to another city depend on the number of his family members living in the city where he currently lives?

For a more detailed description, see Katherine A. McGonagle et al., "The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research," *Longitudinal and Life Course Studies*, 2012, 3(2): pp. 268–284, https://psidonline.isr.umich.edu/llcs2012.pdf.

This description contains much truth. Like the foresighted consumer, we surely think about our wealth and our expected future labor income when deciding how much to consume today. But one cannot help thinking that it assumes too much computation and foresight on the part of the typical consumer.

To get a better sense of what this description implies and what is wrong with it, let's apply this decision process to a problem facing a typical US college student.

An Example

Let's assume you are 19 years old, with three more years of college before you start your first job. You may be in debt today, having taken out a loan to go to college. You may own a car and a few other worldly possessions. For simplicity, let's assume your debt and your possessions roughly offset each other, so that your nonhuman wealth is equal to zero. Your only wealth therefore is your human wealth, the present value of your expected after-tax labor income.

You expect that your starting annual salary in three years will be around \$40,000 (in 2018 dollars) and will increase by an average of 3% per year in real terms, until your retirement at age 60. About 25% of your income will go to taxes.

You are welcome to use your own numbers and see where the computation takes you.

Building on what we saw in Chapter 14, let's compute the present value of your labor income as the value of *real* expected after-tax labor income, discounted using *real* interest rates. Let Y_{Lt} denote real labor income in year t. Let T_t denote real taxes in year t. Let $V(Y^e_{Lt} - T^e_t)$ denote your human wealth; that is, the expected present value of your after-tax labor income—expected as of year t.

To make the computation simple, assume the real interest rate at which you can borrow is equal to zero, so the expected present value is simply the sum of expected labor income over your working life and is therefore given by

$$V(Y^e_{Lt} - T^e_t) = (\$40{,}000)(0.75)[1 + (1.03) + (1.03)^2 + \cdots + (1.03)^{38}]$$

The first term (\$40,000) is your initial level of labor income, in year 2018 dollars.

The second term (0.75) comes from the fact that, because of taxes, you keep only 75% of what you earn.

The third term $[1 + (1.03) + (1.03)^2 + \cdots + (1.03)^{38}]$ reflects the fact that you expect your real income to increase by 3% a year for 39 years (you will start earning income at age 22 and work until age 60).

The computation of the consumption level you can sustain is made much easier by our assumption that the real interest rate you face equals zero. In this case, if you consume one less good today, you can consume exactly one more good next year, and the condition you must satisfy is simply that the sum of consumption over your lifetime is equal to your wealth. So if you want to consume a constant amount each year, you just need to divide your wealth by the remaining number of years you expect to live.

One of the goals of the field of "behavioral economics" has been to assess whether actual behavior looks more like the behavior of the foresighted consumer or more like the cartoon.

Using the properties of geometric series to solve for the sum in brackets gives

$$V(Y^e_{Lt} - T^e_t) = (\$40{,}000)(0.75)(72.2) = \$2{,}166{,}000$$

Your wealth today, the expected value of your lifetime after-tax labor income, is around $2 million.

How much should you consume? You can expect to live about 20 years after you retire, so that your expected remaining life today is 62 years. If you want to consume the same amount every year, the constant level of consumption that you can afford equals your total wealth divided by your expected remaining life, or $\$2{,}166{,}000/62 = \$34{,}935$ a year. Given that your income until you get your first job is equal to zero, this implies you will have to borrow $34,935 a year for the next three years and begin to save when you get your first job.

Toward a More Realistic Description

Your first reaction to this computation may be that this is a stark and slightly sinister way of summarizing your life prospects. You might find yourself more in agreement with the retirement plans described in the cartoon on the next page.

Your second reaction may be that although you agree with most of the ingredients that went into the computation, you surely do not intend to borrow $\$34{,}935 \times 3 = \$104{,}805$ over the next three years. For example:

1. You might not want to plan for constant consumption over your lifetime. Instead you may be quite happy to defer higher consumption until later. Student life usually does not leave much time for expensive activities. You may want to defer trips to the Galapagos Islands to later in life. You also have to think about the additional expenses that will come with having children, sending them to nursery school, summer camp, college, and so on.

2. You might find that the amount of computation and foresight involved in the computation we just went through far exceeds the amount you use in your own decisions. You may never have thought until now about exactly how much income you are going to earn and for how many years. You might feel that most consumption decisions are made in a simpler, less forward-looking fashion.

3. The computation of total wealth is based on forecasts of what is expected to happen. But things can turn out better or worse. What happens if you are unlucky and you become unemployed or sick? How will you pay back what you borrowed? You might want to be prudent, making sure that you can adequately survive even the worst outcomes, and thus decide to borrow much less than $104,805.

4. Even if you decide to borrow $104,805, you might have a hard time finding a bank willing to lend it to you. Why? The bank may worry that you are taking on a commitment you will not be able to afford if times turn bad and that you may not be able or willing to repay the loan. In other words, if you want to borrow this much money, the borrowing rate you face may be much higher than assumed in the computation.

These reasons, all good ones, suggest that to characterize consumers' actual behavior, we must modify the description we gave previously. The last three reasons in particular suggest that consumption depends not only on total wealth but also on current income.

Take the second reason: You may, because it is a simple rule, decide to let your consumption follow your income and not think about what your wealth might be. In that case your consumption will depend on your current income, not on your wealth.

Now take the third reason: It implies that a safe rule may be to consume no more than your current income. This way, you do not run the risk of accumulating debt that you cannot repay if times turn bad.

Roz Chast / The New Yorker Collection/The Cartoon Bank.

Or take the fourth reason: It implies that you may have little choice anyway. Even if you wanted to consume more than your current income, you might be unable to do so because no bank will give you a loan.

If we want to allow for a direct effect of current income on consumption, what measure of current income should we use? A convenient measure is after-tax labor income, which we introduced when we defined human wealth. This leads to a consumption function of the form

$$C_t = C(\text{Total wealth}_t, Y_{Lt} - T_t) \qquad (15.2)$$
$$(\quad + \quad , \quad + \quad)$$

FOCUS

Do People Save Enough for Retirement?

How carefully do people look forward when making consumption and saving decisions? One way to answer this question is to look at how much people save for retirement.

Table 1, taken from a study by James Poterba of MIT, Steven Venti of Dartmouth, and David Wise of Harvard, gives some basic numbers. They are based on a panel dataset called the *Health and Retirement Study*, a panel study run by the University of Michigan that surveys a representative sample of approximately 20,000 Americans over the age of 50 every two years. The table shows the mean level and the composition of (total) wealth for people aged between 65 and 69 years in 2008—so, most of them retired. It also distinguishes between people who reach that age as singles or as a couple; in this case the numbers refer to the wealth of the couple.

The first three components of wealth capture the various sources of retirement income. The first is the present value of Social Security benefits. The second is the value of the retirement plans provided by employers. And the third is the value of personal retirement plans. The last three components include the other assets held by consumers, such as bonds and stocks, and housing.

A mean wealth of $1.1 million dollars for a couple is substantial. It gives an image of forward-looking individuals making careful saving decisions and retiring with enough wealth to enjoy a comfortable retirement.

We must be careful, however. The high average may hide important differences across individuals. Some individuals may save a lot, others little. Another study, by John Scholz, Ananth Seshadri, and Surachai Khitatrakun, from the University of Wisconsin, sheds light on this aspect, also using data from the Health and Retirement Study. The authors construct a target level of wealth for each household (i.e., the wealth level that each household should have if it wants to maintain a roughly constant level of consumption after retirement). The authors then compare the actual wealth level to the target level for each household.

The first conclusion of their study is similar to the conclusion reached by Poterba, Venti, and Wise: On average, people save enough for retirement. More specifically, the authors find that more than 80% of households have wealth above the target level. Put the other way around, only 20% of households have wealth below the target. But these numbers hide important differences across income levels.

Among those in the top half of the income distribution, more than 90% have wealth that exceeds the target often by a large amount. This suggests that these households plan to leave bequests and so save more than what is needed for retirement.

Among those in the bottom 20% of the income distribution, however, fewer than 70% have wealth above the target. For the 30% of households below the target, the difference between actual and target wealth is typically small. But the relatively large proportion of individuals with wealth below the target suggests that there are a number of individuals who, through bad planning or bad luck, do not save enough for retirement. For most of these individuals, nearly all their wealth comes from the present value of Social Security benefits (the first component of wealth in Table 1), and it is reasonable to think that the proportion of people with wealth below the target would be even larger if Social Security did not exist. This is indeed what the Social Security system was designed to do: to make sure that people have enough to live on when they retire. In that regard, it appears to be a success.[1]

Table 1 Mean Wealth of People, Age 65–69, in 2008 (in thousands of 2008 dollars)

	Married couples	Single-person household
Social Security pension	262	134
Employer-provided pension	129	63
Personal retirement assets	182	47
Other financial assets	173	83
Home equity	340	188
Other equity	69	18
Total	1,155	533

Source: James M. Poterba, Steven F. Venti, and David A. Wise, "The composition and drawdown of wealth in retirement," *Journal of Economic Perspectives*, 25(4), pages 95–118, Fall 2011.

[1]Sources: *James M. Poterba, Steven F. Venti, and David A. Wise, "The Composition and Drawdown of Wealth in Retirement,"* Journal of Economic Perspectives, *2011, 25(4): pp. 95–118. John Scholz, Ananth Seshadri, and Surachai Khitatrakun, "Are Americans Saving 'Optimally' for Retirement?"* Journal of Political Economy, *2006, 114(4): pp. 607–643.*

In words, consumption is an increasing function of total wealth and also an increasing function of current after-tax labor income. Total wealth is the sum of nonhuman wealth—financial wealth plus housing wealth—and human wealth—the present value of expected after-tax labor income.

How much does consumption depend on total wealth (and therefore on expectations of future income) and how much does it depend on current income? The evidence is that most consumers look forward in the spirit of the theory developed by Modigliani and Friedman. (See the Focus Box "Do People Save Enough for Retirement?") But some consumers, especially those who have temporarily low income and poor access to credit, are likely to consume their current income, regardless of what they expect will happen to them in the future. A worker who becomes unemployed and has no financial wealth may have a hard time borrowing to maintain his level of consumption, even if he is fairly confident that he will soon find another job. Consumers who are richer and have easier access to credit are more likely to give more weight to the expected future and to try to maintain roughly constant consumption over time.

Putting Things Together: Current Income, Expectations, and Consumption

Let's go back to the importance of expectations in the determination of spending. Note first that, with consumption behavior described by equation (15.2), expectations affect consumption in two ways:

- Expectations affect consumption directly through *human wealth*: To compute their human wealth, consumers must form their own expectations about future labor income, real interest rates, and taxes.
- Expectations affect consumption indirectly through *nonhuman wealth*—stocks, bonds, and housing. Consumers do not need to do any computation here and can just take the value of these assets as given. As you saw in Chapter 14, the computation is in effect done for them by participants in financial markets. The price of their stocks, for example, itself depends on expectations of future dividends and interest rates.

How expectations of higher output in the future affect consumption today:

Expected future output increases
$\Rightarrow$ Expected future labor income increases
$\Rightarrow$ Human wealth increases
$\Rightarrow$ Consumption increases

Expected future output increases
$\Rightarrow$ Expected future dividends increase
$\Rightarrow$ Stock prices increase
$\Rightarrow$ Nonhuman wealth increases
$\Rightarrow$ Consumption increases

This dependence of consumption on expectations has in turn two main implications for the relation between consumption and income:

- *Consumption is likely to respond less than one-for-one to fluctuations in current income.* When deciding how much to consume, a consumer looks at more than her current income. If she concludes that a decrease in her income is permanent, she is likely to decrease consumption one-for-one with the decrease in income. But if she concludes that the decrease in her current income is transitory, she will adjust her consumption by less. In a recession, consumption adjusts less than one-for-one to decreases in income. This is because consumers know that recessions typically do not last for more than a few quarters and that the economy will eventually recover. The same is true in expansions. Faced with an unusually rapid increase in income, consumers are unlikely to increase consumption by as much as income. They are likely to assume that the boom is transitory and that things will return to normal.
- *Consumption may move even if current income does not change.* The election of a charismatic president who articulates the vision of an exciting future may lead people to become more optimistic about the future in general, and about their own future income, leading them to increase consumption even if their current income does not change. Other events may have the opposite effect.

  The effects of the Great Recession are particularly striking in this respect. Using data from a survey of consumers, Figure 15-1 shows the evolution of expectations about family income growth over the following year, for each year since 1990. Note

Looking at the short run (Chapter 3), we assumed $C = c_0 + c_1Y$ (ignoring taxes here). This implied that, when income increased, consumption increased less than proportionately with income (C/Y went down). This was appropriate because our focus was on fluctuations, on transitory movements in income.

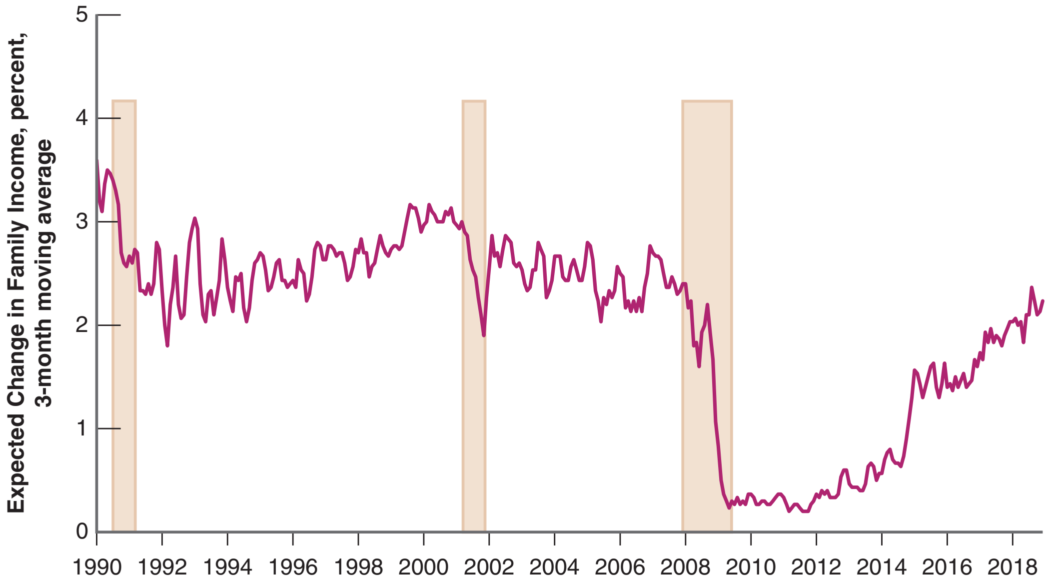
Looking at the long run (Chapter 10), we assumed that $S = sY$, or equivalently $C = (1 - s)Y$. This implied that, when income increased, consumption increased proportionately with income (C/Y remained the same). This was appropriate because our focus was on permanent—long-run—movements in income.

Figure 15-1

Expected Change in Family Income since 1990

After falling sharply in 2008 and 2009, expectations of income growth remained low for a long time.

Source: Survey of Consumers, Table 14, University of Michigan, https://data.sca.isr.umich.edu. Shaded areas denote recessions.



how relatively stable expectations remained until 2008, how sharply they dropped in 2008 and 2009, and how long they remained low after that. Only in 2014 did they start increasing.

The drop at the start of the crisis is not surprising. As consumers saw output falling, it was normal for them to expect a drop in income over the following year. But its magnitude is striking. Both previous recessions, in 1991 and in 2000, also had a drop in expected income growth, but it was much smaller and much less persistent. In 2008, consumers got very scared and remained scared for a long time. This led them to limit their consumption, and this in turn led to a slow recovery.

15-2 INVESTMENT

How do firms make investment decisions? In our first pass at the answer in the core (Chapter 5), we took investment to depend on the current interest rate and the current level of sales. We refined that answer in Chapter 6 by pointing out that what mattered was the real interest rate, not the nominal interest rate. It should now be clear that investment decisions, like consumption decisions, depend on more than current sales and the current real interest rate. They also depend very much on expectations of the future. We now explore how those expectations affect investment decisions.

Just like the basic theory of consumption, the basic theory of investment is straightforward. A firm deciding whether to invest—say, whether to buy a new machine—must make a simple comparison. It must first compute the present value of profits it can expect from having this additional machine. It must then compare the present value of profits to the cost of buying the machine. If the present value exceeds the cost, the firm should buy the machine—invest; if the present value is less than the cost, then the firm should not buy the machine—not invest. This, in a nutshell, is the theory of investment. Let's look at it in more detail.

Investment and Expectations of Profit

Let's go through the steps a firm must take to determine whether to buy a new machine. (Although we refer to a machine, the same reasoning applies to the other components of investment—the building of a new factory, the renovation of an office complex, and so on.)

Depreciation

To compute the present value of expected profits, the firm must first estimate how long the machine will last. Most machines are like cars. They can last nearly forever, but as time passes, they become more and more expensive to maintain and less and less reliable.

Look at cars in Cuba.

Assume a machine loses its usefulness at rate δ (the Greek lowercase letter delta) per year. A machine that is new this year is worth only $(1 - \delta)$ machine next year, $(1 - \delta)^2$ machine in two years, and so on. The *depreciation rate*, δ, measures how much usefulness the machine loses from one year to the next. What are reasonable values for δ? This is a question that the statisticians in charge of measuring the US capital stock have had to answer. Based on their studies of depreciation of specific machines and buildings, US statisticians use numbers from 2.5% for office buildings to 15% for communication equipment to 55% for prepackaged software.

If the firm has a large number of machines, we can think of δ as the proportion of machines that die every year (think of light bulbs, which work perfectly until they die). If the firm starts the year with K working machines and does not buy new ones, it will have $K(1 - \delta)$ machines left one year later, and so on.

The Present Value of Expected Profits

The firm must then compute the present value of expected profits.

To capture the fact that it takes some time to put machines in place (and even more time to build a factory or an office building), let's assume that a machine bought in year t becomes operational—and starts depreciating—only one year later, in year $t + 1$. Denote profit per machine in real terms by Π.

This is an uppercase Greek pi as opposed to the lowercase Greek pi, which we use to denote inflation.

If the firm buys a machine in year t, the machine will generate its first profit in year $t + 1$; denote this expected profit by Π^e_{t+1}. The present value, in year t, of this expected profit in year $t + 1$ is given by

For simplicity, and to focus on the role of expectations as opposed to risk, let us assume again that the risk premium is equal to zero, so we do not have to carry it along in the formulas below.

$$\frac{1}{1 + r_t}\Pi^e_{t+1}$$

This term is represented by the arrow pointing left in the upper line of Figure 15-2. Because we are measuring profit in real terms, we are using real interest rates to discount future real profits.

Denote expected profit per machine in year $t + 2$ by Π^e_{t+2}. Because of depreciation, only $(1 - \delta)$ of the machine is left in year $t + 2$, so the expected profit from the machine is equal to $(1 - \delta)\Pi^e_{t+2}$. The present value of this expected profit as of year t is equal to

$$\frac{1}{(1 + r_t)(1 + r^e_{t+1})}(1 - \delta)\,\Pi^e_{t+2}$$

This computation is represented by the arrow pointing left in the lower line of Figure 15-2.

The same reasoning applies to expected profits in the following years. Putting the pieces together gives us *the present value of expected profits* from buying the machine in year t, which we shall call $V(\Pi^e_t)$:

$$V(\Pi^e_t) = \frac{1}{1 + r_t}\Pi^e_{t+1} + \frac{1}{(1 + r_t)(1 + r^e_{t+1})}(1 - \delta)\,\Pi^e_{t+2} + \cdots \qquad (15.3)$$

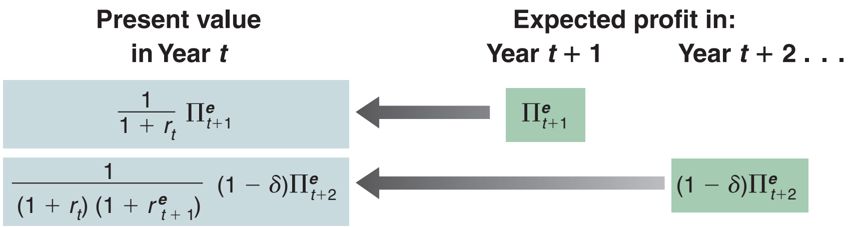


Figure 15-2

Computing the Present Value of Expected Profits

The expected present value is equal to the discounted value of expected profit next year, plus the discounted value of expected profit two years from now (taking into account the depreciation of the machine), and so on.

The Investment Decision

The firm must now decide whether to buy the machine. This decision depends on the relation between the present value of expected profits and the price of the machine. To simplify notation, let's assume the real price of a machine—that is, the price of the machine in terms of the basket of goods produced in the economy—equals 1. What the firm must then do is to compare the present value of profits to 1.

If the present value is less than 1, the firm should not buy the machine. If it did, it would be paying more for the machine than it expects to get back in profits later. If the present value exceeds 1, the firm has an incentive to buy the new machine.

Let's now go from this one-firm, one-machine example to investment in the economy as a whole.

Let I_t denote aggregate investment.

Denote profit per machine or, more generally, profit per unit of capital (where capital includes machines, factories, office buildings, and so on) for the economy as a whole by Π_t.

Denote the expected present value of profit per unit of capital by $V(\Pi_t^e)$, as defined in equation (15.3).

Our discussion suggests an investment function of the form

$$I_t = I\,[V(\Pi_t^e)] \qquad (15.4)$$
$$(\;+\;)$$

In words: Investment depends positively on the expected present value of future profits (per unit of capital). The higher the expected profits, the higher the expected present value and the level of investment. The higher expected real interest rates, the lower the expected present value, and thus the lower the level of investment.

If the present value computation the firm has to make strikes you as quite similar to the present value computation we saw in Chapter 14 for the fundamental value of a stock, you are right. This relation was first explored by James Tobin, of Yale University, who argued that, for this reason, there should be a tight relation between investment and the value of the stock market. His argument and the evidence are presented in the Focus Box "Investment and the Stock Market."

Tobin received the Nobel Prize in economics in 1981 for this and many other contributions.

A Convenient Special Case

Before exploring further implications and extensions of equation (15.4), it is useful to go through a special case where the relation between investment, profit, and interest rates becomes simple.

Suppose firms expect both future profits (per unit of capital) and future interest rates to remain at the same level as today, so that

$$\Pi_{t+1}^e = \Pi_{t+2}^e = \cdots = \Pi_t$$

and

$$r_{t+1}^e = r_{t+2}^e = \cdots = r_t$$

Economists call such expectations—that the future will be like the present—**static expectations**. Under these two assumptions, equation (15.3) becomes (the derivation is given in the appendix to this chapter)

$$V(\Pi_t^e) = \frac{\Pi_t}{r_t + \delta} \qquad (15.5)$$

Investment and the Stock Market

Suppose a firm has 100 machines and 100 shares outstanding—one share per machine. Suppose the price per share is $2, and the purchase price of a machine is only $1. Obviously the firm should invest—buy a new machine—and finance it by issuing a share. Each machine costs the firm $1 to purchase, but stock market participants are willing to pay $2 for a share corresponding to this machine when it is installed in the firm.

This is an example of the more general argument made by Tobin that there should be a tight relation between the stock market and investment. When deciding whether to invest, he argued firms might not need to go through the type of complicated computation you saw in the text. In effect, the stock price tells firms how much the stock market values each unit of capital already in place. The firm then has a simple problem: Compare the purchase price of an additional unit of capital to the price the stock market is willing to pay for it. *If the stock market value exceeds the purchase price, the firm should buy the machine; otherwise, it should not.*

Tobin then constructed a variable corresponding to the value of a unit of capital in place relative to its purchase price and looked at how closely it moved with investment. He used the symbol q to denote the variable, and the variable has become known as **Tobin's q**. Its construction is as follows:

1. Take the total value of US corporations as assessed by financial markets. That is, compute the sum of their stock market value (the price of a share times the number of shares). Compute also the total value of their bonds outstanding (remember that firms finance themselves not only through stocks but also through bonds). Add together the value of stocks and bonds. Subtract the firms' financial assets, the value of the cash, bank accounts, and any bonds the firms might hold.
2. Divide this total value by the value of the capital stock of US corporations at replacement cost (the price firms would have to pay to replace their machines, their plants, and so on).

The ratio gives us, in effect, the value of a unit of capital in place relative to its current purchase price. This ratio is *Tobin's q*. Intuitively, the higher q, the higher the value of capital relative to its current purchase price, and the higher investment should be. (In the example at the start of the box, Tobin's q is equal to 2, so the firm should definitely invest.)

How tight is the relation between Tobin's q and investment? The answer is given in Figure 1, which plots two variables for each year since 1960 for the United States.

Measured on the left vertical axis is the change in the ratio of investment to capital for nonfinancial US corporations. Measured on the right vertical axis is the change of Tobin's q. This variable is lagged once. For 2000, for example, the figure shows the change in the ratio of investment to capital for 2000, and the change in Tobin's q for 1999—that is, a year earlier. The reason for presenting the two variables this way is that the strongest relation in the data appears to be between investment this year and

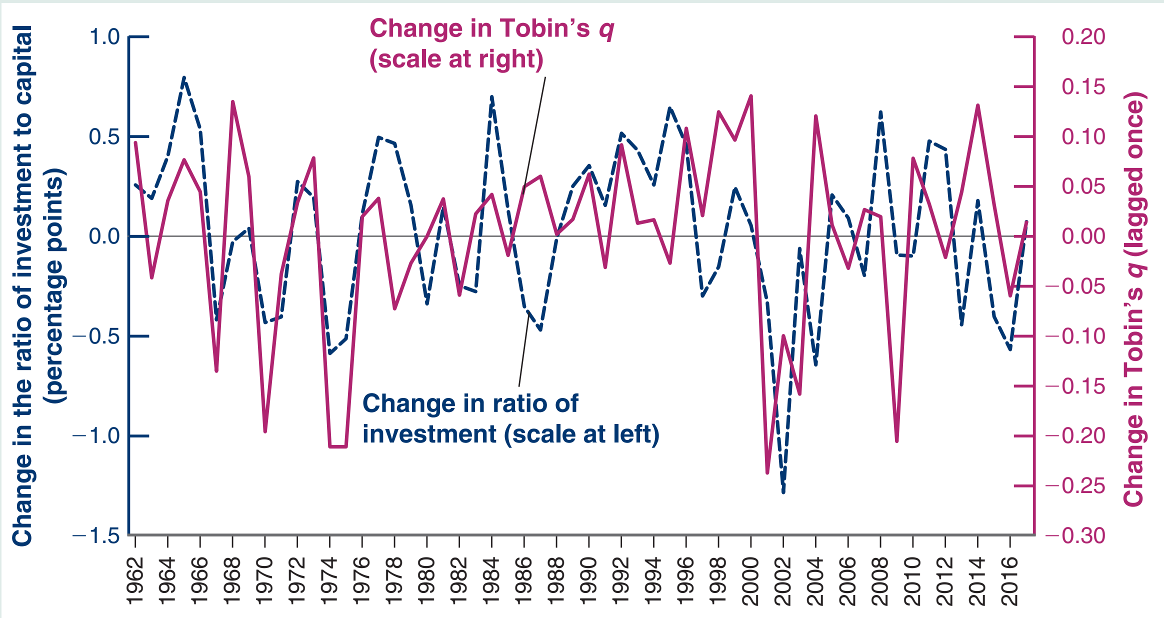


Figure 1

Tobin's q versus the Ratio of Investment to Capital. Annual Rates of Change since 1962

Source: Flow of Funds, Table s5a. Numerator of q: Market value of equity + Debt + Loans – Financial Assets, nonfinancial US corporations. Denominator: Nonfinancial assets of nonfinancial US corporations.

Tobin's q last year. Put another way, movements in investment this year are more closely associated with movements in the stock market last year rather than with movements in the stock market this year; a plausible explanation is that it takes time for firms to make investment decisions, build new factories, and so on.

The figure shows that there is a clear relation between Tobin's q and investment. This is not because firms blindly follow the signals from the stock market, but because investment decisions and stock market prices depend very much on the same factors—expected future profits and expected future interest rates.

The present value of expected profits is simply the ratio of the profit rate—that is, profit per unit of capital—to the sum of the real interest rate and the depreciation rate. Replacing (15.5) in equation (15.4), investment is given by:

$$I_t = I\left(\frac{\Pi_t}{r_t + \delta}\right) \tag{15.6}$$

Investment is a function of the ratio of the profit rate to the sum of the interest rate and the depreciation rate.

The sum of the real interest rate and the depreciation rate is called the **user cost** or the **rental cost** of capital. To see why, suppose the firm, instead of buying the machine, rented it from a rental agency. How much would the rental agency have to charge per year? Even if the machine did not depreciate, the agency would have to ask for an interest charge equal to r_t times the price of the machine (we have assumed the price of a machine to be 1 in real terms, so r_t times 1 is just r_t). The agency has to get at least as much from buying and renting out the machine as it would from, say, buying bonds. In addition, the rental agency would have to charge for depreciation, δ, times the price of the machine, 1. Therefore:

Such arrangements exist. For example, many firms lease cars and trucks from leasing companies.

$$\text{Rental cost} = (r_t + \delta)$$

Even though firms typically do not rent the machines they use, $(r_t + \delta)$ still captures the firms' implicit cost—sometimes called the *shadow cost*—of using the machine for one year.

The investment function given by equation (15.6) then has a simple interpretation. *Investment depends on the ratio of profit to the user cost. The higher the profit, the higher the level of investment. The higher the user cost, the lower the level of investment.*

This relation between profit, the real interest rate, and investment hinges on a strong assumption: that the future is expected to be the same as the present. It is a useful relation to remember—and one that macroeconomists keep handy in their toolbox. It is time, however, to relax this assumption and return to the role of expectations in determining investment decisions.

Current versus Expected Profit

The theory we have developed implies that investment should be forward looking and should depend primarily on *expected future profits*. (Under our assumption that it takes a year for investment to generate profits, current profit does not even appear in equation (15.3).)

One striking empirical fact about investment, however, is how strongly it moves with fluctuations in *current profit*. This relation is shown in Figure 15-3, which plots yearly changes in investment and profit since 1960 for the US economy. Profit is constructed as the ratio of the sum of *after-tax profits plus interest payments paid by US nonfinancial corporations*, divided by their capital stock. Investment is constructed as the ratio of investment by US nonfinancial corporations to their capital stock. Profit is lagged once. For 2000, for example, the figure shows the change in investment for 2000, and the change in profit

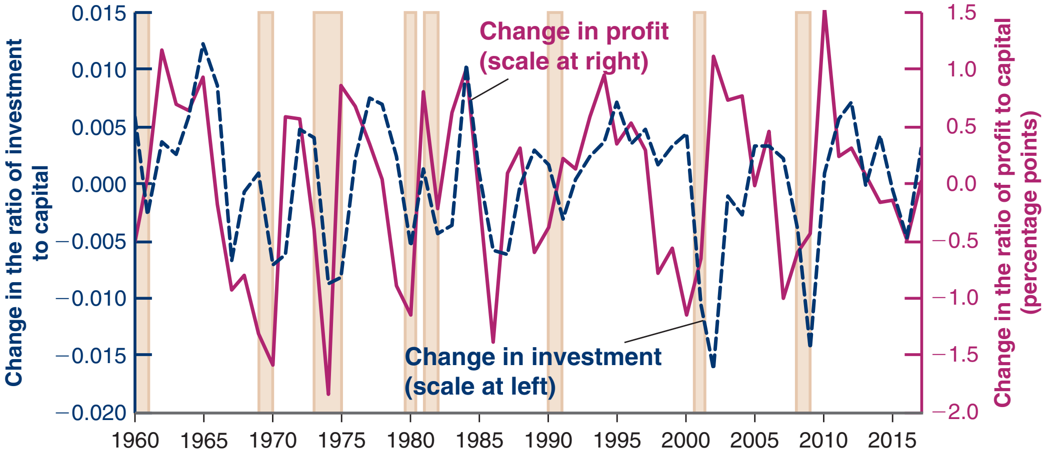


Figure 15-3

Changes in Investment and Profit in the United States since 1960

Investment and profit move very much together.

Source: Gross investment: Federal Reserve Board, Flow of Funds, series FA105013005.A. Capital stock: BEA Fixed Assets Tables, net stock of private nonresidential fixed assets, nonfinancial. Profit: BEA, NIPA Table 1.14, Net operating surplus minus taxes, minus transfers, minus net interest payments.

for 1999—that is, a year earlier. The reason for presenting the two variables this way is that the strongest relation in the data appears to be between investment in a given year and profit the year before—a lag plausibly due to the fact that it takes time for firms to decide on new investment projects in response to higher profit. The shaded areas in the figure represent years in which there was a recession—a decline in output for at least two consecutive quarters of the year.

There is a clear positive relation between changes in investment and changes in current profit in Figure 15-3. Is this relation inconsistent with the theory we have just developed, which holds that investment should be related to the present value of expected future profits rather than to current profit? Not necessarily. If firms expect future profits to move very much like current profit, then the present value of those future profits will move very much like current profit, and so will investment.

Economists who have looked at the question more closely have concluded, however, that the effect of current profit on investment is stronger than would be predicted by the theory developed so far. How they have gathered some of the evidence is described in the Focus Box "Profitability versus Cash Flow." On the one hand, some firms with highly profitable investment projects but low current profits appear to be investing too little. On the other hand, some firms that have high current profit appear sometimes to invest in projects of doubtful profitability. In short, current profit appears to affect investment, even after controlling for the expected present value of profits.

Why does current profit play a role in the investment decision? The answer parallels our discussion of consumption in Section 15-1, where we discussed why consumption depends directly on current income. Some of the reasons we used to explain the behavior of consumers also apply to firms:

- If its current profit is low, a firm that wants to buy new machines can get the funds it needs only by borrowing. But it may be reluctant to borrow. Although expected profits might look good, things may turn bad, leaving the firm unable to repay the debt. If current profit is high, however, the firm might be able to finance its investment just by retaining some of its earnings and without having to borrow. The bottom line is that higher current profit may lead the firm to invest more.

FOCUS

Profitability versus Cash Flow

How much does investment depend on the expected present value of future profits, and how much does it depend on current profit? In other words: Which is more important for investment decisions: **Profitability** (the expected present discounted value of future profits) or **cash flow** (current profit, the net flow of cash the firm is receiving now)?

The difficulty in answering this question is that, most of the time, cash flow and profitability move together. Firms that do well typically have both large cash flows and good future prospects. Firms that suffer losses often have poor future prospects.

The best way to isolate the effects of cash flow and profitability on investment is to identify times or events when cash flow and profitability move in different directions, and then look at what happens to investment. This is the approach taken by Owen Lamont, an economist at Harvard University. An example will help you understand Lamont's strategy.

Think of two firms, A and B. Both are involved in steel production. Firm B is also involved in oil exploration.

Suppose there is a sharp drop in the price of oil, leading to losses in oil exploration. This shock decreases firm B's cash flow. If the losses in oil exploration are large enough to offset the profits from steel production, firm B might even show an overall loss.

As a result of the drop in the price of oil, will firm B invest less in its steel operation than firm A does? If only the profitability of steel production matters, there is no reason for firm B to invest less in its steel operation than firm A. But if current cash flow also matters, the fact that firm B has a lower cash flow may prevent it from investing as much as firm A in its steel operation. Looking at investment in the steel operations of the two firms can tell us how much investment depends on cash flow versus profitability.

This is the empirical strategy followed by Lamont. He focused on what happened in 1986 when the price of oil in the United States dropped by 50%, leading to large losses in oil-related activities. He then looked at whether firms that had substantial oil activities cut investment in their nonoil activities relatively more than other firms in the same nonoil activities. He concluded that they did. He found that for every \$1 decrease in cash flow as a result of the decrease in the price of oil, investment spending in nonoil activities was reduced by 10 to 20 cents. In short: Current cash flow matters.[2]

- A firm that wants to invest might have difficulty borrowing. Potential lenders may not be convinced the project is as good as the firm says it is, and they may worry the firm will be unable to repay. If the firm has large current profits, it does not have to borrow and so does not need to convince potential lenders. It can proceed and invest as it pleases and is more likely to do so.

In summary, to fit the investment behavior we observe in practice, the investment equation is better specified as

$$I_t = I\,[V(\Pi_t^e),\ \Pi_t] \qquad (15.7)$$
$$(\ +\quad ,\ +\ )$$

In words: *Investment depends both on the expected present value of future profits and on the current level of profit.*

Profit and Sales

Let's take stock of where we are. We have argued that investment depends both on current profit and on expected profit or, more specifically, on current and expected profit per unit of capital. We need to take one last step. What determines profit per unit of capital? Primarily two factors: (1) the level of sales and (2) the existing capital stock. If sales are low relative to the capital stock, profits per unit of capital are likely to be low as well.

[2]Source: *Owen Lamont, "Cash Flow and Investment: Evidence from Internal Capital Markets,"* Journal of Finance, *1997, 52(1): pp. 83–109.*

Let's write this more formally. Ignore the distinction between sales and output, and let Y_t denote output (equivalently, sales). Let K_t denote the capital stock at time t. Our discussion suggests the following relation:

$$\Pi_t = \Pi\left(\frac{Y_t}{K_t}\right) \quad (15.8)$$
$$(+)$$

Profit per unit of capital is an increasing function of the ratio of sales to the capital stock. For a given capital stock, the higher the sales, the higher the profit per unit of capital. For given sales, the higher the capital stock, the lower the profit per unit of capital.

How well does this relation hold in practice? Figure 15-4 plots yearly changes in profit per unit of capital (measured on the right vertical axis) and changes in the ratio of output to capital (measured on the left vertical axis) for the United States since 1960. As in Figure 15-3, profit per unit of capital is defined as the sum of after-tax profits plus interest payments by US nonfinancial corporations, divided by their capital stock measured at replacement cost. The ratio of output to capital is constructed as the ratio of GDP to the aggregate capital stock.

Figure 15-4 shows that there is a strong relation between changes in profit per unit of capital and changes in the ratio of output to capital. Given that most of the year-to-year changes in the ratio of output to capital come from movements in output, and most of the year-to-year changes in profit per unit of capital come from movements in profit (capital moves slowly over time; the reason is that capital is large compared to yearly investment, so even large movements in investment lead to small changes in the capital stock), we can state the relation as follows: Profit decreases in recessions (shaded areas are periods of recession) and increases in expansions.

Why is this relation between output and profit relevant here? Because it implies a link between *current output and expected future output*, on the one hand, and *investment*, on the other. Current output affects current profit, expected future output affects expected future profit, and current and expected future profits affect investment. For example, anticipation of a sustained economic expansion leads firms to expect high profits, now

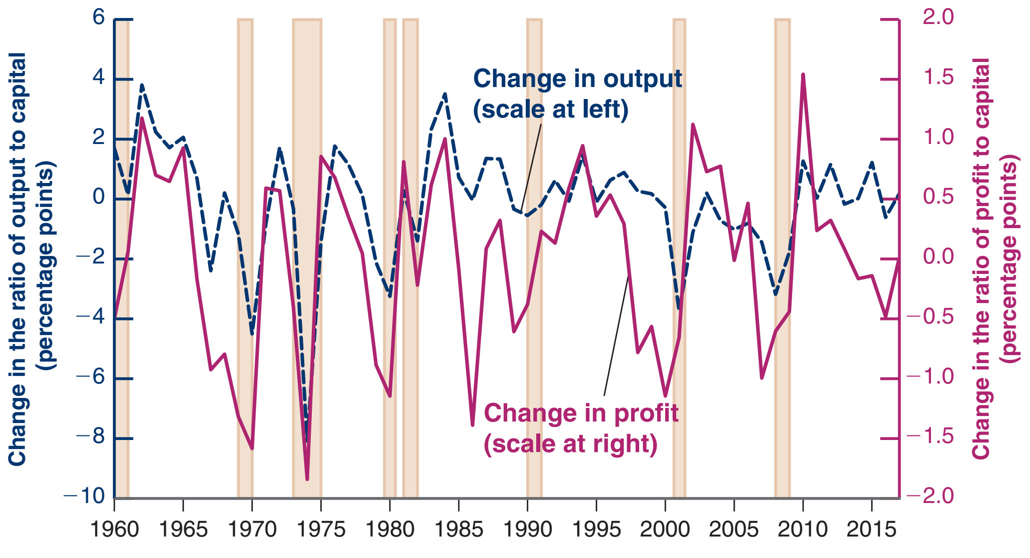


Figure 15-4

Changes in Profit per Unit of Capital versus Changes in the Ratio of Output to Capital in the United States since 1960

Profit per unit of capital and the ratio of output to capital move largely together.

Source: Capital stock: BEA Fixed Assets Tables. Net stock of private nonresidential fixed assets, nonfinancial assets; Profit: BEA, NIPA Table 1.14, net operating surplus minus taxes minus transfers minus net interest and miscellaneous payments; Output: BEA, Gross value added of nonfinancial corporate business sector.

and for some time in the future. These expectations in turn lead to higher investment. The effect of current and expected output on investment, together with the effect of investment on demand and output, will play a crucial role when we return to the determination of output in Chapter 16.

High expected output ⇒ High expected profit ⇒ High investment today.

15-3 THE VOLATILITY OF CONSUMPTION AND INVESTMENT

You will surely have noticed the similarities between our treatment of consumption and of investment behavior in Sections 15-1 and 15-2:

- Whether consumers perceive current movements in income to be transitory or permanent affects their consumption decisions. The more transitory they expect a current increase in income to be, the less they will increase their consumption.
- In the same way, whether firms perceive current movements in sales to be transitory or permanent affects their investment decisions. The more transitory they expect a current increase in sales to be, the less they revise their assessment of the present value of profits, and thus the less likely they are to buy new machines or build new factories. This is why, for example, the boom in sales that happens every year between Thanksgiving and Christmas does not lead to a boom in investment every December. Firms understand that this boom is transitory.

In the United States, retail sales are 24% higher on average in December than in other months. In France and Italy, sales are 60% higher in December.

But there are also important differences between consumption decisions and investment decisions:

- The theory of consumption developed previously implies that when consumers perceive an increase in income as permanent, they respond with, *at most*, an equal increase in consumption. The permanent nature of the increase in income implies that they can afford to increase consumption now and in the future by the same amount as the increase in income. Increasing consumption more than one-for-one would require cuts in consumption later, and there is no reason for consumers to want to plan consumption this way.
- Now consider the behavior of firms faced with an increase in sales they believe to be permanent. The present value of expected profits increases, leading to an increase in investment. In contrast to consumption, however, this does not imply that the increase in investment should be at most equal to the increase in sales. Rather, once a firm has decided that an increase in sales justifies the purchase of a new machine or the building of a new factory, it may want to proceed quickly, leading to a large but short-lived increase in investment spending. This increase in investment spending may exceed the increase in sales.

More concretely, take a firm that has a ratio of capital to its annual sales of, say, 3. An increase in sales of $10 million this year, if expected to be permanent, requires the firm to spend $30 million on additional capital if it wants to maintain the same ratio of capital to output. If the firm buys the additional capital right away, the increase in investment spending this year will equal *three times* the increase in sales. Once the capital stock has adjusted, the firm will return to its normal pattern of investment. This example is extreme because firms do not adjust their capital stock right away. But even if they adjust their capital stock over a few years, the increase in investment might still exceed the increase in sales for a while.

We can tell the same story in terms of equation (15.8). Because we make no distinction here between output and sales, the initial increase in sales leads to an equal increase in output, Y, so that Y/K—the ratio of the firm's output to its existing capital stock—also increases. The result is higher profit, which leads the firm to undertake more investment. Over time, the higher level of investment leads to a higher capital stock, K, so that Y/K decreases back to normal. Profit per unit of capital returns to normal and so does investment. Thus, in response to a permanent increase in sales, investment may increase a lot initially and then return to normal over time.

These differences suggest that investment should be more volatile than consumption. How much more? The answer is given in Figure 15-5, which plots yearly rates of change in US consumption and investment since 1960. The shaded areas are years during which the US economy was in recession. To make the figure easier to interpret, both rates of change are plotted as deviations from the average rate of change over the period, so that they are, on average, equal to zero.

Figure 15-5 yields three conclusions:

- Consumption and investment usually move together. Recessions, for example, are typically associated with decreases in *both* investment and consumption. Given our discussion, which has emphasized that consumption and investment depend largely on the same determinants, this should not come as a surprise.
- Investment is much more volatile than consumption. Relative movements in investment range from −29% to +24%, whereas relative movements in consumption range only from −5% to +3%.
- Because, however, the level of investment is much smaller than the level of consumption (recall that investment accounts for about 15% of GDP, whereas consumption accounts for close to 70%), changes in investment from one year to the next end up being of the same overall magnitude as changes in consumption. In other words, both components contribute roughly equally to fluctuations in output over time.

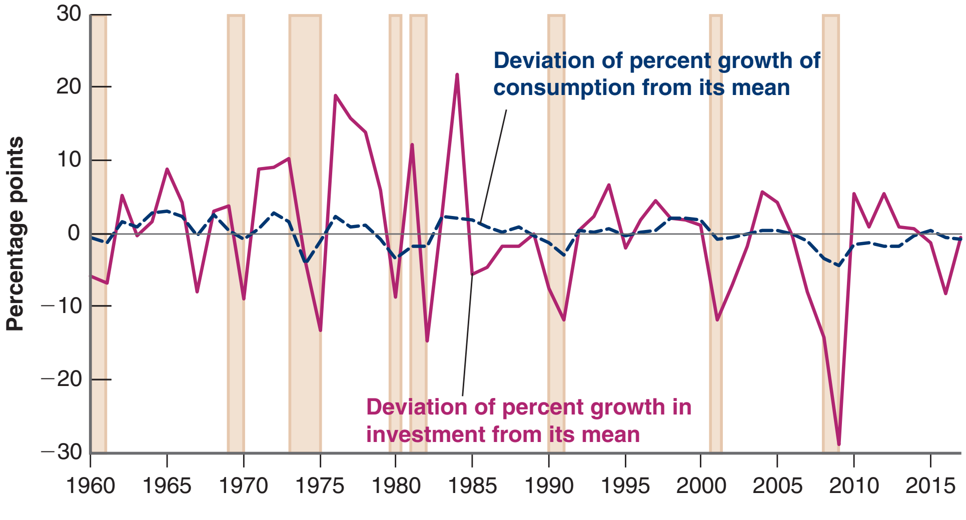


Figure 15-5

Rates of Change in Consumption and Investment in the United States since 1960

Relative movements in investment are much larger than relative movements in consumption.

Source: FRED, series PCECC96, GPDI

SUMMARY

- Consumption depends on both wealth and current income. Wealth is the sum of nonhuman wealth (financial wealth and housing wealth) and human wealth (the present value of expected after-tax labor income).
- The response of consumption to changes in income depends on whether consumers perceive these changes as transitory or permanent.
- Consumption is likely to respond less than one-for-one to movements in income. Consumption might move even if current income does not change.
- Investment depends on both current profit and the present value of expected future profits.
- Under the simplifying assumption that firms expect profits and interest rates to be the same in the future as they are today, we can think of investment as depending on the ratio of profit to the user cost of capital, where the user cost is the sum of the real interest rate and the depreciation rate.
- Movements in profit are closely related to movements in output. Hence, we can think of investment as depending indirectly on current and expected future output movements. Firms that anticipate a long output expansion, and thus a long sequence of high profits, will invest. Movements in output that are not expected to last will have a small effect on investment.
- Investment is much more volatile than consumption. But because investment accounts only for 15% of GDP and consumption accounts for 70%, movements in investment and consumption are of roughly equal importance in accounting for movements in aggregate output.

KEY TERMS

permanent income theory of consumption, 314
life-cycle theory of consumption, 314
financial wealth, 314
housing wealth, 314
human wealth, 314
nonhuman wealth, 314
total wealth, 314
panel datasets, 315
static expectations, 322
Tobin's *q*, 323
user cost, 324
rental cost, 324
profitability, 326
cash flow, 326

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. For a typical college student, human wealth and nonhuman wealth are approximately equal.
b. Natural experiments, such as retirement, do not suggest that expectations of future income are a major factor affecting consumption.
c. At the start of the financial crisis, expected future income growth fell.
d. Buildings and factories depreciate much faster than machines.
e. A high value for Tobin's *q* indicates that the stock market believes that capital is overvalued, and thus investment should be lower.
f. Unless current profit affects expectations of future profit, it should have no impact on investment.
g. Data from the past three decades in the United States suggest that corporate profits are closely tied to the business cycle.
h. Changes in consumption and investment typically occur in the same direction and are roughly of the same magnitude.

2. A consumer has nonhuman wealth equal to $100,000. She earns $40,000 this year and expects her salary to increase by 5% in real terms each year for the following two years. She will then retire. The real interest rate is equal to 0% and is expected to remain at 0% in the future. Labor income is taxed at a rate of 25%.

a. What is this consumer's human wealth?
b. What is her total wealth?
c. If she expects to live for seven more years after retiring and wants her consumption to remain the same (in real terms) every year from now on, how much can she consume this year?
d. If she received a bonus of $20,000 in the current year only, with all future salary payments remaining as stated earlier, by how much could this consumer increase consumption now and in the future?
e. Suppose now that at retirement, Social Security will start paying benefits each year equal to 60% of this consumer's earnings during her last working year. Assume that benefits are not taxed. How much can she consume this year and still maintain constant consumption over her lifetime?

3. A pretzel manufacturer is considering buying another pretzel-making machine that costs $100,000. The machine will depreciate *by 8% per year. It will generate real profits equal to \$18,000 next year, \$18,000 (1 – 8%) two years from now (that is, the same real profits but adjusted for depreciation), \$18,000 $(1 - 8\%)^2$ three years from now, and so on. Determine whether the manufacturer should buy the machine if the real interest rate is assumed to remain constant at each of the values in parts a through c.*

a. 5%
b. 10%
c. 15%

4. Suppose that at age 22, you have just finished college and have been offered a job with a starting salary of \$40,000. Your salary will remain constant in real terms. However, you have also been admitted to a professional school. The school can be completed in two years. Upon graduation, you expect your starting salary to be 10% higher in real terms and to remain constant in real terms thereafter. The tax rate on labor income is 40%.

a. If the real interest rate is zero and you expect to retire at age 60 (i.e., if you do not go to professional school, you expect to work for 38 years total), what is the maximum you should be willing to pay in tuition to attend this professional school?
b. What is your answer to part a if you expect to pay 30% in taxes?

DIG DEEPER

5. Individual saving and aggregate capital accumulation

Suppose that every consumer is born with zero financial wealth and lives for three periods: youth, middle age, and old age. Consumers work in the first two periods and retire in the last one. Their income is \$5 in the first period, \$25 in the second, and \$0 in the last one. Inflation and expected inflation are equal to zero, and so is the real interest rate.

a. What is the present discounted value of labor income at the beginning of the first period of life? What is the highest sustainable level of consumption such that consumption is equal in all three periods?
b. For each age group, what is the amount of saving that allows consumers to maintain the constant level of consumption you found in part a? (*Hint*: Saving can be a negative number if the consumer needs to borrow to maintain a certain level of consumption.)
c. Suppose there are *n* people born each period. What is total saving in the economy? (*Hint*: Add up the saving of each age group. Remember that some age groups may have negative saving.) Explain.
d. What is total financial wealth in the economy? (*Hint*: Compute the financial wealth of people at the beginning of the first period of life, the beginning of the second period, and the beginning of the third period. Add the three numbers. Remember that people can be in debt, so financial wealth can be negative.)

6. Borrowing constraints and aggregate capital accumulation

Continue with the setup from Problem 5, but suppose now that borrowing restrictions do not allow young consumers to borrow. If we call the sum of income and total financial wealth “cash on hand,” then the borrowing restriction means that consumers cannot consume more than their cash on hand. In each age group, consumers compute their total wealth and then determine their desired level of consumption as the highest level that allows their consumption to be equal in all three periods. However, if at any time, desired consumption exceeds cash on hand, then consumers are constrained to consume exactly their cash on hand.

a. Calculate consumption in each period of life. Compare this answer to your answer to part a of Problem 5 and explain any differences.
b. Calculate total saving for the economy. Compare this answer to your answer to part c of Problem 5 and explain any differences.
c. Derive total financial wealth for the economy. Compare this answer to your answer to part d of Problem 5 and explain any differences.
d. Consider the following statement: “Financial liberalization may be good for individual consumers, but it is bad for overall capital accumulation.” Discuss.

7. Saving with uncertain future income

Consider a consumer who lives for three periods: youth, middle age, and old age. When young, the consumer earns \$20,000 in labor income. Earnings during middle age are uncertain; there is a 50% chance that the consumer will earn \$40,000 and a 50% chance that the consumer will earn \$100,000. When old, the consumer spends savings accumulated during the previous periods. Assume that inflation, expected inflation, and the real interest rate equal zero. Ignore taxes for this problem.

a. What is the expected value of earnings in the middle period of life? Given this number, what is the present discounted value of expected lifetime labor earnings? If the consumer wishes to maintain constant expected consumption over her lifetime, how much will she consume in each period? How much will she save in each period?
b. Now suppose the consumer wishes, above all else, to maintain a minimum consumption level of \$20,000 in each period of her life. To do so, she must consider the worst outcome. If earnings during middle age turn out to be \$40,000, how much should the consumer spend when she is young to guarantee consumption of at least \$20,000 in each period? How does this level of consumption compare to the level you obtained for the young period in part a?
c. Given your answer in part b, suppose that the consumer’s earnings during middle age turn out to be \$100,000. How much will she spend in each period of life? Will consumption be constant over the consumer’s lifetime? (*Hint*: When the consumer reaches middle age, she will try to maintain constant consumption for the last two periods of life, as long as she can consume at least \$20,000 in each period.)
d. What effect does uncertainty about future labor income have on saving (or borrowing) by young consumers?

EXPLORE FURTHER

8. The movements of consumption and investment

Go to the FRED database operated by the Federal Reserve Bank of St. Louis. Find annual data for personal consumption expenditures and gross private domestic investment as well as for real GDP. The data are measured in real dollars. Place values starting in 1960 and ending with the most recent year of data in a spreadsheet. (FRED allows you to directly download to a spreadsheet.) As of the time of writing the series names are: Real GDP (Chained 2012 dollars), GDPCA; Real Personal Consumption Expenditures (Chained 2012 dollars), PCECCA; Real Gross Private Domestic Investment (Chained 2012 dollars), GPDICA. You should be able to search these names but be careful to download the levels of these variables at an annual rate. Pay attention to whether the variables are measured in millions or billions of dollars.

a. On average, how much larger is consumption than investment? Calculate both as a percent of GDP.
b. Compute the change in the levels of consumption and investment from one year to the next, and graph them for the period 1961 to the latest available date. Are the year-to-year changes in consumption and investment of similar magnitude?
c. Compute the annual percentage change in real consumption and real investment from 1961. Which is more volatile?

9. Consumer confidence, disposable income, and recessions

Go to the Web site of the FRED economic data base and download the quarterly series for real personal disposable income per capita (series name A229RX0Q048SBEA (billions of chained 2012 dollars)), the University of Michigan Survey of Consumers Index of Consumer Sentiment (series UMCSENT)). The Consumer Sentiment data are monthly and you will need to create a quarterly average from the monthly observations by averaging. We will use this data series as our measure of consumer confidence. The monthly observations start in January 1978, so begin your study at that date.

a. Before you look at the data, can you think of any reason to expect consumer confidence to be related to disposable income? Can you think of reasons why consumer confidence would be unrelated to disposable income?
b. Plot the level of the index of consumer sentiment against the growth rate of disposable income per person. Is the relationship positive?
c. Plot the change in the index of consumer sentiment against the growth rate of disposable income per person. What does that relation look like? Focus on observations where the change in disposable income is less than 0.2% in absolute value. Is the level of consumer sentiment changing? How would we interpret such observations?
d. Focus on the years 2007, 2008, and 2009. How does the behavior of consumer sentiment from 2007 to 2008 compare to the usual behavior in consumer sentiment? Why? (*Hint*: The bankruptcy of Lehmann Brothers occurred in September 2008.) Does the fall in consumer sentiment anticipate the decline in real personal disposable income that accompanied the crisis?
e. Donald Trump was rather unexpectedly elected President in November 2017. Is there any evidence of a change in consumer confidence in the first quarter of 2017?
f. There was a large tax cut, signed into law on December 22, 2017 and implemented in 2018. Is there a substantial change in personal disposable income per capita from 2017 to 2018? Is there a change in the index of consumer sentiment between 2017 and 2018?

APPENDIX: Derivation of the Expected Present Value of Profits under Static Expectations

You saw in the text (equation (15.3)) that the expected present value of profits is given by

$$V(\Pi_t^e) = \frac{1}{1 + r_t}\Pi_{t+1}^e + \frac{1}{(1 + r_t)(1 + r_{t+1}^e)}(1 - \delta)\,\Pi_{t+2}^e + \cdots$$

If firms expect both future profits (per unit of capital) and future interest rates to remain at the same level as today, so that $\Pi_{t+1}^e = \Pi_{t+2}^e = \cdots = \Pi_t$ and $r_{t+1}^e = r_{t+2}^e = \cdots = r_t$, the equation becomes

$$V(\Pi_t^e) = \frac{1}{1 + r_t}\,\Pi_t + \frac{1}{(1 + r_t)^2}\,(1 - \delta)\,\Pi_t + \cdots$$

Factoring out $[1/(1 + r_t)]\Pi_t$,

$$V(\Pi_t^e) = \frac{1}{1 + r_t}\,\Pi_t\left(1 + \frac{1 - \delta}{1 + r_t} + \cdots\right) \qquad (15.A1)$$

The term in parentheses in this equation is a geometric series of the form $1 + x + x^2 + \cdots$. So, from Proposition 2 in Appendix 2 at the end of the book,

$$(1 + x + x^2 + \cdots) = \frac{1}{1 - x}$$

Here x equals $(1 - \delta)/(1 + r_t)$, so

$$\left(1 + \frac{1 - \delta}{1 + r_t} + \left(\frac{1 - \delta}{1 + r_t}\right)^2 + \cdots\right)$$

$$= \frac{1}{1 - (1 - \delta)/(1 + r_t)} = \frac{1 + r_t}{r_t + \delta}$$

Replacing the term in parentheses in equation (15.A1) with the expression above and manipulating gives:

$$V(\Pi_t^e) = \frac{1}{1 + r_t}\,\frac{1 + r_t}{r_t + \delta}\,\Pi_t$$

Simplifying gives equation (15.5) in the text:

$$V(\Pi_t^e) = \frac{\Pi_t}{(r_t + \delta)}$$

This page is intentionally left blank

16

Expectations, Output, and Policy

In Chapter 14, we saw how expectations affected asset prices, from bonds to stocks to houses. In Chapter 15, we saw how expectations affected consumption decisions and investment decisions. In this chapter we put the pieces together and take another look at the effects of monetary and fiscal policy.

Section 16-1 draws the major implication of what we have learned, namely that expectations of both future output and future interest rates affect current spending, and therefore current output.

Section 16-2 looks at monetary policy. It shows that the effects of monetary policy depend crucially on how changes in the policy rate today lead people and firms to change their expectations of future interest rates and future income and, by implication, to change their spending decisions.

Section 16-3 turns to fiscal policy. It shows how, in contrast to the simple model you saw in the core, a fiscal contraction can sometimes lead to an increase in output, even in the short run. Again, how expectations respond to policy is at the center of the story.

If you remember one basic message from this chapter, it should be: The effects of monetary and fiscal policy depend very much on how they affect expectations. ▶▶▶

16-1 EXPECTATIONS AND DECISIONS: TAKING STOCK

Let's start by reviewing what we have learned, and then discuss how we should modify the characterization of goods and financial markets—the IS-LM model—developed in the core.

Expectations, Consumption, and Investment Decisions

The theme of Chapter 15 was that both consumption and investment decisions depend very much on expectations of future income and interest rates. The channels through which expectations affect consumption and investment spending are summarized in Figure 16-1.

Note the many channels through which expected future variables affect current decisions, both directly and through asset prices:

- An increase in current and expected future after-tax real labor income or a decrease in current and expected future real interest rates increase human wealth (the expected present discounted value of after-tax real labor income), which in turn leads to an increase in consumption.
- An increase in current and expected future real dividends or a decrease in current and expected future real interest rates increases stock prices, which leads to an increase in nonhuman wealth and, in turn, an increase in consumption.
- A decrease in current and expected future nominal interest rates leads to an increase in bond prices, which leads to an increase in nonhuman wealth and, in turn, an increase in consumption.
- An increase in current and expected future real after-tax profits or a decrease in current and expected future real interest rates increases the present value of real after-tax profits, which leads, in turn, to an increase in investment.

Note that in the case of bonds, it is nominal rather than real interest rates that matter because bonds are claims to future dollars rather than to future goods.

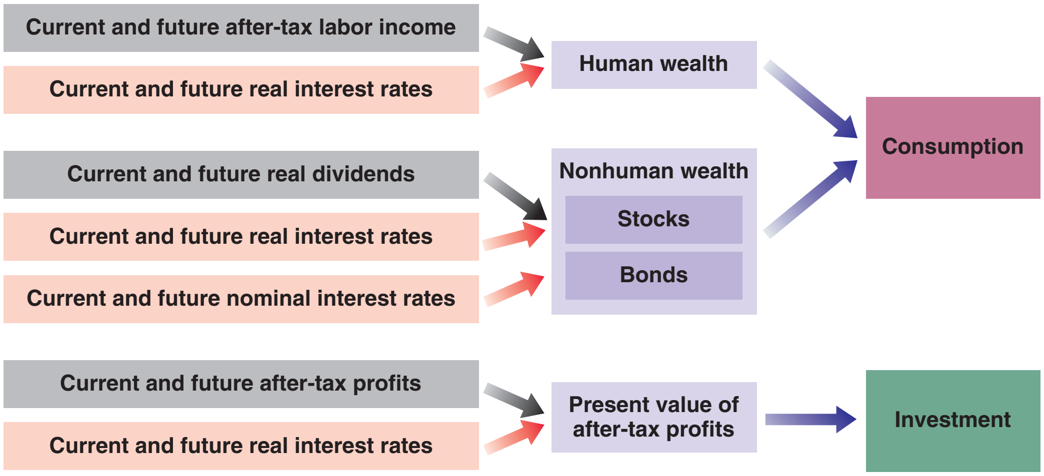
Expectations and the IS Relation

A model that gave a detailed treatment of consumption and investment along the lines suggested in Figure 16-1 would be complicated. It can be done, and indeed it is done in the large models that macroeconomists build to understand the economy and analyze

Figure 16-1

Expectations and Private Spending: The Channels

Expectations affect consumption and investment decisions both directly and through asset prices.



policy; but this is not the place for such complications. We want to capture the essence of what you have learned so far, how consumption and investment depend on expectations of the future, without getting lost in the details.

To do so, let's make a major simplification. Let's reduce the present and the future to only two periods: (1) a *current* period, which you can think of as the current year, and (2) a *future* period, which you can think of as all future years lumped together. This way we do not have to keep track of expectations about each future year.

This way of dividing time between "today" and "later" is the way many of us organize our own lives. Think of "things to do today" versus "things that can wait." You can also think of the future period as the combination of the "medium run" and the "long run" we saw in the core.

Having made this assumption, the question becomes: How should we write the IS relation for the current period? In Chapter 6 (equation (6.5)), we wrote the following equation for the IS relation:

$$Y = C(Y - T) + I(Y, r + x) + G$$

We assumed that consumption depended only on current income, and that investment depended only on current output and the current borrowing rate, equal to the policy rate plus a risk premium. We now want to modify this to take into account how expectations affect both consumption and investment. We proceed in two steps:

First, we rewrite the equation in more compact form without changing its content. For that purpose, let's define aggregate private spending as the sum of consumption and investment spending.

The reason for doing so is to group together the two components of demand, C and I, which both depend on expectations.

$$A(Y, T, r, x) \equiv C(Y - T) + I(Y, r + x)$$

where A stands for **aggregate private spending**, or simply, **private spending**. With this notation we can rewrite the IS relation as

$$Y = A(\underset{(+,-,-,-)}{Y, T, r, x}) + G \qquad (16.1)$$

The properties of aggregate private spending, A, follow from the properties of consumption and investment that we derived in previous chapters: Aggregate private spending is

- An increasing function of income Y: Higher income (equivalently, output) increases both consumption and investment.
- A decreasing function of taxes T: Higher taxes decrease consumption.
- A decreasing function of the real policy rate r: A higher real policy rate decreases investment.
- A decreasing function of the risk premium x: A higher risk premium increases the borrowing rate and decreases investment.

The first step only simplified notation. The second step is to modify equation (16.1) to take into account the role of expectations. Because the focus in this chapter is on expectations rather than on the risk premium, I shall assume that it is constant, and so to save on notation, I shall ignore it in the model. With the focus on expectations, the natural extension of equation (16.1) is to allow spending to depend not only on current variables but also on their expected values in the future.

Introducing uncertainty and risk formally in the model would make it too heavy. I give an informal discussion in a Focus Box at the end of the chapter called "Uncertainty and Fluctuations."

$$Y = A(\underset{(+,-,-,+,\ -,\ -)}{Y, T, r, Y'^e, T'^e, r'^e}) + G \qquad (16.2)$$

Primes denote future values and the superscript e denotes an expectation, so Y'^e, T'^e, and r'^e denote future expected income, future expected taxes, and the future expected real interest rate, respectively. The notation is a bit heavy, but what it captures is straightforward.

Notation: Primes stand for values of the variables in the future period. The superscript e stands for *expected*.

If Y or Y'^e increases $\Rightarrow A$ increases.

If T or T'^e increases $\Rightarrow A$ decreases.

If r or r'^e increases $\Rightarrow A$ decreases.

- Increases in either current or expected future income increase private spending.
- Increases in either current or expected future taxes decrease private spending.
- Increases in either the current or expected future real policy rate decrease private spending.

With the goods market equilibrium now given by equation (16.2), Figure 16-2 shows the new IS curve for the current period. As usual, to draw the curve we take all variables other than current output, Y, and the current real policy rate, r, as given. Thus, the IS curve is drawn for given values of current and future expected taxes, T and T'^e, for given values of expected future output, Y'^e, and for given values of the expected future real policy rate, r'^e.

The new IS curve, based on equation (16.2), is still downward sloping, for the same reason it was in Chapter 6: A decrease in the current policy rate leads to an increase in private spending. This increase in private spending leads, through a multiplier effect, to an increase in output. We can say more, however. The new IS curve is much steeper than the IS curve we drew in Chapter 6. Put another way, *everything else the same*, a large decrease in the current policy rate is likely to have only a small effect on equilibrium output.

To see why the effect is small, take point A on the IS curve in Figure 16-2 and consider the effects of a decrease in the real policy rate, from r_A to r_B. The effect of this decrease on output depends on the strength of the effects of the real policy rate on (1) spending given income and (2) the size of the multiplier.

Let's examine each one:

Suppose you have a 30-year loan and the 1-year interest rate goes down from 5% to 2%. All future 1-year rates remain the same. By how much will the 30-year interest rate come down? (Answer: from 5% to 4.9%. To see why, extend equation (14.11) to the 30-year yield, which is the average of 30 one-year rates.)

- A decrease in the current real policy rate, *given unchanged expectations of the future real policy rate*, does not have much effect on private spending. We saw why in Chapters 14 and 15: A change in only the current real interest rate does not lead to large changes in present values and therefore does not lead to large changes in spending. For example, firms are not likely to change their investment plans very much in response to a decrease in the current real interest rate if they do not expect future real interest rates to be lower as well.

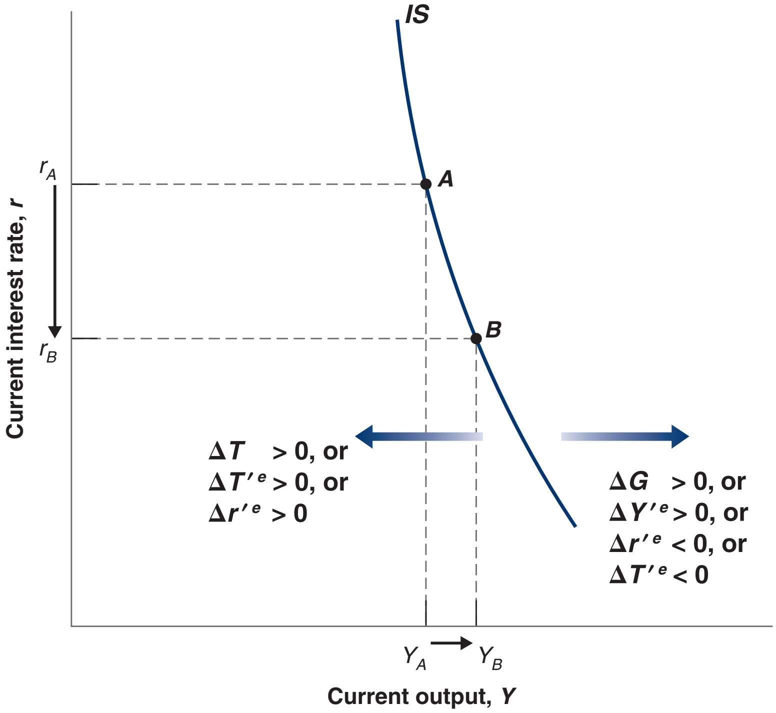


Figure 16-2

The New IS Curve

Given expectations, a de-crease in the real policy rate leads to a small increase in output. The *IS* curve is steeply downward sloping. Increases in government spending or in expected future output shift the *IS* curve to the right. Increases in taxes, in expected future taxes, or in the expected future real policy rate shift the IS curve to the left.

- The multiplier is likely to be small. Recall that the size of the multiplier depends on the size of the effect that a change in current income (output) has on spending. But a change in current income, *given unchanged expectations of future income*, is unlikely to have a large effect on spending. The reason: Changes in income that are not expected to last have only a limited effect on either consumption or investment. Consumers who expect their income to be higher for only a year will increase consumption, but by much less than the increase in their income. Firms that expect sales to be higher for only a year are unlikely to change their investment plans much, if at all.

Suppose the firm where you work decides to give all employees a one-time bonus of $10,000. You do not expect it to happen again. By how much will you increase your consumption this year? (If you need to, look at the discussion of consumption behavior in Chapter 15.)

Putting things together, a large decrease in the current real policy rate—from r_A to r_B in Figure 16-2—leads to only a small increase in output, from Y_A to Y_B. Put another way: The IS curve, which goes through points *A* and *B*, is steeply downward sloping.

Let's look at the effects of the other variables in equation (16.2). A change in any variable in equation (16.2) other than *Y* and *r* shifts the IS curve:

- Changes in current taxes (*T*) or current government spending (*G*) shift the IS curve
  An increase in current government spending increases spending at a given interest rate, shifting the *IS curve* to the right; an increase in taxes shifts the *IS curve* to the left. These shifts are represented in Figure 16-2.
- Changes in expected future variables (Y'^e, T'^e, r'^e) also shift the IS curve
  An increase in expected future output, Y'^e, shifts the IS curve to the right. Higher expected future income leads consumers to feel wealthier and spend more; higher expected future output implies higher expected profits, leading firms to invest more. Higher spending by consumers and firms leads, through the multiplier effect, to higher output. By a similar argument, an increase in expected future taxes leads consumers to decrease their current spending and shifts the IS curve to the left. And an increase in the expected future real policy rate decreases current spending, also leading to a decrease in output and shifting the IS curve to the left. These shifts are also represented in Figure 16-2.

We are now ready to look at the effects of monetary and fiscal policy. This is where the hard work of the two previous chapters will pay off.

16-2 MONETARY POLICY, EXPECTATIONS, AND OUTPUT

The interest rate that the Fed affects directly is the *current real interest rate, r*. So, the LM curve is still given by a horizontal line at the real policy rate chosen by the Fed, called $\bar{r}$. The IS and LM relations are thus given by:

$$IS: \quad Y = A(Y, T, r, Y'^e, T'^e, r'^e) + G \qquad (16.3)$$

$$LM: \quad r = \bar{r} \qquad (16.4)$$

The corresponding IS and LM curves are drawn in Figure 16-3. Equilibrium in goods and financial markets implies that the economy is at point *A*.

Monetary Policy Revisited

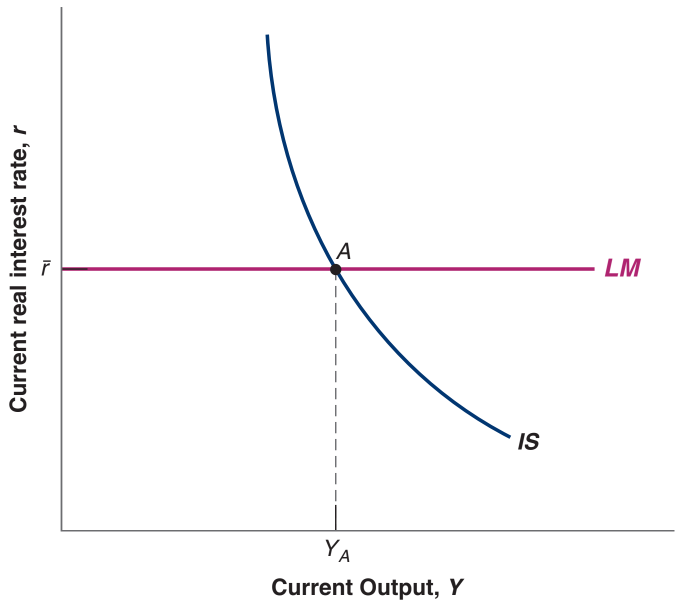
Now suppose the economy is in recession and the Fed decides to lower the real policy rate.

Assume first that this expansionary monetary policy does not change expectations of either the future real policy rate or future output. In Figure 16-4, the LM curve shifts down, from *LM* to *LM''*. (Because I have already used primes to denote future values of the variables,

Figure 16-3

The New IS-LM

The IS curve is steeply downward sloping. Other things being equal, a change in the current interest rate has a small effect on output. Given the current real interest set by the central bank, $\bar{r}$, the equilibrium is at point *A*.



I have to use double primes to denote shifts in curves in this chapter.) The equilibrium moves from point *A* to point *B*, with higher output and a lower real interest rate. The steep IS curve, however, implies that the decrease in the current interest rate has only a small effect on output. Changes in the current interest rate, if not accompanied by changes in expectations, have only a small effect on spending and in turn a small effect on output.

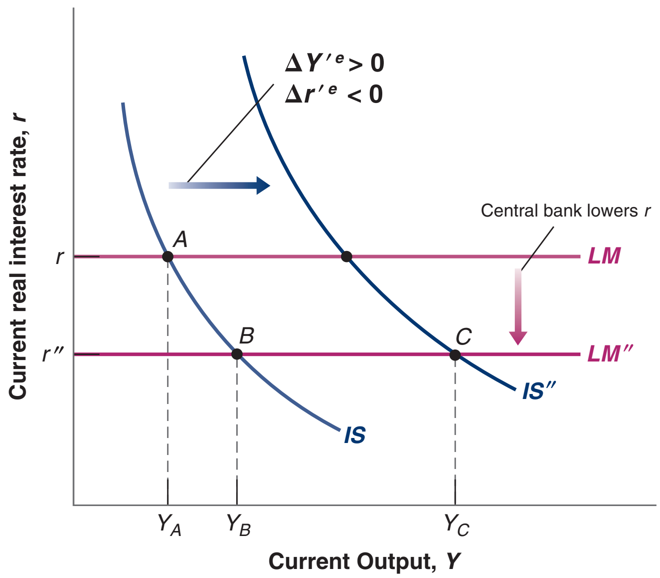
Is it reasonable, however, to assume that expectations are unaffected by an expansionary monetary policy? Isn't it likely that, as the Fed lowers the current real policy rate, financial markets now anticipate lower real interest rates in the future as well, along with higher future output stimulated by the lower future interest rates? What happens if they do anticipate these changes?

At a given current real policy rate, prospects of a lower future real policy rate and of higher future output both increase spending and output, shifting the IS curve to the right, from *IS* to *IS″*. The new equilibrium is given by point *C*. Thus, although the direct effect of the monetary expansion on output is limited, the full effect, once changes in expectations are taken into account, is much larger.

Figure 16-4

The Effects of an Expansionary Monetary Policy

The effects of monetary policy on output depend very much on whether and how monetary policy affects expectations.



Rational Expectations

FOCUS

The importance of expectations is an old theme in macroeconomics. But until the early 1970s, macroeconomists thought of expectations in one of two ways:

- One was as **animal spirits** (from an expression Keynes introduced in the *General Theory* to refer to movements in investment that could not be explained by movements in current variables). In other words, shifts in expectations were considered important but were left largely unexplained.
- The other was as the result of simple, backward-looking rules. For example, people were assumed to have **static expectations**; that is, to expect the future to be like the present (we used this assumption when discussing the Phillips curve in Chapter 8 and when exploring investment decisions in Chapter 15). Or people were assumed to have **adaptive expectations**. If, for example, their forecast of a given variable in a given period turned out to be too low, people were assumed to "adapt" by raising their expectation for the value of the variable for the following period; so seeing an inflation rate higher than they had expected led people to revise upward their forecast of inflation in the future.

In the early 1970s, a group of macroeconomists led by Robert Lucas (at the University of Chicago) and Thomas Sargent (at the University of Minnesota) argued that these assumptions do not do justice to the way people form expectations. (Lucas received the Nobel Prize in 1995; Sargent received the Nobel Prize in 2011.) They argued that, in thinking about the effects of alternative policies, economists should assume that people have rational expectations—that they look into the future and do the best job they can in predicting it. This is not the same as assuming that people know the future, but rather that they use the information they have in the best possible way.

Using the popular macroeconomic models of the time, Lucas and Sargent showed that replacing traditional assumptions about expectation formation with the assumption of rational expectations could fundamentally alter the results. For example, Lucas challenged the notion that disinflation, achieved through tight monetary policy, necessarily led to an increase in unemployment for some time. Under rational expectations, he argued, a credible disinflation policy might be able to decrease inflation without any increase in unemployment. More generally, the research of Lucas and Sargent showed the need for a rethinking of macroeconomic models under the assumption of rational expectations, and this is what happened over the next two decades.

Most macroeconomists today use rational expectations as a working assumption in their models and analyses of policy. It is not because they believe that people truly have rational expectations. Surely there are times when adaptive expectations may be a better description of reality; there are also times when people, firms, or financial market participants lose sight of reality and become too optimistic or too pessimistic. (Recall our discussion of bubbles and fads in Chapter 14.) But, when thinking about the likely effects of a specific economic policy, the best assumption to make is that financial markets, people, and firms will do the best they can to work out the implications of that policy: Designing a policy on the assumption that people will make systematic mistakes in responding to it seems unwise.

At the same time, it is clear that the assumption of rational expectations overstates the ability of people and firms to think about the future and that we have to go beyond this assumption. Indeed, much current research focuses on how to take into account some of the behavioral limits and biases that determine how people form expectations. A workable and reliable alternative does not yet exist, however, and for the moment, rational expectations remain the default assumption of most macroeconomic models.

You have just learned an important lesson. The effects of monetary policy—the effects of any type of macroeconomic policy, for that matter—depend crucially on its effect on expectations:

- If a monetary expansion leads financial investors, firms, and consumers to revise their expectations of future interest rates and output, then the effects of the monetary expansion on output may be large.
- But if expectations remain unchanged, the effects of the monetary expansion on output will be limited.

This is why central banks think of their task as not only to adjust the policy rate but also to "manage expectations," to lead to predictable effects of changes in the policy rate on the economy. Giving an indication to financial markets about future policy rates is known as "forward guidance" by the central bank. More on this in Chapters 21 and 23.

We can link this to the discussion in Chapter 14 about the effects of changes in monetary policy on the stock market. Many of the same issues were present there. If, when the change in monetary policy takes place, it comes as no surprise to investors, firms, and consumers, then expectations will not change; the stock market will react only a little, if at all, and thus demand and output will change only a little, if at all. But if the change comes as a surprise and is expected to last, expectations of future output will go up, expectations of future interest rates will come down, the stock market will boom, and output will increase.

At this stage, you may have become skeptical that macroeconomists can say much about the effects of policy or the effects of other shocks. If the effects depend so much on what happens to expectations, can macroeconomists have any hope of predicting what will happen? The answer is yes.

Saying that the effect of a particular policy depends on its effect on expectations is not the same as saying that anything can happen. Expectations are not arbitrary. The manager of a mutual fund who must decide whether to invest in stocks or bonds, the firm thinking about whether to build a new plant, the consumer thinking about how much she should save for retirement, all give a lot of thought to what might happen in the future. We can think of each of them as forming expectations about the future by assessing the likely course of future expected policy and then working out the implications for future activity. If they do not do it themselves (surely most of us do not spend our time solving macroeconomic models before making decisions), they do so indirectly by watching TV, reading newspapers, or finding public information on the internet, all of which in turn rely on the forecasts of public and private forecasters. Economists refer to expectations formed in this forward-looking manner as **rational expectations**. The introduction of the assumption of rational expectations, starting in the 1970s, has largely shaped the way macroeconomists think about policy. It is discussed further in the Focus Box "Rational Expectations."

We could go back and think about the implications of rational expectations in the case of the monetary expansion we have just studied. It will be more fun to do this in the context of a change in fiscal policy, and this is what we now turn to.

16-3 DEFICIT REDUCTION, EXPECTATIONS, AND OUTPUT

We discussed the short- and medium-run effects of changes in fiscal policy in Section 9–3. We discussed the long-run effects of changes in fiscal policy in Section 11–2.

Recall the conclusions we reached in the core about the effects of a budget deficit reduction:

- In the short run, a reduction in the budget deficit, unless it is offset by a monetary expansion, leads to lower private spending and to a contraction in output.
- In the medium run, a lower budget deficit implies higher saving and higher investment.
- In the long run, higher investment translates into higher capital and thus higher output.

It is the adverse short-run effect that—in addition to the unpopularity of increases in taxes or reductions in government programs—often deters governments from tackling their budget deficits. Why take the risk of a recession now for benefits that will accrue only in the future?

A number of economists have argued, however, that under some conditions a deficit reduction might actually increase output even in the *short run*. Their argument is that, if people take into account the future beneficial effects of deficit reduction, their expectations about the future might improve enough to lead to an increase—rather than a decrease—in current spending, thereby increasing current output. This section explores their argument. The Focus Box "Can a Budget Deficit Reduction Lead to an Output Expansion? Ireland in the 1980s" reviews some of the supporting evidence.

Assume the economy is described by equation (16.3) for the IS relation and equation (16.4) for the LM relation. Now suppose the government announces a program to reduce the deficit, through decreases in both current spending, G, and future spending, G'^e. What will happen to output in *the current period*?

The Role of Expectations about the Future

Suppose first that expectations of future output (Y'^e) and of the future interest rate (r'^e) do not change. Then we get the standard answer: The decrease in government spending in the current period leads to a shift of the IS curve to the left, and so to a decrease in output.

The crucial question therefore is what happens to expectations. To answer, let us go back to what we learned in the core about the effects of a deficit reduction in the medium run and the long run.

- In the medium run, a deficit reduction has no effect on output. It leads, however, to a lower interest rate and to higher investment. These were two of the main lessons of Chapter 9.

  Let's review the logic behind each.

  Recall that, when we look at the medium run, we ignore the effects of capital accumulation on output. So in the medium run, the natural level of output depends on the level of productivity (taken as given) and on the natural level of employment. The natural level of employment depends in turn on the natural rate of unemployment. If spending by the government on goods and services does not affect the natural rate of unemployment—and there is no obvious reason why it should—then changes in spending will not affect the natural level of output. Therefore, a deficit reduction has no effect on the level of output in the medium run.

  Now recall that output must be equal to spending, which is the sum of public and private spending. Given that output is unchanged and that public spending is lower, private spending must therefore be higher. To achieve higher private spending requires a lower equilibrium interest rate: The lower interest rate leads to higher investment and thus to higher private spending, which offsets the decrease in public spending and output is unchanged.
- In the long run—that is, taking into account the effects of capital accumulation on output—higher investment leads to a higher capital stock and therefore a higher level of output.

In the medium run: Output, Y, does not change; investment, I, is higher.

In the long run:
I increases ⇒
K increases ⇒
Y increases.

This was the main lesson of Chapter 11. The higher the proportion of output saved (or invested; investment and saving must be equal for the goods market to be in equilibrium in a closed economy), the higher the capital stock, and thus the higher the level of output in the long run.

We can think of our *future period* as including both the medium and the long run. If people, firms, and financial market participants have *rational expectations*, then, in response to the announcement of a deficit reduction, they will expect these developments to take place in the future. Thus, they will revise their expectation of future output (Y'^e) up, and their expectation of the future interest rate (r'^e) down.

The way this is likely to happen in practice: Forecasts by economists will show that these lower deficits are likely to lead to higher output and lower interest rates in the future. In response to these forecasts, long-term interest rates will decrease and the stock market will increase. People and firms, reading these forecasts and seeing the increase in bond and stock prices, will revise their spending plans and increase spending.

Back to the Current Period

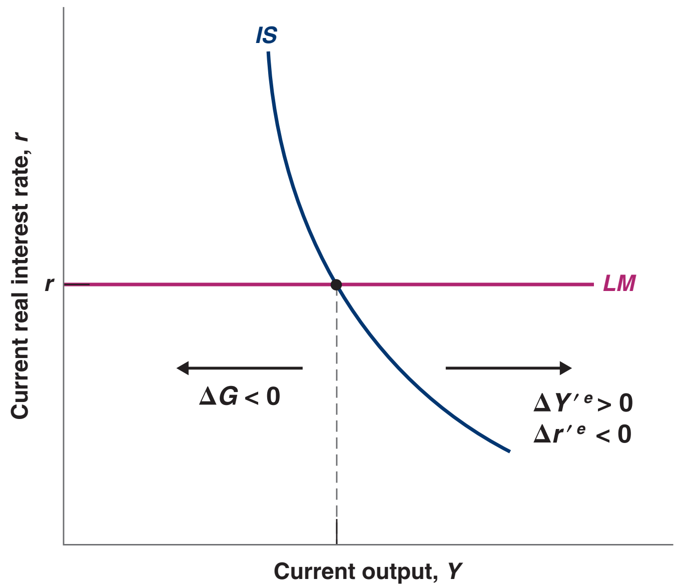
We can now return to the question of what happens *this period* in response to the announcement and start of the deficit reduction program. Figure 16-5 draws the IS and LM curves for the current period. In response to the announcement of the deficit reduction, there are now three factors shifting the IS curve:

- Current government spending (G) goes down, leading the IS curve to shift to the left. At a given interest rate (r), the decrease in government spending leads to a decrease

Figure 16-5

The Effects of a Deficit Reduction on Current Output

When account is taken of its effect on expectations, the decrease in government spending need not lead to a decrease in output.



in total spending and so a decrease in output. This is the standard effect of a reduction in government spending, and the only one taken into account in the basic IS-LM model.

- Expected future output (Y'^e) goes up, leading the IS curve to shift to the right. At a given interest rate, the increase in expected future output leads to an increase in private spending, increasing output.
- The expected future interest rate (r'^e) goes down, leading the IS curve to shift to the right. At a given current interest rate, a decrease in the future interest rate stimulates spending and increases output.

What is the net effect of these three shifts in the IS curve? Can the effect of expectations on consumption and investment spending offset the decrease in government spending? Without much more information about the exact form of the IS relation and about the details of the deficit reduction program, we cannot tell which shifts will dominate and whether output will go up or down. But our analysis suggests that both cases are possible, that output may go up in response to the deficit reduction. And it gives us a few hints as to when this might happen:

- Timing matters. Note that the smaller the decrease in current government spending (G), the smaller the adverse effect on spending today. Note also that the larger the decrease in expected future government spending (G'^e), the larger the effect on expected future output and interest rates, thus the larger the favorable effect on spending today. This suggests that credibly **backloading** the deficit reduction program toward the future, with small cuts today and larger cuts in the future, is more likely to lead to an increase in output. On the other hand, backloading raises an obvious issue. Announcing the need for painful cuts in spending, and then leaving them to the future, is likely to decrease the program's **credibility**—the perceived probability that the government will do what it has promised when the time comes to do it. The government must play a delicate balancing act: enough cuts in the current period to show a commitment to deficit reduction, enough cuts left to the future to reduce the adverse effects on the economy in the short run.

Can a Budget Deficit Reduction Lead to an Output Expansion? Ireland in the 1980s

FOCUS

Ireland went through two major deficit reduction programs in the 1980s.

1. The first program was started in 1982. In 1981, the budget deficit had reached a high 13% of GDP. Government debt, the result of the accumulation of current and past deficits, was 77% of GDP, also a high level. The Irish government clearly had to regain control of its finances. Over the next three years, it embarked on a program of deficit reduction, based mostly on tax increases. This was an ambitious program. Had output continued to grow at its normal growth rate, the program would have reduced the deficit by 5% of GDP.

   The results, however, were dismal. As shown in line 2 of Table 1, output growth was low in 1982 and negative in 1983. Low output growth was associated with a major increase in unemployment, from 9.5% in 1981 to 15% in 1984 (line 3). Because of low output growth, tax revenues—which depend on the level of economic activity—were lower than anticipated. The actual deficit reduction from 1981 to 1984, shown in line 1, was only 3.5% of GDP. And the result of continuing high deficits and low GDP growth was a further increase in the ratio of debt to GDP to 97% in 1984.

2. A second attempt to reduce budget deficits was made starting in February 1987. Things were still very bad. The 1986 deficit was 10.7% of GDP; debt stood at 116% of GDP, a record high in Europe at the time. The new program of deficit reduction was different from the first: It focused more on reducing the role of government and cutting government spending than on increasing taxes. Tax increases were achieved through a tax reform that widened the tax base—increasing the number of households paying taxes—rather than through an increase in the marginal tax rate. The program was again ambitious. Had output grown at its normal rate, the reduction in the deficit would have been 6.4% of GDP.

   The results of the second program could not have been more different from the results of the first. The years 1987 to 1989 were years of strong growth, with average GDP growth exceeding 5%. The unemployment rate was reduced by almost 2%. Because of strong output growth, tax revenues were higher than anticipated, and the deficit was reduced by nearly 9% of GDP.

A number of economists have argued that the striking difference between the results of the two programs can be traced to the different reaction of expectations in each case. The first program, they argue, focused on tax increases and did not change what many people saw as too large a role of government in the economy. The second program, with its focus on cuts in spending and on tax reform, had a much more positive impact on expectations, and so a positive impact on spending and output.

Are these economists right? One variable, the household saving rate—defined as disposable income minus consumption, divided by disposable income—strongly suggests that expectations are an important part of the story. To interpret the behavior of the saving rate, recall the lessons from Chapter 15 about consumption behavior. When disposable income grows unusually slowly or falls—as it does in a recession—consumption typically slows down or declines by less than disposable income because people expect things to improve in the future. Put another way, when the growth of disposable income is unusually low, the saving rate typically comes down. Now look (in line 4) at what happened from 1981 to 1984. Despite low growth throughout the period and a recession in 1983, the household saving rate actually increased slightly during the period. Put another way, people reduced their consumption by more than the reduction in their disposable income: The reason must be that they were pessimistic about the future.

Now turn to the period 1987 to 1989. During that period, economic growth was unusually strong. By the same argument as in the previous paragraph, we would have expected consumption to increase less strongly and thus the saving rate to increase. Instead, the saving rate dropped sharply, from 15.7% in 1986 to 12.6% in 1989. Consumers must have become much more optimistic about the future to increase their consumption by more than the increase in their disposable income.

Table 1 Fiscal and Other Macroeconomic Indicators in Ireland, 1981 to 1984, and 1986 to 1989

		1981	1982	1983	1984	1986	1987	1988	1989
1	Budget deficit (% of GDP)	–13.0	–13.4	–11.4	–9.5	–10.7	–8.6	–4.5	–1.8
2	Output growth rate (%)	3.3	2.3	–0.2	4.4	–0.4	4.7	5.2	5.8
3	Unemployment rate (%)	9.5	11.0	13.5	15.0	16.1	16.9	16.3	15.1
4	Household saving rate (% of disposable income)	17.9	19.6	18.1	18.4	15.7	12.9	11.0	12.6

Source: OECD Economic Outlook, June 1998.

The next question is whether this difference in the adjustment of expectations over the two episodes can be attributed fully to the differences in the two fiscal programs. The answer is no. Ireland was changing in many ways at the time of the second fiscal program. Productivity was increasing much faster than real wages, reducing the cost of labor for firms. Attracted by tax breaks, low labor costs, and an educated labor force, many foreign firms were relocating to Ireland and building new plants. These factors played a major role in the expansion of the late 1980s, and Irish growth remained very strong—usually more than 5% per year—until the crisis in 2007. The long expansion was doubtless due to many factors, but the change in fiscal policy in 1987 probably played an important role in convincing people, firms—including foreign firms—and financial markets that the government was regaining control of its finances. And the fact remains that the substantial deficit reduction of 1987–1989 was accompanied by a strong output expansion, not by the recession predicted by the basic IS-LM model.

For a more detailed discussion, look at Francesco Giavazzi and Marco Pagano, "Can Severe Fiscal Contractions Be Expansionary? Tales of Two Small European Countries," NBER Macroeconomics Annual (MIT Press, 1990), Olivier Jean Blanchard and Stanley Fischer, editors.

For a more systematic look at whether and when fiscal consolidations have been expansionary (and a mostly negative answer), see "Will It Hurt? Macroeconomic Effects of Fiscal Consolidation," Chapter 3, World Economic Outlook, International Monetary Fund, October 2010.

- Composition matters. How much of the reduction in the deficit is achieved by raising taxes and how much by cutting spending may be important. If some government spending programs are perceived as "wasteful," cutting these programs today will allow the government to cut taxes in the future. Expectations of lower future taxes and lower distortions could induce firms to invest today, thus raising output in the short run.
- The initial situation matters. Take an economy where the government appears to have, in effect, lost control of its budget. Government spending is high, tax revenues are low, and the deficit is large. Government debt is increasing fast. In such an environment, a credible deficit reduction program is more likely to increase output in the short run. Before the announcement of the program, people may expect major political and economic troubles in the future. The announcement of a program of deficit reduction may reassure them that the government has regained control and that the future is less bleak than they anticipated. This decrease in pessimism about the future may lead to an increase in spending and output, even if taxes are increased as part of the deficit reduction program. Investors who thought that the government might default on the debt and were asking for a large risk premium may conclude that the risk of default is much lower and ask for much lower interest rates. Lower interest rates for the government are likely to translate into lower interest rates for firms and people.
- Monetary policy matters. The three previous arguments focused on the direction of the shift in the IS curve, with no change in monetary policy. But as we have discussed before, even if it cannot fully offset the effect of an adverse shift in the IS curve, monetary policy can, by decreasing the policy rate, help reduce the adverse effects of the shift on output.

Let's summarize.

A program of deficit reduction may increase output even in the short run. Whether it does or does not depends on many factors:

Note how far we have moved from the results of Chapter 3, where, by choosing spending and taxes wisely, the government could achieve any level of output it wanted. Here, even the sign of the effect of a deficit reduction on output is ambiguous. More on current fiscal policy issues in Chapter 22.

- The credibility of the program: Will spending be cut or taxes increased in the future as announced?
- The composition of the program: Does the program remove some of the distortions in the economy?
- The state of government finances in the first place: How large is the initial deficit? Is this a "last chance" program? What will happen if it fails?

- Monetary and other policies: Will they help offset the direct adverse effect on demand in the short run?

This gives you a sense of both the importance of expectations in determining the outcome and the complexities involved in the use of fiscal policy in such a context. And it is far more than an illustrative example. This has been a major bone of contention in the euro area since the beginning of 2010.

By 2010, the sharp economic downturn, together with the fiscal measures to limit the fall in demand during 2009, had led to large budget deficits and large increases in government debt. There was little question that the large deficits could not go on forever, and debt had to be eventually stabilized. The question was when and at what pace?

Some economists, and most of the policymakers in the euro area, believed that fiscal consolidation had to start immediately and be strong. They argued that this was essential to convince investors that the fiscal situation was under control. They argued that, if coupled with structural reforms to increase future output, the effect through anticipations of higher output later would dominate the direct adverse effects of consolidation. They argued for what became known as **fiscal austerity**. For example, the president of the European Central Bank, Jean Claude Trichet, said in September 2010:

> "[Fiscal consolidation] is a prerequisite for maintaining confidence in the credibility of governments' fiscal targets. Positive effects on confidence can compensate for the reduction in demand stemming from fiscal consolidation, when fiscal adjustment strategies are perceived as credible, ambitious and focused on the expenditure side. The conditions for such positive effects are particularly favourable in the current environment of macroeconomic uncertainty."

Others were skeptical that, in a depressed environment, the positive expectation effects would be strong. They pointed out that the policy rate was already at the zero lower bound, and so monetary policy could not help much, if at all. They argued for a slow and steady fiscal consolidation, even if it were to lead to higher levels of debt until debt stabilized.

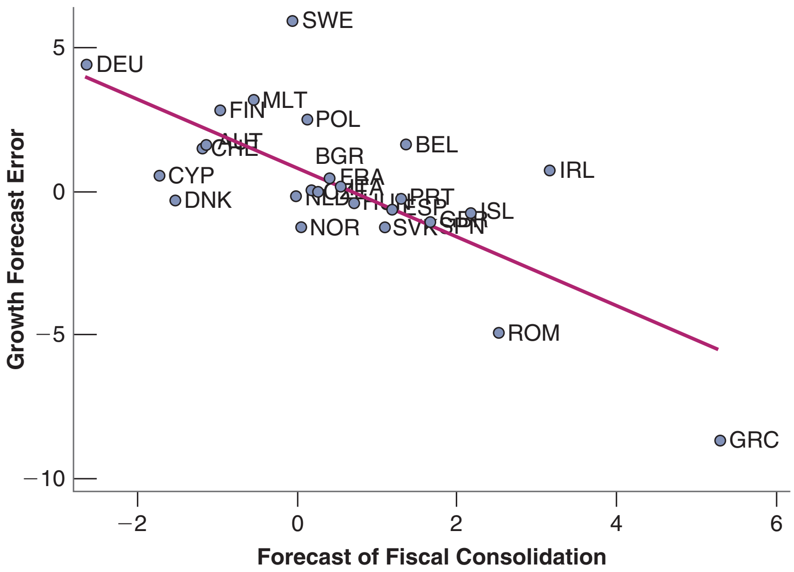
The debate became known as the **fiscal multipliers** debate. Those in favor of strong consolidation argued that the fiscal multipliers—the net effects of fiscal consolidation once direct and expectation effects were taken into account—were likely to be *negative*. Smaller deficits would lead, other things being equal, to *an increase* in output. Those against consolidation argued that fiscal multipliers were likely to be *positive* and possibly large. Smaller deficits would lead to a decrease in output, or at least slow down the recovery.

The skeptics turned out, unfortunately, to be right. As evidence accumulated, it became clear that the net effect of fiscal consolidation was contractionary. The strongest evidence was the relation between forecast errors and the size of fiscal consolidation across countries. In most euro countries, growth in 2010 and 2011 turned out to be much lower than had been forecast. Looking across countries, the negative forecast errors were closely correlated with the size of fiscal consolidation. As shown in Figure 16-6, which plots growth forecast errors against a measure of fiscal consolidation, countries with larger fiscal consolidations showed a larger (negative) forecast error. This was particularly striking in the case of Greece, but was true of other countries as well. Given that the forecasts had been constructed using models that implied small positive multipliers, this evidence implied that the fiscal multipliers were in fact not only positive, but larger than had been assumed. Expectation effects did not offset the adverse direct effects of lower spending and higher taxes.

Figure 16-6

Growth Forecast Errors and Fiscal Consolidation in Europe, 2010–2011

European countries with stronger fiscal consolidations in 2010 and 2011 had larger negative growth forecast errors.



FOCUS

Uncertainty and Fluctuations

The future is, by its nature, uncertain. We hold expectations about the outcome, but we realize that the outcome is uncertain. So far, we have focused on the effect of expectations on consumption and investment decisions and ignored this uncertainty. But, for given expectations, uncertainty is likely to affect our decisions as well.

Consider three examples.

Picture two scenarios. In the first, you are sure that, a year from now, you will be promoted and your salary will be increased by 10%. In the second, you think that there is a 50% chance you will be promoted and your wage will be increased by 40%, and a 50% chance you will be fired, in which case you are confident that you can find another job, but at a wage 20% below your current wage. Your expected income is the same in both cases, namely higher by 10%. Will your consumption this year be the same under both scenarios? Probably not. In the second scenario, you are likely to be more careful and increase consumption by less.

Suppose now that you are the CEO of a firm and you hear that there may be a tariff war that may affect some of your sales abroad. The rumors are vague, the list of products that might be affected is not clear. In other words, and in contrast to the previous case, you are unable to even form an opinion about probabilities and outcomes. This is very likely to lead you to be extremely prudent and to cancel a number of investment plans that you were considering earlier.

Continue with the previous example but assume that you are confident that the uncertainty will be resolved within a year: you will know whether there is a tariff and whether and how it affects your business. Then, if investment decisions are difficult to reverse, it makes a lot of sense for you to wait to make those decisions until the uncertainty will be resolved. Economists refer to such situations as reflecting the **option value of waiting**. Making what turns out to be the wrong investment decision can be very costly; waiting until the uncertainty is resolved and the right decision can be made may be much less costly.

I have focused on individual decisions. But if many people or many firms are subject to such uncertainty, this may have major macroeconomic effects. Here are three examples.

The first dates back to the Great Depression. When the stock market collapsed in October 1929, people realized that the future was going to be worse than they had anticipated. But they also perceived the future as being much more uncertain. Was the stock market collapse a temporary event, or was it going to trigger a long and painful recession? Christina Romer, of Berkeley, argues that both lower expectations and higher perceived uncertainty were at work. Her argument: Sales of automobiles plummeted in October and November. But sales of department stores barely budged. If only lower expectations had been at work, both sales of automobiles and sales at department stores should have come down. The fact that sales of automobiles went down a lot, while sales at department stores remained roughly the same, suggests that uncertainty and the option value of waiting were at work. Not knowing what was going to happen, consumers decided to wait and see before buying a car. This behavior in turn contributed to the fall in demand and the speed at which the US economy collapsed.

The second refers to the Great Recession. After the collapse of Lehman in September 2008 (see Chapter 6), it became clear that the financial system was in serious trouble. But assessing how bad things were, and guessing what might happen, was extremely

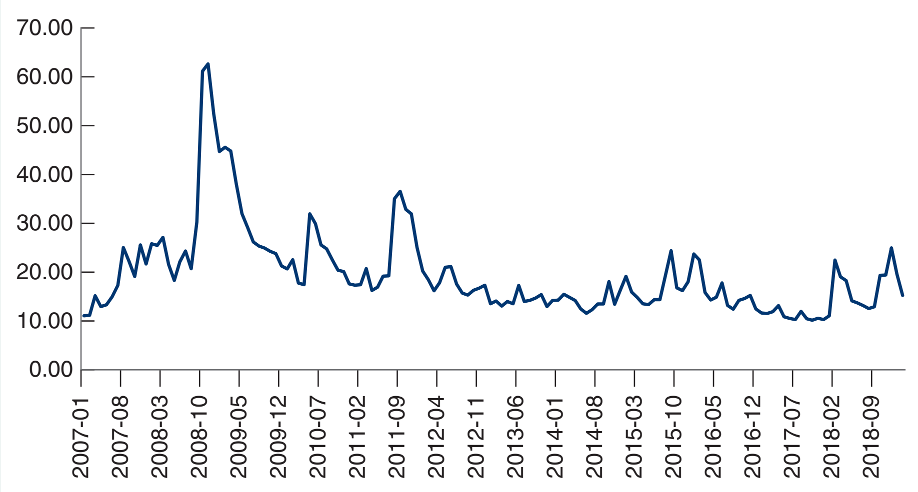


Figure 1

The Evolution of the VIX since 2007

Source: FRED, VIXCLS

difficult. (I know; I was at the IMF at the time.) A measure of the degree of uncertainty that prevailed in financial markets is given in Figure 1, which plots the evolution of a stock market volatility index, called the **VIX.** The VIX is computed using the prices of options on buying or selling stocks at a given price in the future. If investors are worried that the stock market may collapse, they will be willing to pay a high price to have the option to sell at a given price in the future. Conversely, if they expect the stock market to recover rapidly, they will be willing to pay a high price to have the option to buy at a given price in the future. The VIX is constructed by aggregating this information, and gives a sense of the uncertainty in the market at a given point in time. Figure 1 shows how sharply the VIX went up in September and October 2008, reflecting the large increase in uncertainty at the time and contributing to a sharp decline in demand and a deep recession. It also shows how fast the smoke cleared, and how fast it went down and has remained low since then.

The third example deals with current trade tensions. Since the 2016 election, the Trump administration has argued that tariffs on US exports were too high and should be reduced. Talks with Mexico, China, and the European Union are ongoing. There is, however, substantial uncertainty as to what the final outcome will be. In that context, you would expect firms, in particular exporting firms, to put some investment on hold and to wait until the outcome of the talks is clearer. This is indeed what the evidence shows. A survey of firms by the Federal Reserve Bank of Atlanta in January 2019 shows that 19% of them were reassessing their investment plans as a result of the tariff discussions. Of those, 36% were postponing some of their plans. The option value of waiting was clearly at work, not enough to lead to a large decrease in overall investment but enough to be noticeable in aggregate data.

SUMMARY

- Private spending in the goods market depends on current and expected future output and on current and expected future real interest rates.
- Expectations affect demand and, in turn, output. Changes in expected future output or in the expected future real interest rate lead to changes in spending and in output today.
- By implication, the effects of fiscal and monetary policy on spending and output depend on how the policy affects expectations of future output and real interest rates.
- The assumption of rational expectations is the assumption that people, firms, and participants in financial markets form expectations of the future by assessing the course of future expected policy and then working out the implications for future output, future interest rates, and so on. Although it is clear that most people do not go through this exercise themselves, we can think of them as doing so indirectly by relying on the predictions of public and private forecasters.
- Although there are surely cases in which people, firms, or financial investors do not have rational expectations, the assumption of rational expectations seems to be the best benchmark to evaluate the potential effects of alternative policies. Designing a policy on the assumption that people will make systematic mistakes in responding to it would be unwise.

- The central bank controls the short-term real interest rate. Spending, however, depends instead on current and expected future real interest rates. Thus, the effect of monetary policy on activity depends crucially on whether and how changes in the short-term real interest rate lead to changes in current and expected future real interest rates.
- A budget deficit reduction may lead to an increase rather than a decrease in output. This is because expectations of higher output and lower interest rates in the future may lead to an increase in spending that more than offsets the reduction in spending from the direct effect of the deficit reduction on total spending. Whether it does depends on the pace, credibility, nature of the deficit reduction, and the ability of monetary policy to accommodate and sustain demand.

KEY TERMS

aggregate private spending, or private spending, 337
animal spirits, 341
static expectations, 341
adaptive expectations, 341
rational expectations, 342
backloading, 344
credibility, 344
fiscal austerity, 347
fiscal multipliers, 347
option value of waiting, 348
VIX, 349

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. Changes in the current one-year real interest rate are likely to have a much larger effect on spending than changes in expected future one-year real interest rates.
b. The introduction of expectations in the goods market model makes the *IS* curve flatter, although it is still downward sloping.
c. Investment depends on current and expected future interest rates.
d. The rational expectations assumption implies that consumers take into account the effects of future fiscal policy on output.
e. Expected future fiscal policy affects expected future economic activity but not current economic activity.
f. Depending on its effect on expectations, a fiscal contraction may actually lead to an economic expansion.
g. Ireland's experience with deficit reduction programs in 1982 and 1987 provides strong evidence against the hypothesis that deficit reduction can lead to an output expansion.
h. The Euro area experience in 2010 and 2011 suggests that fiscal consolidations, through expectations, lead to substantial increases in output growth.

2. Consider these quotes concerning Federal Reserve policy.

a. *On January 25, 2012 the Federal Reserve made two announcements. Here are excerpts from each:*

The announcement about the path of short-term interest rates:
To support a stronger economic recovery and to help ensure that inflation, over time, is at levels consistent with the dual mandate, the Committee expects to maintain a highly accommodative stance for monetary policy. In particular, the Committee decided today to keep the target range for the federal funds rate at 0 to 1/4% and currently anticipates that economic conditions—including low rates of resource utilization and a subdued outlook for inflation over the medium run—are likely to warrant exceptionally low levels for the federal funds rate at least through late 2014.

The formal announcement of an inflation target:
The inflation rate over the longer run is primarily determined by monetary policy, and hence the Committee has the ability to specify a longer-run goal for inflation. The Committee judges that inflation at the rate of 2%, as measured by the annual change in the price index for personal consumption expenditures, is most consistent over the longer run with the Federal Reserve's statutory mandate. Communicating this inflation goal clearly to the public helps keep longer-term inflation expectations firmly anchored, thereby fostering price stability and moderate long-term interest rates and enhancing the Committee's ability to promote maximum employment in the face of significant economic disturbances.

What do you think was the combined goal of the two announcements? What do they imply about real interest rates?

b. *The monetary policy announcement December 19, 2018: Information received since the Federal Open Market Committee met in November indicates that the labor market has continued to strengthen and that economic activity has been rising at a strong rate. Job gains have been strong, on average, in recent months, and the unemployment rate has remained low. Household spending has continued to grow strongly, while growth of business fixed investment has moderated from its rapid pace earlier in the year. On a 12-month basis, both overall inflation and inflation for items other than food and energy remain near 2%. Indicators of longer-term inflation expectations are little changed, on balance.*

Consistent with its statutory mandate, the Committee seeks to foster maximum employment and price stability. The Committee judges that some further gradual increases in the *target range for the federal funds rate will be consistent with sustained expansion of economic activity, strong labor market conditions, and inflation near the Committee's symmetric 2% objective over the medium term.*

In view of realized and expected labor market conditions and inflation, the Committee decided to raise the target range for the federal funds rate to 2-1/4 to 2-1/2%.

What happened to the target range for the policy interest rate in this announcement? What information was given about the future course of policy interest rates? How was the decision justified? Can you make an inference about the value of the neutral policy rate of interest from this announcement?

3. *For each of the changes in expectations in parts a through d, determine whether there is a shift in the IS curve, the LM curve, both curves, or neither. In each case assume that no other exogenous variable is changing.*

a. a decrease in the expected future real interest rate
b. an increase in the current real policy interest rate
c. an increase in expected future taxes
d. a decrease in expected future income

4. *Consider the following statement: "The rational expectations assumption is unrealistic because, essentially, it amounts to the assumption that every consumer has perfect knowledge of the economy."* Discuss.

5. *A new president, who promised during the campaign that she would cut taxes, has just been elected. People trust that she will keep her promise, but expect that the tax cuts will be implemented only in the future. Determine the impact of the election victory on current output, the current interest rate, and current private spending under each of the assumptions in parts a through c. In each case, indicate what you think will happen to Y'^e, r'^e, and T'^e, and then how these changes in expectations affect output today.*

a. The Fed will not change its current policy rate, but does not make a commitment as to the direction of rates in the future.
b. The Fed will act to prevent any change in current and future output.
c. The Fed will not change either the current real policy interest rate or the future real policy interest rate.

6. *The Irish deficit reduction packages*

The Focus Box "Can a Budget Deficit Reduction Lead to an Output Expansion? Ireland in the 1980s" provides an example of fiscal consolidation. Ireland had a large budget deficit in 1981 and 1982.

a. What does a deficit reduction imply for the medium run and the long run? What are the advantages of reducing the deficit?
b. The box discusses two deficit reduction programs. How did they differ?
c. The box presents evidence that the two deficit reduction programs had different effects on household expectations. What is that evidence?
d. Although the data show strong output growth from 1987 to 1989, there is some evidence of continued macroeconomic weakness in Ireland during the second fiscal consolidation. What is that evidence?

DIG DEEPER

7. *A new Federal Reserve chairman*

Suppose, in a hypothetical economy, that the chairman of the Fed unexpectedly announces that he will retire in one year. At the same time, the President announces her nominee to replace the retiring Fed chair. Financial market participants expect the nominee to be confirmed by Congress. They also believe that the nominee will conduct a more contractionary monetary policy in the future. In other words, market participants expect the policy interest rate to increase in the future.

a. Consider the present to be the last year of the current Fed chair's term and the future to be the time after that. Given that monetary policy will be more contractionary in the future, what will happen to future interest rates and future output? Given these expected changes in future output and future interest rates, what will happen to output and the interest rate in the present? What will happen to the yield curve on the day of the announcement that the current Fed chair will retire in one year?

Now suppose that instead of making an unexpected announcement, the Fed chair is required by law to retire in one year (there are limits on the term of the Fed chair), and financial market participants have been aware of this for some time. Suppose, as in part a, that the president nominates a replacement who is expected to raise interest rates more than the current Fed chair.

b. Suppose financial market participants are not surprised by the president's choice. In other words, market participants had correctly predicted whom the president would choose as nominee. Under these circumstances, is the announcement of the nominee likely to have any effect on the yield curve?
c. Suppose instead that the identity of the nominee is a surprise and that financial market participants had expected the nominee to be someone who favored an even more contractionary policy than the actual nominee. Under these circumstances, what is likely to happen to the yield curve on the day of the announcement? (*Hint:* Be careful. Compared to what was expected, is the actual nominee expected to follow a more contractionary or more expansionary policy?)
d. On November 2, 2017, Jerome Powell was nominated to succeed Janet Yellen as chair of the Federal Reserve. You can find the 10-year bond yield on a daily basis (FRED variable DGS10) as well as the daily level of the Dow Jones average (FRED variable DJCA). Are there changes in stock prices or bond yields that suggest financial market participants were surprised by the choice of Powell? You could also do a yield curve analysis of the kind described in Problem 8 for the period around Powell's nomination. If you do this, use the one- and 10-year interest rates.)

EXPLORE FURTHER

8. Deficits and fiscal consolidation

As seen in the following table, the 2008 financial crisis left the United States with an enormous federal budget deficit in 2009.

There was a substantial fiscal consolidation from 2011 onward, yet real output continued to grow.

Fiscal Consolidation in the United States 2009–2014				
Year	**Receipts (% of GDP)**	**Outlays (% of GDP)**	**Surplus or Deficit (–) (% of GDP)**	**Growth in Real GDP (%)**
2008	17.1	20.2	–3.1	–0.3
2009	14.6	24.4	–9.8	–2.8
2010	14.6	23.4	–8.7	2.5
2011	15.0	23.4	–8.5	1.6
2012	15.3	22.1	–6.8	2.3
2013	16.7	20.8	–4.1	2.2
2014	17.5	20.3	–2.8	2.4

(*Source:* Table B-1, Table B-20, Economic Report of the President 2015.)

a. Which played a larger role in the fiscal consolidation: raising taxes, or reducing outlays?

b. In terms of the language of the text, if this fiscal consolidation were anticipated as of 2009, was it "backloaded?" How might this help minimize the effects of the fiscal consolidation on output growth?

c. We know from Question 2 and from Chapters 4 and 6 that monetary policy maintained the nominal policy rate of interest of close to 0% throughout this period and promised to maintain low interest rates into the future. Why did it do so?

d. The Federal Reserve introduced a target rate of inflation during the consolidation period on January 25, 2012. What is one advantage of introducing a policy where inflation is targeted at 2% during a period of zero interest rates and fiscal consolidation?

e. We used the University of Michigan's Index of Consumer Sentiment in the previous chapter as a measure of expectations of households about the future. You can look at the values of this index at the FRED database maintained by the Federal Reserve Bank of St. Louis (*series UMCSENT*). Find this index and comment on its evolution from 2010 to 2014 as the fiscal consolidation proceeded.

FURTHER READINGS

- For a general discussion of the effects of uncertainty on the macroeconomy, read "Fluctuations in Uncertainty" by Nicolas Bloom, *Journal of Economic Perspectives*, Spring 2014.
- For more evidence on the effects of uncertainty, go to the site www.policyuncertainty.com, which gives the evolution of a number of indexes of uncertainty for a large number of countries.
- The study of the effects of tariffs on US firms' decisions is available at https://macroblog.typepad.com/macroblog/2019/02/tariff-worries-and-us-business-investment-take-two.html.

The Open Economy

The next four chapters cover the second extension of the core. They look at the implications of openness—the fact that most economies trade both goods and assets with the rest of the world.

EXTENSIONS

Chapter 17

Chapter 17 discusses the implications of openness in goods markets and financial markets. Openness in goods markets allows people to choose between domestic goods and foreign goods. An important determinant of their decisions is the real exchange rate—the relative price of domestic goods in terms of foreign goods. Openness in financial markets allows people to choose between domestic assets and foreign assets. This imposes a tight relation between the exchange rate, both current and expected, and domestic and foreign interest rates—a relation known as the *interest parity condition*.

Chapter 18

Chapter 18 focuses on equilibrium in the goods market in an open economy: the demand for domestic goods now depends also on the real exchange rate. It shows how fiscal policy affects both output and the trade balance. It discusses the conditions under which a real depreciation improves the trade balance and increases output.

Chapter 19

Chapter 19 characterizes the equilibrium in goods and financial markets in an open economy. In other words, it gives an open economy version of the IS-LM model we saw in the core. It shows how, under flexible exchange rates, monetary policy affects output through its effect not only on the interest rate but also on the exchange rate. Fixing the exchange rate also implies giving up the ability to change the interest rate.

Chapter 20

Chapter 20 looks at the properties of different exchange rate regimes. It shows that, in the medium run, the real exchange rate can adjust even under a fixed exchange rate regime. The chapter then looks at exchange rate crises under fixed exchange rates, and at movements in exchange rates under flexible exchange rates. It ends by discussing the pros and cons of various exchange rate regimes, including the adoption of a common currency such as the euro.

This page is intentionally left blank

17

Openness in Goods and Financial Markets

We have assumed until now that the economy we looked at was *closed*—that it did not interact with the rest of the world. We had to start this way to keep things simple and to build up intuition for the basic macroeconomic mechanisms. Figure 17-1, which repeats for convenience Figure 1-1, shows how bad, in fact, this assumption is. The figure plots the growth rates for advanced and emerging economies since 2005. What is striking is how the growth rates have moved together. Even though the crisis originated in the United States, the outcome was a worldwide recession with negative growth in advanced economies and much lower growth in emerging economies. It is therefore time to dismiss our closed economy assumption. Understanding the macroeconomic implications of openness will occupy us for this and the next three chapters.

Openness has three distinct dimensions:

1. **Openness in goods markets**—the ability of consumers and firms to choose between domestic goods and foreign goods. In no country is this choice completely free of restrictions. Even the countries most committed to free trade have **tariffs**—taxes on imported goods—and **quotas**—restrictions on the quantity of goods that can be imported—on at least some foreign goods. At the same time, in most countries, average tariffs are low and, until recently, were getting lower.

2. **Openness in financial markets**—the ability of financial investors to choose between domestic assets and foreign assets. Not long ago, even some of the richest countries in the world, such as France and Italy, had strong **capital controls**—restrictions on the foreign assets their domestic residents could hold and the domestic assets foreigners could hold. Capital controls are much more limited today. As a result, world financial markets are becoming more closely integrated.

3. **Openness in factor markets**—the ability of firms to choose where to locate production and of workers to choose where to work. Here also trends have been clear. Multinational companies operate plants in many countries and move their operations around the world to take advantage of low costs. Much of the debate about the **North American Free Trade Agreement (NAFTA)** signed in 1993 by the United States, Canada, and Mexico centered on how it would affect the relocation of US firms to Mexico. Similar fears now center on China. And immigration from low-wage countries is a hot political issue throughout Europe and in the United States.

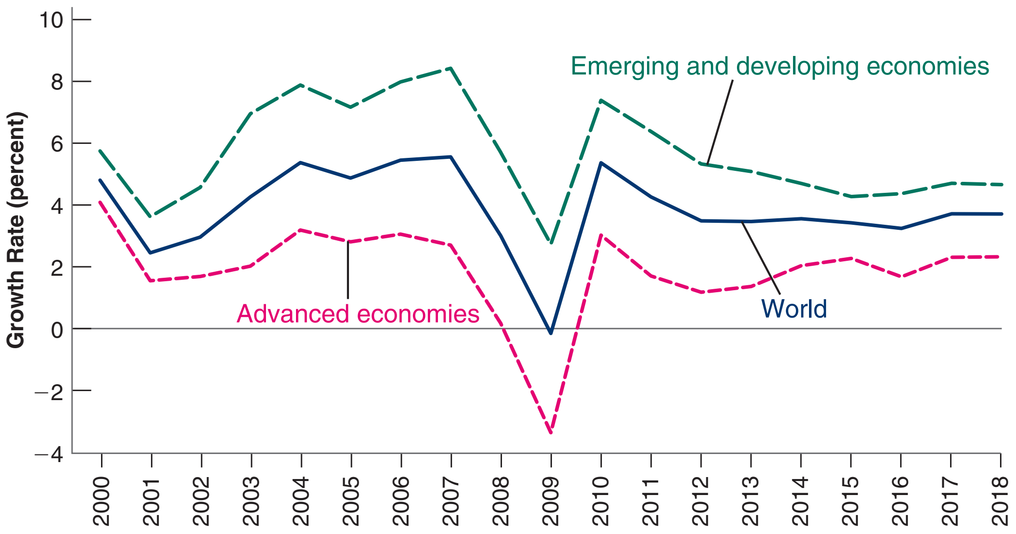
In the short run and in the medium run—the focus of this and the next three chapters—openness in factor markets plays much less of a role than openness in either goods markets or financial markets. Thus, we shall ignore openness in factor markets and focus on the implications of the first two dimensions of openness here.

Figure 17-1

Growth in Advanced and Emerging Economies since 2000

Advanced, emerging, and developing economies move very much together.

Source: IMF, World Economic Outlook, Oct 2015. Used courtesy of IMF.



Section 17-1 looks at openness in goods markets, determinants of the choice between domestic and foreign goods, and the role of the real exchange rate.

Section 17-2 looks at openness in financial markets, determinants of the choice between domestic and foreign assets, and the role of interest rates and exchange rates.

Section 17-3 gives a map to the next three chapters.

If you remember one basic message from this chapter, it should be: When looking at an open economy, we have to think about both the choice of domestic versus foreign goods and that of domestic versus foreign assets. ▶▶▶

17-1 OPENNESS IN GOODS MARKETS

Let's start by looking at how much the United States sells to and buys from the rest of the world. Then we shall be better able to think about the choice between domestic goods and foreign goods, and the role of the relative price of domestic goods in terms of foreign goods—the real exchange rate.

Exports and Imports

Figure 17-2 plots the evolution of US exports and US imports, as ratios to GDP, since 1960 ("US exports" means exports *from* the United States; "US imports" means imports *to* the United States). The figure suggests two main conclusions:

From Chapter 3: The trade balance is the difference between exports and imports: If exports exceed imports, there is a trade surplus (equivalently, a positive trade balance). If imports exceed exports, there is a trade deficit (equivalently, a negative trade balance). ▶

- The US economy is becoming more open over time. Exports and imports, which were equal to 5% of GDP in the early 1960s, are now equal to about 14% of GDP (12.3% for exports, 15.4% for imports). In other words, the United States trades roughly three times as much (relative to its GDP) with the rest of the world than it did 50 years ago.
- Although imports and exports have followed the same upward trend, since the late 1970s imports have consistently exceeded exports. Put another way, for the last 40 years, the United States has consistently run a trade deficit. For four years in a row in the mid-2000s, the ratio of the trade deficit to GDP exceeded 5% of GDP. Although it has decreased since the beginning of the crisis, it remains large today, at 3%.

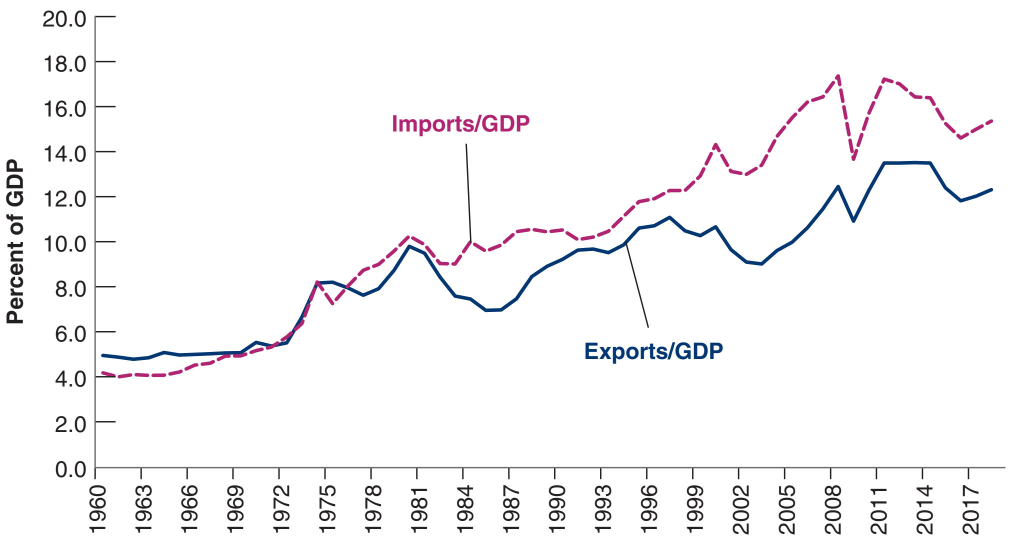


Figure 17-2

US Exports and Imports as Ratios to GDP since 1960

Since 1960, US exports and imports have roughly tripled in relation to GDP. The United States has become a much more open economy.

Source: Series GDP, EXPGS, IMPGS. Federal Reserve Economic Data (FRED) https://research.stlouisfed.org/fred2/

Understanding the sources and implications of this large deficit is an important issue and one to which we shall return later.

Given all the talk in the media about *globalization*, a volume of trade (measured by the average of the ratios of exports and imports to GDP) around 14% of GDP might strike you as small. However, the volume of trade is not necessarily a good measure of openness. Many firms are exposed to foreign competition, but by being competitive and keeping their prices low enough, these firms are able to retain their domestic market share and limit imports. This suggests that a better index of openness than export or import ratios is the proportion of aggregate output composed of **tradable goods**—goods that compete with foreign goods in either domestic markets or foreign markets. Estimates are that tradable goods represent about 60% of aggregate output in the United States today.

Tradable goods: cars, computers, pharmaceuticals, etc. Nontradable goods: housing, most medical services, haircuts, etc.

With exports around 12.3% of GDP, it is true that the United States has one of the smallest ratios of exports to GDP among the rich countries of the world. Table 17-1 gives ratios for a number of OECD countries.

For more on the OECD, see Chapter 1.

The United States is at the low end of the range of export ratios. Japan's ratio is a bit higher, the United Kingdom's nearly three times as large, and Germany's four times as large. The smaller European countries have very large ratios, from 65.0% in Switzerland to 86.4% in the Netherlands. (The high Netherlands ratio of exports to GDP raises an odd possibility: Can a country have exports larger than its GDP; in other words, can a country have an export ratio greater than one? The answer is yes. The reason why is given in the Focus Box "Can Exports Exceed GDP?".)

Table 17-1 Ratios of Exports to GDP for Selected OECD Countries, 2017

Country	Export Ratio	Country	Export Ratio
United States	12.3%	Germany	47.2%
Japan	16.1%	Austria	53.9%
Chile	28.7%	Switzerland	65.0%
United Kingdom	30.5%	Netherlands	86.4%

Source: World Bank database, exports

FOCUS

Can Exports Exceed GDP?

Can a country have exports larger than its GDP—that is, can it have an export ratio greater than one?

It would seem that the answer must be no: A country cannot export more than it produces, so that the export ratio must be less than one. Not so. The key to the answer is to realize that exports and imports include intermediate goods.

Take, for example, a country that imports intermediate goods for $1 billion. Suppose it transforms them into final goods using only labor. Say labor is paid $200 million and that there are no profits. The value of these final goods is thus equal to $1,200 million. Assume that $1 billion worth of final goods is exported and the rest, $200 million, is consumed domestically.

Exports and imports therefore both equal $1 billion. What is GDP in this economy? Remember that GDP is value added in the economy (see Chapter 2). So in this example, GDP equals $200 million, and the ratio of exports to GDP equals $\$1000/\$200 = 5$.

Hence, exports can exceed GDP. This is actually the case for a number of small countries where most economic activity is organized around a harbor and import-export activities. This is even the case for small countries such as Singapore, where manufacturing plays an important role. In 2017, the ratio of exports to GDP in Singapore was 173%!

Do these numbers indicate that the United States has higher trade barriers than, say, the United Kingdom or the Netherlands? No. The main factors behind these differences are geography and size. Distance from other markets explains a good part of the lower Japanese ratio. Size also matters; the smaller the country, the more it must specialize in producing and exporting only a few products and rely on imports for the other products. The Netherlands can hardly afford to produce the range of goods produced by the United States, a country roughly 20 times its economic size.

Iceland is both isolated and small. What would you expect its export ratio to be? (Answer: 47%.)

The Choice between Domestic Goods and Foreign Goods

How does openness in goods markets force us to rethink the way we look at equilibrium in the *goods market*?

Until now, when we were thinking about consumers' decisions in the goods market, we focused on their decision to save or to consume. When goods markets are open, domestic consumers face a second decision: whether to buy domestic or foreign goods. Indeed, all buyers—including domestic and foreign firms and governments—face the same decision. This decision has a direct effect on domestic output. If buyers decide to buy more domestic goods, the demand for domestic goods increases, and so does domestic output. If they decide to buy more foreign goods, then foreign output increases instead of domestic output.

In a closed economy, people face one spending decision: Save or buy (consume). In an open economy, they face two spending decisions: Save or buy. And buy domestic or buy foreign.

Central to this second decision (to buy domestic or foreign goods) is the price of domestic goods relative to foreign goods. We call this relative price the **real exchange rate**. The real exchange rate is not directly observable, and you will not find it in the newspapers. What you will find in newspapers are *nominal exchange rates*, the relative prices of currencies. So we start by looking at nominal exchange rates and then see how we can use them to construct real exchange rates.

Nominal Exchange Rates

Nominal exchange rates between two currencies can be quoted in one of two ways:

- As the price of the domestic currency in terms of the foreign currency. If, for example, we look at the United States and the United Kingdom, and think of the dollar as the domestic currency and the pound as the foreign currency, we can express the nominal exchange rate as the price of a dollar in terms of pounds. In December 2018, the exchange rate defined this way was 0.79. In other words, one dollar was worth 0.79 pounds.

- As the price of the foreign currency in terms of the domestic currency. Continuing with the same example, we can express the nominal exchange rate as the price of a pound in terms of dollars. In December 2018, the exchange rate defined this way was 1.26. In other words, one pound was worth 1.26 dollars.

Warning: There is unfortunately no agreed-upon rule among economists or among newspapers as to which of the two definitions to use. You will encounter both. Always check which definition is used.

Either definition is fine; the important thing is to remain consistent. In this text, we shall adopt the first definition: we shall define the **nominal exchange rate** as *the price of the domestic currency in terms of foreign currency*, and denote it by E. When looking, for example, at the exchange rate between the United States and the United Kingdom (from the viewpoint of the United States, so the dollar is the domestic currency), E will denote the price of a dollar in terms of pounds (thus E was 0.79 in December 2018).

E: Nominal exchange rate—Price of domestic currency in terms of foreign currency. (From the point of view of the United States looking at the United Kingdom, the price of a dollar in terms of pounds.)

Exchange rates between the dollar and most foreign currencies are determined in foreign exchange markets and change every day—indeed every minute of the day. These changes are called *nominal appreciations* or *nominal depreciations*—appreciations or depreciations for short.

- An **appreciation** of the domestic currency is an increase in the price of the domestic currency in terms of a foreign currency. Given our definition of the exchange rate, an appreciation corresponds to an *increase* in the exchange rate.
- A **depreciation** of the domestic currency is a decrease in the price of the domestic currency in terms of a foreign currency. So given our definition of the exchange rate, a depreciation of the domestic currency corresponds to a decrease in the exchange rate, E.

Appreciation of the domestic currency ⇔ Increase in the price of the domestic currency in terms of foreign currency ⇔ Increase in the exchange rate.

Depreciation of the domestic currency ⇔ Decrease in the price of the domestic currency in terms of foreign currency ⇔ Decrease in the exchange rate.

You may have encountered two other words to denote movements in exchange rates: "revaluations" and "devaluations." These two terms are used when countries operate under **fixed exchange rates**—a system in which two or more countries maintain a constant exchange rate between their currencies. Under such a system, increases in the exchange rate—which are infrequent by definition—are called **revaluations** (rather than appreciations). Decreases in the exchange rate are called **devaluations** (rather than depreciations).

We shall discuss fixed exchange rates in Chapter 20.

Figure 17-3 plots the nominal exchange rate between the dollar and the pound since 1971. Note the two main characteristics of the figure:

- *The trend increase in the exchange rate.* In 1971, a dollar was worth only 0.41 pounds. In December 2018, a dollar was worth 0.79 pounds. Put another way, there was an appreciation of the dollar relative to the pound over the period.
- *The large fluctuations in the exchange rate.* In the 1980s, a sharp appreciation, in which the dollar more than doubled in value relative to the pound, was followed by a nearly equally sharp depreciation. In the 2000s, a large depreciation was followed by a large appreciation as the Financial Crisis started, and a smaller depreciation. In the 2010s, the exchange rate remained roughly unchanged until the Brexit referendum in June 2016 led to another exchange rate appreciation.

If we are interested, however, in the choice between domestic goods and foreign goods, the nominal exchange rate gives us only part of the information we need. Figure 17-3, for example, tells us only about movements in the relative price of the two currencies, the dollar and the pound. To US tourists thinking of visiting the United Kingdom, the question is not only how many pounds they will get in exchange for their dollars but how much goods will cost in the United Kingdom relative to how much they cost in the United States. This takes us to our next step—the construction of real exchange rates.

From Nominal to Real Exchange Rates

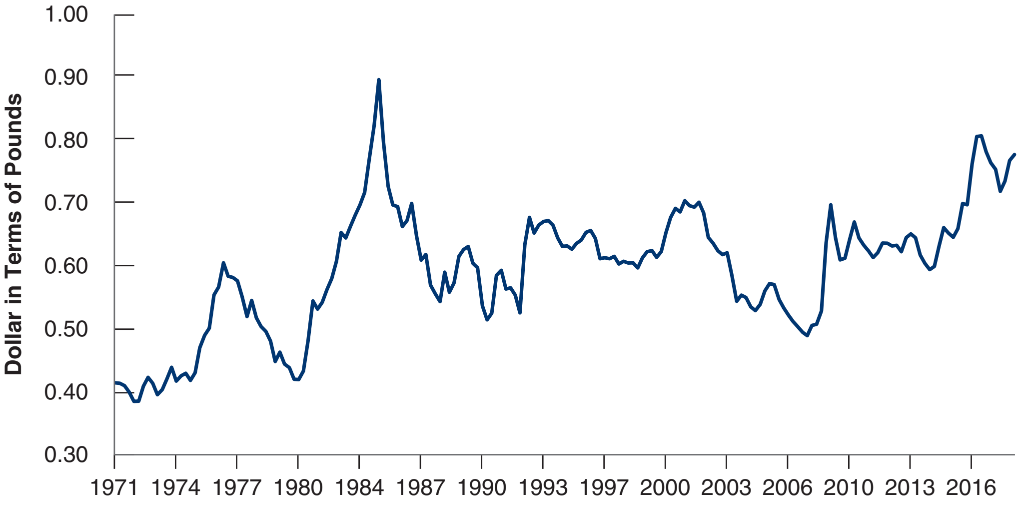
How can we construct the real exchange rate between the United States and the United Kingdom—the price of US goods in terms of British goods?

Figure 17-3

The Nominal Exchange Rate between the Dollar and the Pound since 1971

Although the dollar has appreciated relative to the pound over the past four decades, this appreciation has come with large swings in the nominal exchange rate between the two currencies.

Source: FRED DEXUSUK



Suppose the United States produced only one good, a Cadillac luxury sedan, and the United Kingdom also produced only one good, a Jaguar luxury sedan. (This is one of those "suppose" statements that run completely against the facts, but we shall become more realistic shortly.) Constructing the real exchange rate, the price of the US goods (Cadillacs) in terms of British goods (Jaguars) would be straightforward. We would express both goods in terms of the same currency and then compute their relative price.

Check that if we expressed both goods in terms of dollars instead, we would get the same result for the real exchange rate. ▶

Suppose, for example, we expressed both goods in terms of pounds.

- The first step would be to take the price of a Cadillac in dollars and convert it to a price in pounds. The price of a Cadillac in the United States is, say, \$40,000. The dollar is worth, say, £0.79, so the price of a Cadillac in pounds is $\$40{,}000 \times 0.79 = £31{,}600$.
- The second step would be to compute the ratio of the price of the Cadillac in pounds to the price of the Jaguar in pounds. The price of a Jaguar in the United Kingdom is, say, £30,000. So the price of a Cadillac in terms of Jaguars—that is, the real exchange rate between the United States and the United Kingdom—would be $£31{,}600/£30{,}000 = 1.05$. A Cadillac would be 5% more expensive than a Jaguar.

This example is straightforward, but how do we generalize it? The United States and the United Kingdom produce more than Cadillacs and Jaguars, and we want to construct a real exchange rate that reflects the relative price of *all* the goods produced in the United States in terms of *all* the goods produced in the United Kingdom.

The computation we just went through tells us how to proceed. Rather than using the price of a Jaguar and the price of a Cadillac, we must use a price index for all goods produced in the United Kingdom and a price index for all goods produced in the United States. This is exactly what the GDP deflators we introduced in Chapter 2 do. They are, by definition, price indexes for the set of final goods and services produced in the economy.

Let P be the GDP deflator for the United States, P^* be the GDP deflator for the United Kingdom (as a rule, we shall denote foreign variables by an asterisk, and E be the dollar-pound nominal exchange rate. Figure 17-4 goes through the steps needed to construct the real exchange rate.

ε: The real exchange rate is the price of domestic goods in terms of foreign goods. (For example, from the point of view of the United States looking at the United Kingdom, the price of US goods in terms of British goods.) ▶

- The price of US goods in dollars is P. Multiplying it by the exchange rate, E—the price of dollars in terms of pounds—gives us the price of US goods in pounds, EP.
- The price of British goods in pounds is P^*. The *real exchange rate*, the price of US goods in terms of British goods, which we shall call ε (the Greek lowercase epsilon), is thus given by

Figure 17-4

The Construction of the Real Exchange Rate

$$\varepsilon = \frac{EP}{P^*} \qquad (17.1)$$

The real exchange rate is constructed by multiplying the domestic price level by the nominal exchange rate and then dividing by the foreign price level—a straightforward extension of the computation we made in our Cadillac/Jaguar example.

Note, however, that there is an important difference between our Cadillac/Jaguar example and this more general computation. Unlike the price of Cadillacs in terms of Jaguars, the real exchange rate is an index number; that is, its level is arbitrary, and therefore uninformative. It is uninformative because the GDP deflators used to construct the real exchange rate are themselves index numbers. As we saw in Chapter 2, they are equal to 1 (or 100) in whatever year is chosen as the base year.

But all is not lost. Although the level of the real exchange rate is uninformative, the rate of change of the real exchange rate is informative. If, for example, the real exchange rate between the United States and the United Kingdom increases by 10%, this tells us that US goods are now 10% more expensive relative to British goods than they were before.

Like nominal exchange rates, real exchange rates move over time. These changes are called *real appreciations* or *real depreciations*.

- An increase in the real exchange rate—that is, an increase in the relative price of domestic goods in terms of foreign goods—is called a **real appreciation**.
- A decrease in the real exchange rate—that is, a decrease in the relative price of domestic goods in terms of foreign goods—is called a **real depreciation**.

Real appreciation ⇔ Increase in the price of the domestic goods in terms of foreign goods ⇔ Increase in the real exchange rate.

Real depreciation ⇔ Decrease in the price of the domestic goods in terms of foreign goods ⇔ Decrease in the real exchange rate.

Figure 17-5 plots the evolution of the real exchange rate between the United States and the United Kingdom since 1971, constructed using equation (17.1). For

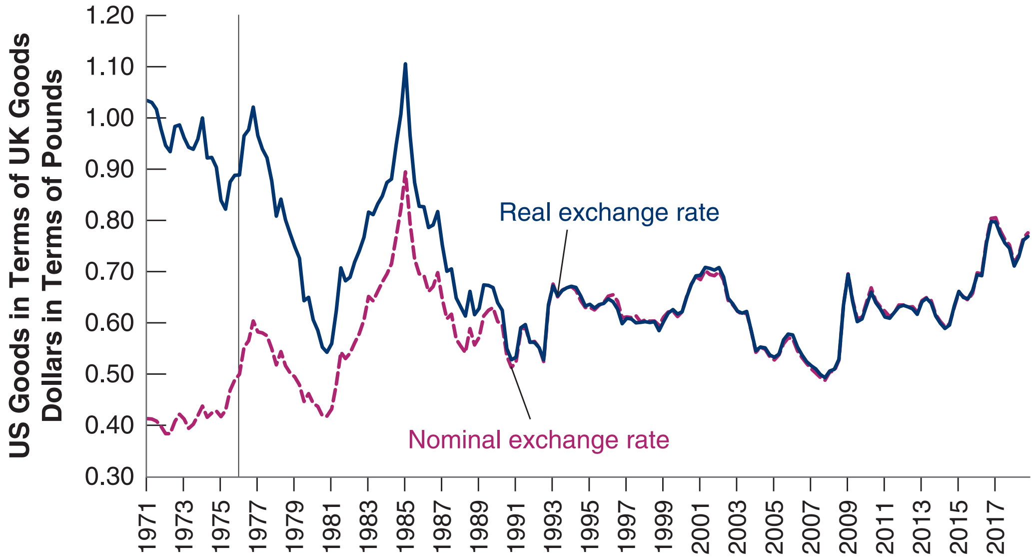


Figure 17-5

Real and Nominal Exchange Rates between the United States and the United Kingdom since 1971

Except for the difference in trend reflecting higher average inflation in the United Kingdom than in the United States until the early 1990s, the nominal and real exchange rates have largely moved together.

Source: FRED. GDPDEF, GBRGDPDEFQISMEI, DEXUSUK.

convenience, it also reproduces the evolution of the *nominal* exchange rate from Figure 17-3. The GDP deflators have both been set equal to 1 in the year 2000, so the nominal exchange rate and the real exchange rate are equal in that year by construction.

You should draw two lessons from Figure 17-5.

- The nominal and the real exchange rate can move in opposite directions. Note, for example, that, from 1971 to 1976, the nominal exchange rate went up, whereas the real exchange rate went down.

How do we reconcile the fact that there was both a nominal appreciation (of the dollar relative to the pound) and a real depreciation (of US goods relative to British goods) during the period? To see why, return to the definition of the real exchange rate in equation (17.1) and rewrite it as:

$$\varepsilon = E\frac{P}{P^*}$$

Two things happened in the 1970s:

First, E increased. The dollar went up in terms of pounds—this is the nominal appreciation we saw previously.

Second, P/P^* decreased. The price level increased less in the United States than in the United Kingdom. Put another way, over the period, average inflation was lower in the United States than in the United Kingdom.

The resulting decrease in P/P^* was larger than the increase in E, leading to a decrease in ε, a real depreciation—a decrease in the relative price of domestic goods in terms of foreign goods.

To get a better understanding of what happened, let's go back to our US tourists thinking about visiting the United Kingdom, circa 1976. They would find that they could buy more pounds per dollar than in 1971 (E had increased). Did this imply their trip would be cheaper? No. When they arrived in the United Kingdom, they would discover that the prices of goods in the United Kingdom had increased much more than the prices of goods in the United States (P^* has increased more than P, so P/P^* has declined), and this more than canceled the increase in the value of the dollar in terms of pounds. They would find that their trip was actually more expensive (in terms of US goods) than it would have been five years earlier.

There is a general lesson here. Over long periods of time, differences in inflation rates across countries can lead to very different movements in nominal exchange rates and real exchange rates. We shall return to this issue in Chapter 20.

Can there be a real appreciation with no nominal appreciation? Can there be a nominal appreciation with no real appreciation? (The answer to both questions: yes.)

- The large fluctuations in the nominal exchange rate we saw in Figure 17-3 also show up in the real exchange rate.
- This not surprising. Price levels move slowly. So year-to-year movements in the price ratio P/P^* are typically small compared to the often sharp movements in the nominal exchange rate E. Thus, from year to year, or even over the course of a few years, movements in the real exchange rate ε tend to be driven largely by movements in the nominal exchange rate E. Note that since the early 1990s, the nominal exchange rate and the real exchange rate have moved nearly together. This reflects the fact that, since the early 1990s, inflation rates have been similar—and low—in both countries.

If inflation rates were exactly equal, P/P^* would be constant, and ε and E would move exactly together.

From Bilateral to Multilateral Exchange Rates

We need to take one last step. So far, we have concentrated on the exchange rate between the United States and the United Kingdom. But the United Kingdom is just one of many countries the United States trades with. Table 17-2 shows the geographic composition of US trade for exports and imports.

The main message of the table is that the United States does most of its trading with three sets of countries. The first includes its immediate neighbors to the North and to the South: Trade with Canada and Mexico accounts for 34% of US exports and 26% of US imports. The second includes the European Union, which accounts for 19% of US exports and 19% of US imports. The third includes the Asian countries—including Japan and China—which together account for 26% of US exports and 36% of US imports.

Note also the sharp imbalance between US imports to China and US exports to China, which has been the source of tensions between China and the Trump administration. This has led the Trump administration to introduce tariffs on Chinese goods to decrease this bilateral deficit. More on this in Chapter 19.

How do we go from **bilateral exchange rates**, like the real exchange rate between the United States and the United Kingdom, to **multilateral exchange rates** that reflect this composition of trade?

Bi means "two." *Multi* means "many."

The principle we want to use is simple, even if the details of construction are complicated. We want the weight of a given country to incorporate not only how much the country trades with the United States but also how much it competes with the United States in other countries. (Why not just look at trade shares between the United States and each individual country? Take two countries, the United States and country A. Suppose the United States and country A do not trade with each other—so trade shares are equal to zero—but they are both exporting to country B. The real exchange rate between the United States and country A will matter very much for how much the United States exports to country B and thus to the US export performance.) The variable constructed in this way is called the **multilateral real US exchange rate** or the US real exchange rate for short.

These are all equivalent names for the relative price of US goods in terms of foreign goods: the multilateral real US exchange rate, the US **trade-weighted real exchange rate**, the US **effective real exchange rate**.

Figure 17-6 shows the evolution of this multilateral real exchange rate, the price of US goods in terms of foreign goods since 1973. Like the bilateral real exchange rates we saw a few pages earlier, it is an index number and its level is arbitrary. The index is roughly today where it was in 1973. What is most noticeable about the figure are the large swings in the multilateral real exchange rate in the 1980s and, to a lesser extent, in the 2000s. These swings are so striking that they have been given various names, from the "dollar cycle" to the more graphic "dance of the dollar." In the coming chapters, we shall examine where these swings come from and their effects on the trade deficit and economic activity.

The figure begins in 1973 because the multilateral real exchange rate, which is constructed by the Federal Reserve Board, is available only as far back as 1973.

Table 17-2 Country Composition of US Exports and Imports, 2018

	Percent of Exports to	Percent of Imports from
Canada	18	12
Mexico	16	14
European Union	19	19
China	7	21
Japan	4	6
Rest of Asia and Pacific	15	9
Others	21	19

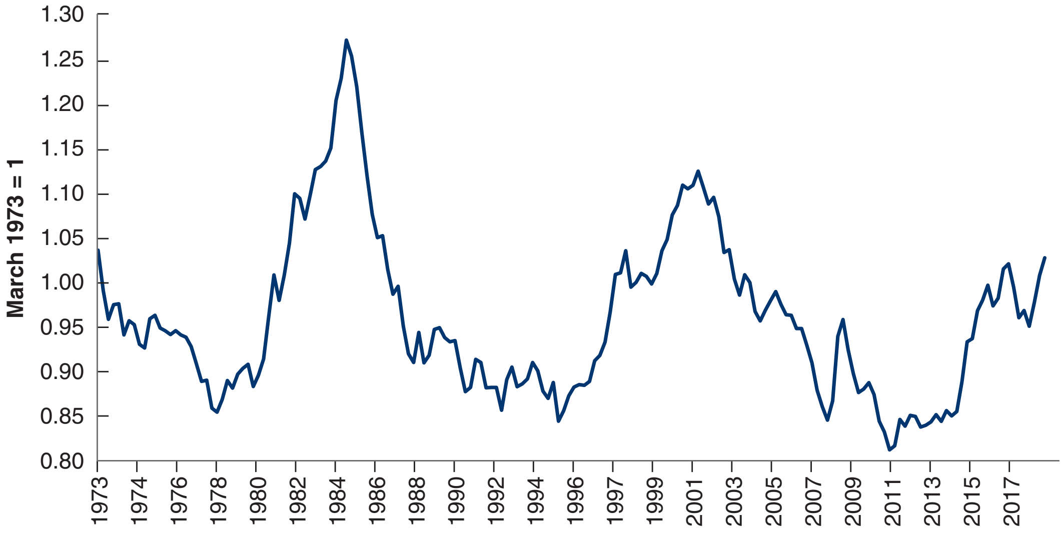
Source: US Census, International Trade Data, FT900, exhibit 14.

Figure 17-6

The US Multilateral Real Exchange Rate, since 1973

Since 1973 there have been two large real appreciations of the US dollar and two large real depreciations.

Source: FRED, TWEXBPA



17-2 OPENNESS IN FINANCIAL MARKETS

Openness in financial markets allows financial investors to hold both domestic assets and foreign assets, to diversify their portfolios, to speculate on movements in foreign interest rates versus domestic interest rates, on movements in exchange rates, and so on.

Diversify and speculate they do. Given that buying or selling foreign assets implies buying or selling foreign currency—sometimes called **foreign exchange**—the volume of transactions in foreign exchange markets gives a sense of the importance of international financial transactions. In 2016, for example, the recorded *daily* volume of foreign exchange transactions in the world was $5.1 trillion, of which 88% involved US dollars on one side of the transaction and 31% involved the euro.

Daily volume of foreign exchange transactions with dollars on one side of the transaction: $4.5 trillion. Daily volume of trade of the United States with the rest of the world: $13 billion (0.29% of the volume of foreign exchange transactions).

To get a sense of the magnitude of these numbers, the sum of US exports and imports in 2016 totaled about $13 billion per day. Suppose the only dollar transactions in foreign exchange markets had been, on one side, by US exporters selling their foreign currency earnings, and on the other side by US importers buying the foreign currency they needed to buy foreign goods. Then the volume of transactions involving dollars in foreign exchange markets would have been $13 billion per day, or about 0.29% of the actual daily total volume of dollar transactions ($\$5.1 \times 0.88 = \$4.5$ trillion) involving dollars in foreign exchange markets. This computation tells us that most of the transactions are associated not with trade but with purchases and sales of financial assets. Moreover, the volume of transactions in foreign exchange markets is not only high but also rapidly increasing: it has more than quintupled since 2001. Again, this increase in activity reflects mostly an increase in financial transactions rather than an increase in trade.

A country's openness in financial markets has an important implication. It allows the country to run trade surpluses and trade deficits. Recall that a country running a trade deficit is buying more from the rest of the world than it is selling to the rest of the world. To pay for the difference between what it buys and what it sells, the country must borrow from the rest of the world. It borrows by making it attractive for foreign financial investors to increase their holdings of domestic assets—in effect, to lend to the country.

Let's start by looking more closely at the relation between trade flows and financial flows. We shall then look at the determinants of these financial flows.

The Balance of Payments

A country's transactions with the rest of the world, including both trade flows and financial flows, are summarized by a set of accounts called the **balance of payments**. Table 17-3 presents the US balance of payments for 2018. The table has two parts, separated by a line. Transactions are referred to as either **above the line** or **below the line**.

The Current Account

The transactions above the line record payments to and from the rest of the world. They are called **current account** transactions.

- The first two lines record the exports and imports of goods and services. Exports lead to payments from the rest of the world, imports to payments to the rest of the world. The difference between exports and imports is the *trade balance*. In 2018, imports exceeded exports, leading to a US *trade deficit* of \$622 billion—roughly 3% of US GDP.
- Exports and imports are not the only sources of payments to and from the rest of the world. US residents receive income on their holdings of foreign assets, and foreign residents receive income on their holdings of US assets. In 2018, income received from the rest of the world was \$1,200 billion and income paid to foreigners was \$1,067 billion for a net **income balance** of \$133 billion.

The sum of these net payments to and from the rest of the world is called the **current account balance**. If net payments from the rest of the world are positive, the country is running a **current account surplus**; if they are negative, the country is running a **current account deficit**. Adding all payments to and from the rest of the world, net payments from the United States to the rest of the world were equal in 2018 to $-\$622 + \$133 = -\$489$ billion. Put another way, in 2018 the United States ran a current account deficit of \$489 billion—roughly 2.4% of its GDP.

Can a country have: A trade deficit and no current account deficit? A current account deficit and no trade deficit? (The answer to both questions: yes.)

Table 17-3 The US Balance of Payments, 2018, in Billions of US Dollars

Current Account		
Exports	2,500	
Imports	3,122	
Trade balance (deficit = minus sign) (1)		− 622
Income received	1,200	
Income paid	1,067	
Net income (2)		133
Current account balance (1) + (2) (deficit = minus sign)		− 489
Financial Account		
Net capital transfers (3)	9	
Increase in foreign holdings of US assets (4)	811	
Increase in US holdings of foreign assets (5)	301	
Financial account balance (7) = (3) + (4) − (5)		519
Statistical discrepancy: financial account − current account balance		30

Source: US Bureau of Economic Analysis, US International Transactions, Table. 17.1.

The Financial Account

Until recently, the financial account was called the **capital account**. It is still called the capital account in many countries and in the press. Technically, in the United States today the "capital account" refers only to the net capital transfers in Table 17-3. Yes, it is confusing!

The fact that the United States had a current account deficit of $489 billion in 2018 implies that it had to borrow $489 billion from the rest of the world—or, equivalently, that net foreign holdings of US assets had to increase by $489 billion. The numbers below the line show how this was achieved. They are called **financial account** transactions.

One way a country can finance its deficit is by receiving gifts from other countries. An example of such gifts may be a cancellation of some of the country's debt. This is reflected in the line called "net capital transfers." For a country like the United States, this is not an important item, and the number for net capital transfers is close to zero. Nearly all the financing comes from an increase in net foreign holdings of US assets.

In the same way that if you spend more than you earn, you have to finance the difference.

The increase in foreign holdings of US assets in 2018 was $811 billion. Foreign investors—be they private investors, governments, or central banks—bought $811 billion worth of US stocks, bonds, and other US assets.

At the same time, however, there was an increase in US holdings of foreign assets of $301 billion. US investors, private and public, bought $301 billion worth of foreign stocks, bonds, and other assets.

The result was an increase in net US foreign indebtedness (the increase in foreign holdings of US assets, plus net capital transfers, minus the increase in US holdings of foreign assets), also called **net capital flows** to the United States, of $\$811 + \$9 - \$301 = \$519$ billion. Another name for net capital flows is the **financial account balance**. Positive net capital flows are called a **financial account surplus**; negative net capital flows are called a **financial account deficit**. So, put another way, in 2018 the United States ran a financial account surplus of $519 billion.

A country that runs a current account deficit must finance it through positive net capital flows. Equivalently, it must run a financial account surplus.

Shouldn't net capital flows (equivalently, the financial account surplus) plus net capital transfers be equal to the current account deficit (which was equal to $489 billion in 2018)? In principle, yes. In practice, no.

The numbers for current and capital account transactions are constructed using different sources; although they should give the same answers, they typically do not. In 2018, the difference between the two—called the **statistical discrepancy**—was $30 billion, about 6% of the current account balance. This is yet another reminder that, even for an advanced country such as the United States, economic data are far from perfect. (This problem of measurement manifests itself in another way as well. The sum of the current account deficits of all the countries in the world should be equal to zero. One country's deficit should show up as a surplus for the other countries taken as a whole. However, this is not the case in the data. If we just add the published current account deficits of all the countries in the world, it would appear that the world is running a large current account deficit!)

Some economists speculate that the explanation lies in unrecorded trade with Martians. Most others believe that mismeasurement is the explanation.

Now that we have looked at the current account, we can return to an issue we touched on in Chapter 2, the difference between gross domestic product (GDP), the measure of output we have used so far, and **gross national product (GNP)**, another measure of aggregate output.

GDP measures *value added domestically*. GNP measures the *value added by domestic factors of production*. When the economy is closed, the two measures are the same. When the economy is open, however, they can differ. Some of the income from domestic production goes to foreigners; and domestic residents receive some foreign income. Thus, to go from GDP to GNP, one must start from GDP, add income received from the rest of the world, and subtract income paid to the rest of the world. Put another way, GNP is equal to GDP plus net payments from the rest of the world. More formally, denoting these net income payments by NI,

$$GNP = GDP + NI$$

In most countries, the difference between the GNP and GDP is small (relative to GDP). For example, in the United States, you can see from Table 17-3 that net income payments were equal to \$133 billion. GNP exceeded GDP by \$133 billion, or about 0.6% of GDP. For some countries, however, the difference can be large. This is explored in the Focus Box "GDP versus GNP: The Example of Kuwait."

The Choice between Domestic and Foreign Assets

Openness in financial markets implies that people (or financial institutions that act on their behalf) face a new financial decision: whether to hold domestic assets or foreign assets.

It would appear that we actually have to think about at least *two* new decisions: the choice of holding *domestic* money versus *foreign* money and that of holding *domestic* interest-paying assets versus *foreign* interest-paying assets. But remember why people hold money: to engage in transactions. For someone who lives in the United States and whose transactions are mostly or fully in dollars, there is little point in holding foreign currency. Foreign currency cannot be used for transactions in the United States, and if the goal is to hold foreign assets, holding foreign currency is clearly less desirable than holding foreign bonds, which pay interest. This leaves us with only one new decision: the choice between domestic and foreign interest-paying assets.

Two qualifications, from Chapter 4:

- Foreigners involved in illegal activities often hold dollars because dollars can be exchanged easily and cannot be traced.
- In times of high inflation, people sometimes switch to a foreign currency—in other countries, often the dollar—for use even in some domestic transactions.

For now, let's think of these assets as domestic and foreign one-year bonds. Consider, for example, the choice between US and UK one-year bonds from the point of view of a US investor.

- Suppose you decide to hold US bonds.
- Let i_t be the one-year US nominal interest rate. Then, as Figure 17-7 shows for every dollar you put in US bonds, you will get $(1 + i_t)$ dollars next year. (This is represented by the arrow pointing to the right at the top of the figure.)
- Suppose you decide instead to hold UK bonds.
- To buy UK bonds, you must first buy pounds. Let E_t be the nominal exchange rate between the dollar and the pound. For every dollar, you get E_t pounds. (This is represented by the arrow pointing downward in the figure.)

Let i_t^* denote the one-year nominal interest rate on UK bonds (in pounds). When next year comes, you will have $E_t(1 + i_t^*)$ pounds. (This is represented by the arrow pointing to the right at the bottom of the figure.)

You will then have to convert your pounds back into dollars. If you expect the nominal exchange rate next year to be E_{t+1}^e, each pound will be worth $(1/E_{t+1}^e)$ dollars. So you can expect to have $E_t(1 + i_t^*)(1/E_{t+1}^e)$ dollars next year for every dollar you invest now. (This is represented by the arrow pointing upward in the figure.)

We shall look at the expression we just derived in more detail soon. But note its basic implication already. In assessing the attractiveness of UK versus US bonds, you cannot look just at the UK interest rate and the US interest rate; you must also assess what you think will happen to the dollar/pound exchange rate between this year and next.

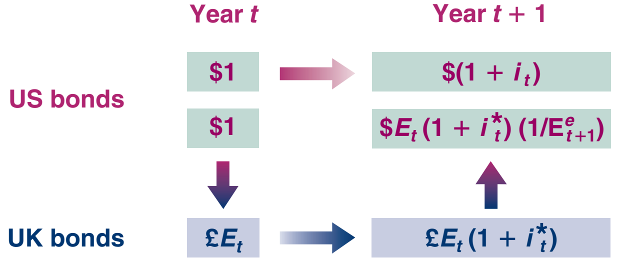


Figure 17-7

Expected Returns from Holding One-Year US Bonds versus One-Year UK Bonds

FOCUS

GDP versus GNP: The Example of Kuwait

When oil was discovered in Kuwait, Kuwait's government decided that a portion of oil revenues would be saved and invested abroad rather than spent, to provide future Kuwaiti generations with income when oil revenues came to an end. Kuwait ran a large current account surplus, steadily accumulating large foreign assets. As a result, it had large holdings of foreign assets and received substantial income from the rest of the world. Table 1 gives GDP, GNP, and net investment income for Kuwait from 1989 to 1994 (you will see the reason for the choice of dates).

Note how much larger GNP was compared to GDP throughout the period. Net income from abroad was 34% of GDP in 1989. But note also how net factor payments steadily decreased after 1989. This is because Kuwait had to pay its allies for part of the cost of the 1990–1991 Gulf War and also had to pay for reconstruction after the war. It did so by running a current account deficit—that is, by decreasing its net holdings of foreign assets. This in turn led to a decrease in the income it earned from foreign assets and, by implication, a decrease in its net factor payments.

Since the Gulf War, Kuwait has rebuilt a sizable net foreign asset position. Net income from abroad was 6% of GDP in 2018.

Table 1 GDP, GNP, and Net Income in Kuwait, 1989–1994

Year	GDP	GNP	Net Income (NI)
1989	7,143	9,616	2,473
1990	5,328	7,560	2,232
1991	3,131	4,669	1,538
1992	5,826	7,364	1,538
1993	7,231	8,386	1,155
1994	7,380	8,321	941

Source: International Financial Statistics, IMF. All numbers are in millions of Kuwaiti dinars. 1 dinar = \$3.0. (2018).

The word *uncovered* is to distinguish this relation from another relation called the *covered interest parity condition*, which is derived by looking at the following choice: Buy and hold US bonds for one year; or buy pounds today, buy one-year UK bonds with the proceeds, and agree to sell the pounds for dollars a year ahead at a predetermined price, called the *forward exchange rate*. The rate of return on these two alternatives, which can both be realized at *no risk today*, must be the same. The covered interest parity condition is a *riskless arbitrage* condition. It typically holds tightly.

Let's now make the same assumption we made in Chapter 14 when first discussing the choice between short- and long-term bonds. Let's assume that you and other financial investors care only about *the expected rate of return*, ignoring differences in risk, and therefore want to hold only the asset with the highest expected rate of return. In this case, if both UK and US bonds are to be held, they must have the same expected rate of return. Arbitrage implies that the following relation must hold:

$$(1 + i_t) = (E_t)(1 + i_t^*)\left(\frac{1}{E_{t+1}^e}\right)$$

Reorganizing,

$$(1 + i_t) = (1 + i_t^*)\left(\frac{E_t}{E_{t+1}^e}\right) \tag{17.2}$$

Equation (17.2) is called the **uncovered interest parity relation** or simply the **interest parity condition**.

The assumption that financial investors will hold only the bonds with the highest expected rate of return is obviously too strong, for two reasons:

- It ignores transaction costs. Buying and selling UK bonds requires three separate transactions, each with a transaction cost.
- It ignores risk. The exchange rate a year from now is uncertain. For the US investor, holding UK bonds is therefore more risky, in terms of dollars, than holding US bonds.

Whether holding UK bonds or US bonds is more risky depends on which investors we are looking at. Holding UK bonds is riskier from the point of view of US investors. Holding US bonds is riskier from the point of view of British investors. (Why?)

But as a characterization of capital movements among the major world financial markets (New York, Frankfurt, London, and Tokyo), the assumption is not far off. Small changes in interest rates and rumors of impending appreciation or depreciation can lead to movements of billions of dollars within minutes. For the rich countries of the world, the arbitrage assumption in equation (17.2) is a good approximation of reality. Other countries whose capital markets are smaller and less developed, or countries that have various forms of capital controls, have more leeway in choosing their domestic interest rate than is implied by equation (17.2). We shall return to this issue at the end of Chapter 20.

Interest Rates and Exchange Rates

Let's get a better sense of what the interest parity condition implies. First rewrite E_t/E^e_{t+1} as $1/(1 + (E^e_{t+1} - E_t)/E_t)$. Replacing in equation (17.2) gives

$$(1 + i_t) = \frac{(1 + i^*_t)}{[1 + (E^e_{t+1} - E_t)/E_t]} \qquad (17.3)$$

This gives us a relation between the domestic nominal interest rate, i_t, the foreign nominal interest rate, i^*_t, and the expected rate of appreciation of the domestic currency, $(E^e_{t+1} - E_t)/E_t$. As long as interest rates or the expected rate of depreciation are not too large—say, below 20% a year—a good approximation to this equation is given by:

This follows from Proposition 3 in Appendix 2 at the end of the book.

$$i_t \approx i^*_t - \frac{E^e_{t+1} - E_t}{E_t} \qquad (17.4)$$

This is the form of the *interest parity condition* you must remember. Arbitrage by investors implies that *the domestic interest rate must be equal to the foreign interest rate minus the expected appreciation rate of the domestic currency.*

Note that the expected appreciation rate of the domestic currency is also the expected depreciation rate of the foreign currency. So equation (17.4) can be equivalently stated as saying that *the domestic interest rate must be equal to the foreign interest rate minus the expected depreciation rate of the foreign currency*.

If the dollar is expected to appreciate by 3% relative to the pound, then the pound is expected to depreciate by 3% relative to the dollar.

Let's apply this equation to US bonds versus UK bonds. Suppose the one-year nominal interest rate is 2.0% in the United States and 5.0% in the United Kingdom. Should you hold UK bonds or US bonds?

- It depends whether you expect the pound to depreciate relative to the dollar over the coming year by more or less than the difference between the US interest rate and the UK interest rate, or 3.0% in this case (5.0% − 2.0%).
- If you expect the pound to depreciate by more than 3.0%, then, despite the fact that the interest rate is higher in the United Kingdom than in the United States, investing in UK bonds is less attractive than investing in US bonds. By holding UK bonds, you will get higher interest payments next year, but the pound will be worth less in terms of dollars, making it less attractive to invest in UK bonds than in US bonds.
- If you expect the pound to depreciate by less than 3.0% or even to appreciate, then the reverse holds, and UK bonds are more attractive than US bonds.

Looking at it another way: If the uncovered interest parity condition holds and the US one-year interest rate is 3% lower than the UK interest rate, it must be that financial investors are expecting, on average, an appreciation of the dollar relative to the pound over the coming year of about 3% and this is why they are willing to hold US bonds despite their lower interest rate. (Another—and more striking—example is provided in the Focus Box "Buying Brazilian Bonds.")

The arbitrage relation between interest rates and exchange rates, in the form of either equation (17.2) or equation (17.4), will play a central role in the following chapters. It suggests that, unless countries are willing to tolerate large movements in their exchange rate, domestic and foreign interest rates are likely to move largely together.

FOCUS

Buying Brazilian Bonds

Put yourself back in September 1993 (the very high interest rate in Brazil at the time helps make the point I want to get across here). Brazilian bonds are paying a monthly interest rate of 36.9%! This seems attractive compared to the annual rate of 3% on US bonds—corresponding to a monthly interest rate of about 0.2%. Shouldn't you buy Brazilian bonds?

The discussion in this chapter tells you that to decide you need one more crucial element, the expected rate of depreciation of the cruzeiro (the name of the Brazilian currency at the time; the currency is now called the real) in terms of dollars.

You need this information because, as we saw in equation (17.4), the return in dollars from investing in Brazilian bonds for a month is equal to one plus the Brazilian interest rate, divided by one plus the expected rate of depreciation of the cruzeiro relative to the dollar:

$$\frac{1 + i_t^*}{[1 + (E_{t+1}^e - E_t)/E_t]}$$

What rate of depreciation of the cruzeiro should you expect over the coming month? A reasonable first pass is to expect the rate of depreciation during the coming month to be equal to the rate of depreciation last month. The dollar was worth 100,000 cruzeiros at the end of July 1993 and worth 134,600 cruzeiros at the end of August 1993, so the rate of appreciation of the dollar relative to the cruzeiro—equivalently, the rate of depreciation of the cruzeiro relative to the dollar—in August was 34.6%. If depreciation is expected to continue at the same rate in September as it did in August, the expected return from investing in Brazilian bonds for one month is

$$\frac{1.369}{1.346} = 1.017$$

The expected rate of return in dollars from holding Brazilian bonds is only $(1.017 - 1) = 1.7\%$ per month, not the 36.9% per month that initially looked so attractive. Note that 1.7% per month is still higher than the monthly interest rate on US bonds (about 0.2%). But think of the risk and the transaction costs—all the elements we ignored when we wrote the arbitrage condition. When these are taken into account, you may well decide not to buy the Brazilian bonds.

Take the extreme case of two countries that commit to maintaining their bilateral exchange rates at a fixed value. If markets have faith in this commitment, they will expect the exchange rate to remain constant and the expected depreciation will be equal to zero. In this case, the arbitrage condition implies that interest rates in the two countries will have to move exactly together. Most of the time, as we shall see, governments do not make such absolute commitments to maintain the exchange rate, but they often do try to avoid large movements in the exchange rate. This puts sharp limits on how much they can allow their interest rate to deviate from interest rates elsewhere in the world.

If $E_{t+1}^e = E_t$, then the interest parity condition implies $i_t = i_t^*$.

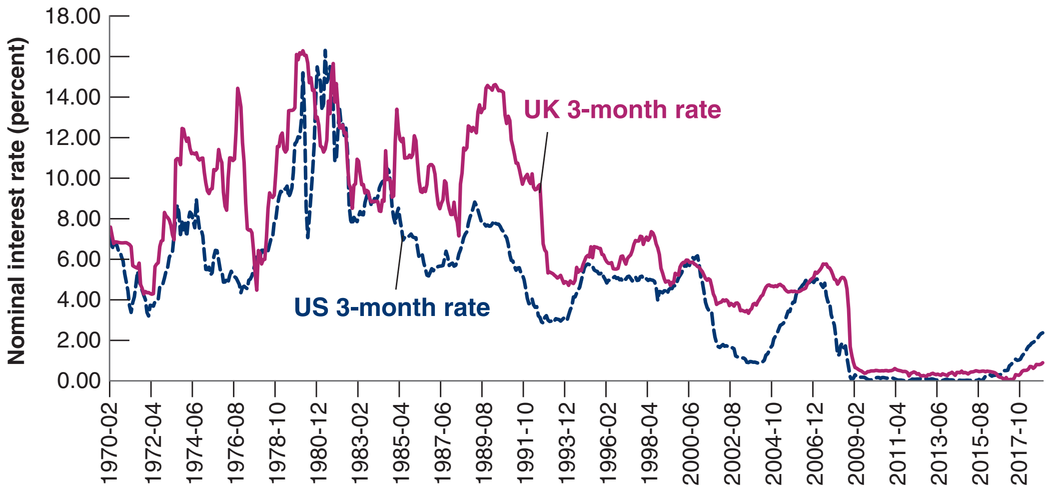
How much do nominal interest rates actually move together in major countries? Figure 17-8 plots the three-month nominal interest rate in the United States and United Kingdom (expressed at annual rates) since 1970. The figure shows that the movements

Figure 17-8

Three-Month Nominal Interest Rates in the United States and United Kingdom since 1970

US and UK nominal interest rates have largely moved together over the last 40 years.

Source: FRED TB3MS; IR3TTS01GBM156N



are related but not identical. Interest rates were high in both countries in the early 1980s, and high again—although much more so in the United Kingdom than in the United States—in the late 1980s. They have been low in both countries since the mid-1990s, and are both very low and close to each other today. (Note how close to the zero lower bound the rate has been in both countries since mid-2009.) At the same time, differences between the two have sometimes been quite large. In 1990, for example, the UK interest rate was nearly 7% higher than the US interest rate. In the coming chapters, we shall return to why such differences emerge and what their implications may be.

Meanwhile, do the following: Look at the back pages of a recent issue of *The Economist* for short-term interest rates in different countries relative to the United States. Assume that uncovered interest parity holds. Which currencies are expected to appreciate against the dollar?

17-3 CONCLUSIONS AND A LOOK AHEAD

We have now set the stage for the study of the open economy:

- Openness in goods markets allows people and firms to choose between domestic goods and foreign goods. This choice depends primarily on the *real exchange rate*—the relative price of domestic goods in terms of foreign goods.
- Openness in financial markets allows investors to choose between domestic assets and foreign assets. This choice depends primarily on their relative rates of return, which depend on domestic and foreign interest rates and on the expected rate of appreciation of the domestic currency.

In Chapter 18 we look at the implications of openness in goods markets. In Chapter 19, we explore the implications of openness in both goods and financial markets. In Chapter 20, we discuss the pros and cons of different exchange rate regimes.

SUMMARY

- Openness in goods markets allows people and firms to choose between domestic and foreign goods. Openness in financial markets allows financial investors to hold domestic or foreign financial assets.
- The nominal exchange rate is the price of the domestic currency in terms of foreign currency. From the viewpoint of the United States, the nominal exchange rate between the United States and the United Kingdom is the price of a dollar in terms of pounds.
- A nominal appreciation (an appreciation, for short) is an increase in the price of the domestic currency in terms of foreign currency. In other words, it corresponds to an increase in the exchange rate. A nominal depreciation (a depreciation, for short) is a decrease in the price of the domestic currency in terms of foreign currency. It corresponds to a decrease in the exchange rate.
- The real exchange rate is the relative price of domestic goods in terms of foreign goods. It is equal to the nominal exchange rate times the domestic price level divided by the foreign price level.
- A real appreciation is an increase in the relative price of domestic goods in terms of foreign goods (i.e., an increase in the real exchange rate). It may come either from a nominal appreciation or from domestic inflation being higher than foreign inflation. A real depreciation is a decrease in the relative price of domestic goods in terms of foreign goods (i.e., a decrease in the real exchange rate). It may come either from a nominal depreciation or from domestic inflation being lower than foreign inflation.
- The multilateral real exchange rate, or real exchange rate for short, is a weighted average of bilateral real exchange rates.
- The balance of payments records a country's transactions with the rest of the world. The current account balance is equal to the sum of the trade balance and net income the country receives from the rest of the world. The financial account balance is equal to net capital transfers plus capital flows from the rest of the world minus capital flows to the rest of the world.
- The current account and the financial account are mirror images of each other. Leaving aside statistical problems, the current account plus the financial account must sum to zero. A current account deficit is financed by net capital flows from the rest of the world, thus by a capital account

surplus. Similarly, a current account surplus corresponds to a capital account deficit.

- Uncovered interest parity, or interest parity for short, is an arbitrage condition stating that the expected rates of return in terms of domestic currency on domestic and foreign bonds must be equal. Interest parity implies that the domestic interest rate approximately equals the foreign interest rate minus the expected nominal appreciation rate of the domestic currency.

KEY TERMS

openness in goods markets, 355
tariffs, 355
quotas, 355
openness in financial markets, 355
capital controls, 355
openness in factor markets, 355
North American Free Trade Agreement (NAFTA), 355
tradable goods, 357
real exchange rate, 358
nominal exchange rate, 359
appreciation (nominal), 359
depreciation (nominal), 359
fixed exchange rates, 359
revaluation, 359
devaluation, 359
real appreciation, 361
real depreciation, 361
bilateral exchange rate, 363
multilateral exchange rate, 363
multilateral real US exchange rate, 363
trade-weighted real exchange rate, 363
effective real exchange rate, 363
foreign exchange, 364
balance of payments, 365
above the line, 365
below the line, 365
current account, 365
income balance, 365
current account balance, 365
current account surplus, 365
current account deficit, 365
financial account, 366
capital account, 366
net capital flows, 366
financial account balance, 366
financial account surplus, 366
financial account deficit, 366
statistical discrepancy, 366
gross national product (GNP), 366
uncovered interest parity relation, 368
interest parity condition, 368

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. If there are no statistical discrepancies, countries with current account deficits must receive net capital inflows.

b. Although the export ratio can be larger than one—as it is in Singapore—the same cannot be true of the ratio of imports to GDP.

c. The fact that a rich country like Japan has such a small ratio of imports to GDP is clear evidence of an unfair playing field for US exporters to Japan.

d. Uncovered interest parity implies that interest rates must be the same across countries.

e. The nominal exchange rate in this chapter is defined as the domestic currency price of a unit of foreign currency.

f. The nominal exchange rate and the real exchange rate always move in the same direction.

g. The nominal exchange rate and the real exchange rate usually move in the same direction.

h. If the dollar is expected to appreciate against the yen, uncovered interest parity implies that the US nominal interest rate must be greater than the Japanese nominal interest rate.

i. Given the definition of the exchange rate adopted in this chapter, if the dollar is the domestic currency and the euro the foreign currency, a nominal exchange rate of 0.75 means that 0.75 dollars is worth one euro.

j. A real appreciation means that domestic goods become less expensive relative to foreign goods.

2. Consider two fictional economies, one called the domestic country and the other the foreign country. Given the transactions listed in (a) through (g), construct the balance of payments for each country. If necessary, include a statistical discrepancy.

a. The domestic country purchased \$100 in oil from the foreign country.

b. Foreign tourists spent \$25 on domestic ski slopes.

c. Foreign investors were paid \$15 in dividends from their holdings of domestic equities.

d. Domestic residents gave \$25 to foreign charities.

e. Domestic businesses borrowed \$65 from foreign banks.

f. Foreign investors purchased \$15 of domestic government bonds.

g. Domestic investors sold $50 of their holdings of foreign government bonds.

3. Consider two bonds, one issued in euros (€) in Germany, and one issued in dollars ($) in the United States. Assume that both government securities are one-year bonds—paying the face value of the bond one year from now. The exchange rate, E, stands at 0.75 euros per dollar.

The face values and prices on the two bonds are given by

	Face Value	Price
United States	$10,000	$9,615.38
Germany	€10,000	€9,433.96

a. Compute the nominal interest rate on each of the bonds.
b. Compute the expected exchange rate next year consistent with uncovered interest parity.
c. If you expect the dollar to depreciate relative to the euro, which bond should you buy?
d. Assume that you are a US investor and you exchange dollars for euros and purchase the German bond today. One year from now, it turns out that the exchange rate, E, is actually 0.72 (0.72 euros buys one dollar). What is your realized rate of return in dollars compared to the realized rate of return you would have made had you held the US bond?
e. Are the differences in rates of return in part d consistent with the uncovered interest parity condition? Why or why not?

DIG DEEPER

4. Consider a world with three equal-sized economies (A, B, and C) and three goods (clothes, cars, and computers). Assume that consumers in all three economies want to spend an equal amount on all three goods.

The value of production of each good in the three economies is given in the table below.

	A	B	C
Clothes	10	0	5
Cars	5	10	0
Computers	0	5	10

a. What is GDP in each economy? If the total value of GDP is consumed and no country borrows from abroad, how much will consumers in each economy spend on each of the goods?
b. If no country borrows from abroad, what will be the trade balance in each country? What will be the pattern of trade in this world (i.e., which good will each country export and to whom)?
c. Given your answer to part b, will country A have a zero trade balance with country B? with country C? Will any country have a zero trade balance with any other country?
d. The United States has a large trade deficit. It has a trade deficit with each of its major trading partners, but the deficit is much larger with some countries (e.g., China) than with others. Suppose the United States eliminates its overall trade deficit (with the world as a whole). Do you expect it to have a zero trade balance with every one of its trading partners? Does the especially large trade deficit with China necessarily indicate that China does not allow US goods to compete on an equal basis with Chinese goods?

5. The exchange rate and the labor market

Suppose the domestic currency depreciates (i.e., E falls). Assume that P and P remain constant.*

a. How does the nominal depreciation affect the relative price of domestic goods (i.e., the real exchange rate)? Given your answer, what effect would a nominal depreciation likely have on (world) demand for domestic goods? on the domestic unemployment rate?
b. Given the foreign price level, P^*, what is the price of foreign goods in terms of domestic currency? How does a nominal depreciation affect the price of foreign goods in terms of domestic currency? How does a nominal depreciation affect the domestic consumer price index? (*Hint*: Remember that domestic consumers buy foreign goods (imports) as well as domestic goods.)
c. If the nominal wage remains constant, how does a nominal depreciation affect the real wage?
d. Comment on the following statement. "A depreciating currency puts domestic labor on sale."

EXPLORE FURTHER

6. Retrieve the nominal exchange rates between Japan and the United States from the Federal Reserve Bank of St. Louis FRED data site. It is series AEXJPUS. This exchange rate is written as yen per dollar.

a. In the terminology of the chapter, when the exchange rate is written as yen per dollar, which country is being treated as the domestic country?
b. Plot the number of yen per dollar since 1971. During which time period(s) did the yen appreciate? During which period(s) did the yen depreciate?
c. Given the long-running Japanese slump, one way of increasing demand would be to make Japanese goods more attractive. Does this require an appreciation or a depreciation of the yen?
d. What has happened to the yen during the past few years? Has it appreciated or depreciated? Is this good or bad for Japan?

7. Retrieve the most recent World Economic Outlook (WEO) from the Web site of the International Monetary Fund (www.imf.org). In the Statistical Appendix, find the table titled "Balances on Current Account," which lists current account balances around the world. Use the data for the most recent year available to answer parts a through c.

a. Note the sum of current account balances around the world. As noted in the chapter, the sum of current account balances should equal zero. What does this sum actually equal? Why does this sum indicate some mismeasurement (i.e., if the sum were correct, what would it imply)?
b. Which regions of the world are borrowing and which are lending?
c. Compare the US current account balance to the current account balances of the other advanced economies. Do other advanced economies lend enough to cover the entire US current account balance?
d. The statistical tables in the WEO typically project data into the future. Look at the projected data on current account balances. Do your answers to parts b and c seem likely to change in the near future?

8. Saving and investment throughout the world

Retrieve the most recent World Economic Outlook (WEO) from the Web site of the International Monetary Fund (www.imf.org). In the Statistical Appendix, find the table titled "Summary of Net Lending and Borrowing," which lists saving and investment (as a percentage of GDP) around the world. Use the data for the most recent year available to answer parts a and b.

a. Does world saving equal investment? (You may ignore small statistical discrepancies.) Offer some intuition for your answer.
b. How does US saving compare to US investment? How is the United States able to finance its investment? *(We explain this explicitly in the next chapter, but your intuition should help you figure it out now.)*
c. From the FRED economic database download real GDP (variable GDPC1) and real GNP (variable GNPC96) for the years 1947 to the latest data. Calculate the percentage difference between GNP and GDP in the United States. Which is larger? Why is that the case?

FURTHER READINGS

- If you want to learn more about international trade and international economics, a very good textbook is Paul Krugman, Marc Melitz, and Maurice Obstfeld's *International Economics, Theory and Policy*, 10th ed. (Prentice Hall, 2014).
- If you want to know current exchange rates between nearly any pair of currencies in the world, look at the "currency converter" on www.oanda.com/currency/converter/.

18

The Goods Market in an Open Economy

In 2009, countries around the world worried about the risk of a recession in the United States, but their worries were not so much for the United States as they were for themselves. To them, a US recession meant lower exports to the United States, a deterioration of their trade position, and weaker growth at home.

Were their worries justified? Figure 17-1 from the previous chapter certainly suggested they were. The US recession clearly led to a world recession. To understand what happened, we must expand the treatment of the goods market in Chapter 3 and account for openness in the analysis of goods markets. This is what we do in this chapter.

Section 18-1 characterizes equilibrium in the goods market for an open economy.

Sections 18-2 and 18-3 show the effects of domestic shocks and foreign shocks on the domestic economy's output and trade balance.

Section 18-4 looks at the effect of a real depreciation on output and the trade balance.

Section 18-5 gives an alternative description of the equilibrium that shows the close connection among saving, investment, and the trade balance. (It is a really important section. You will think differently about open economy issues after you have read it.)

If you remember one basic message from this chapter, it should be: Output depends on both domestic and foreign demand. ▶▶▶

18-1 THE IS RELATION IN THE OPEN ECONOMY

When we were assuming the economy was closed to trade, there was no need to distinguish between the *domestic demand for goods* and the *demand for domestic goods*; they were the same thing. Now we must distinguish between the two. Some domestic demand falls on foreign goods, and some of the demand for domestic goods comes from foreigners. Let's look at this distinction more closely.

"The domestic demand for goods" and "the demand for domestic goods" sound similar but are not the same thing. Part of domestic demand falls on foreign goods. Part of foreign demand falls on domestic goods.

The Demand for Domestic Goods

In an open economy, the **demand for domestic goods**, Z, is given by

$$Z = C + I + G - IM/\varepsilon + X \tag{18.1}$$

The first three terms—consumption, C, investment, I, and government spending, G—constitute the total **domestic demand for goods**, domestic or foreign. If the economy were closed, $C + I + G$ would also be the demand for domestic goods. This is why, until now, we have only looked at $C + I + G$. But now we have to make two adjustments:

- First, we must subtract imports—the part of domestic demand that falls on foreign goods rather than on domestic goods.

  We must be careful here. Foreign goods are different from domestic goods, so we cannot just subtract the quantity of imports, IM. If we were to do so, we would be subtracting apples (foreign goods) from oranges (domestic goods). We must first express the value of imports in terms of domestic goods. This is what IM/ε in equation (18.1) stands for. Recall from Chapter 17 that ε, the real exchange rate, is defined as the price of domestic goods in terms of foreign goods. Equivalently, $1/\varepsilon$ is the price of foreign goods in terms of domestic goods. So $IM\,(1/\varepsilon)$—or, equivalently, IM/ε—is the value of imports in terms of domestic goods.
- Second, we must add exports—the part of demand for domestic goods that comes from abroad. This is captured by the term X in equation (18.1).

In Chapter 3, I ignored the real exchange rate and subtracted IM, not IM/ε. But I was cheating; I did not want to have to talk about the real exchange rate—and complicate matters—so early in the book.

Domestic demand for goods $C + I + G$
$-$ Domestic demand for foreign goods (imports), IM/ε
$+$ Foreign demand for domestic goods (exports), X
$=$ Demand for domestic goods $C + I + G - IM/\varepsilon + X$

The Determinants of *C*, *I*, and *G*

Having listed the five components of demand, the next task is to specify their determinants. Let's start with the first three: C, I, and G. Now that we are assuming the economy is open, how should we modify our earlier descriptions of consumption, investment, and government spending? The answer: not very much, if at all. How much consumers decide to spend still depends on their income and their wealth. Although the real exchange rate surely affects the *composition* of consumption spending between domestic and foreign goods, there is no obvious reason why it should affect the overall *level* of consumption. The same is true of investment; the real exchange rate may affect whether firms buy domestic machines or foreign machines, but it should not affect total investment.

This is good news because it implies that we can use the descriptions of consumption, investment, and government spending that we developed earlier. Therefore, I assume that domestic demand is given by:

$$\text{Domestic demand:}\quad C + I + G = \underset{(+)}{C(Y - T)} + \underset{(+,-)}{I(Y, r)} + G$$

Consumption depends positively on disposable income, $Y - T$, and investment depends positively on production, Y, and negatively on the real policy rate, r. Note that I leave aside some of the refinements I introduced earlier (i.e., the presence of a risk premium, which we focused on in Chapters 6 and 14, and the role of expectations, which we focused on in Chapters 14 to 16). I want to take things one step at a time to understand the effects of opening the economy; I shall reintroduce some of those refinements later.

I again cheat a bit here. Income should include not only domestic income but also net income and transfers from abroad. For simplicity, I ignore these two additional terms here.

The Determinants of Imports

Imports are the part of domestic demand that falls on foreign goods. What do they depend on? They clearly depend on domestic income. Higher domestic income leads to a higher domestic demand for all goods, both domestic and foreign. So a higher domestic income leads to higher imports. They also clearly depend on the real exchange rate—the price of domestic goods in terms of foreign goods. The more expensive domestic goods are relative to foreign goods—or, equivalently, the cheaper foreign goods are relative to domestic goods—the higher is the domestic demand for foreign goods. So a higher real exchange rate leads to higher imports. Thus, we write imports as:

Recall the discussion at the start of this chapter. Countries in the rest of the world worry about a US recession because it means a decrease in the US demand for foreign goods.

$$IM = IM(Y, \varepsilon) \qquad (18.2)$$
$$(+, +)$$

- An increase in domestic income, Y (equivalently, an increase in domestic output—income and output are still equal in an open economy), leads to an increase in imports. This positive effect of income on imports is captured by the positive sign under Y in equation (18.2).
- An increase in the real exchange rate, ε (a real appreciation), leads to an increase in imports, IM. This positive effect of the real exchange rate on imports is captured by the positive sign under ε in equation (18.2). (As ε goes up, IM goes up but $1/\varepsilon$ goes down, so what happens to IM/ε, the *value* of imports in terms of domestic goods, is ambiguous. I return to this point below.)

The Determinants of Exports

Exports are the part of foreign demand that falls on domestic goods. What do they depend on? They depend on foreign income: Higher foreign income means higher foreign demand for all goods, foreign and domestic, and so leads to higher US exports. Exports also depend on the real exchange rate. The higher the price of domestic goods in terms of foreign goods, the lower the foreign demand for domestic goods. In other words, the higher the real exchange rate, the lower are exports.

Recall that asterisks refer to foreign variables.

Let Y^* denote foreign income (equivalently, foreign output). We therefore write exports as

$$X = X(Y^*, \varepsilon) \qquad (18.3)$$
$$(+, -)$$

- An increase in foreign income, Y^*, leads to an increase in exports.
- An increase in the real exchange rate, ε, leads to a decrease in exports.

Putting the Components Together

Figure 18-1 puts together what we have learned so far. It plots the various components of demand against output, keeping constant all other variables (the interest rate, taxes, government spending, foreign output, and the real exchange rate) that affect demand.

In Figure 18-1(a), the line DD plots domestic demand, $C + I + G$, as a function of output, Y. This relation between demand and output is familiar from Chapter 3. Under our standard assumptions, the slope of the relation between demand and output is positive but less than one. An increase in output—equivalently, an increase in

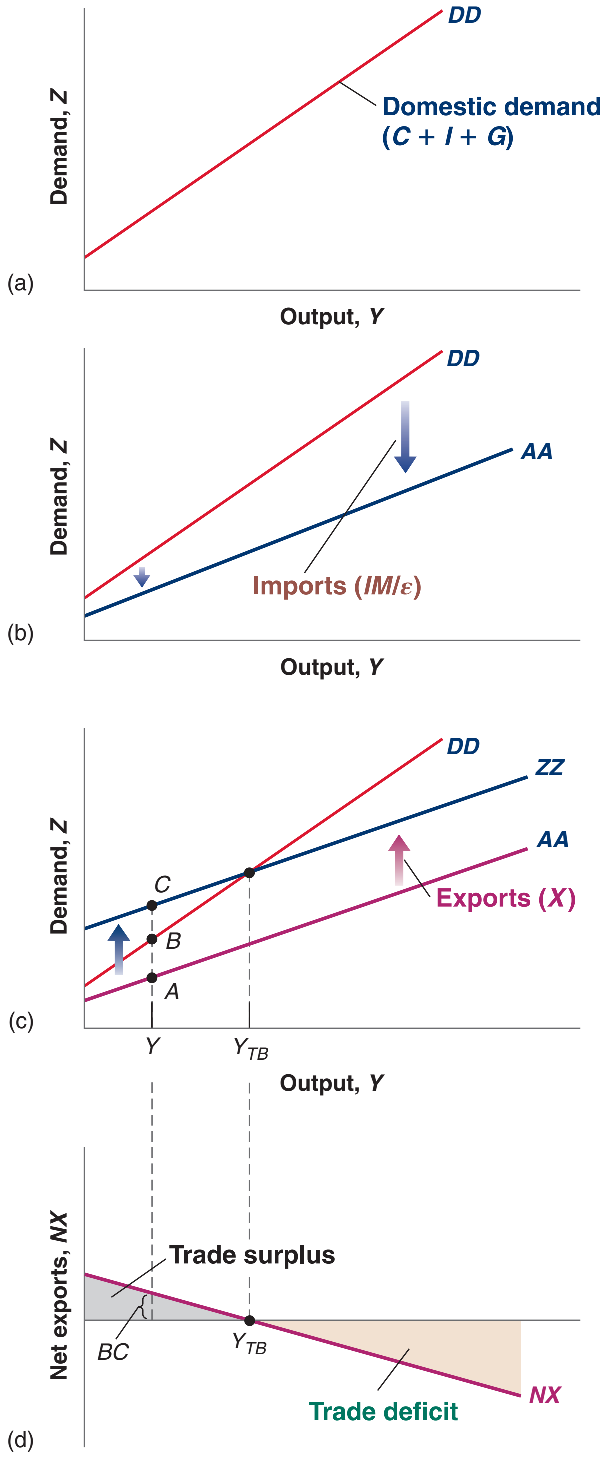
Figure 18-1

The Demand for Domestic Goods and Net Exports

(a) The domestic demand for goods is an increasing function of income (output).

(b) and *(c)* The demand for domestic goods is obtained by subtracting the value of imports from domestic demand and then adding exports.

(d) The trade balance is a decreasing function of output.



income—increases demand but less than one-for-one. (In the absence of good reasons to the contrary, we draw the relation between demand and output, and the other relations in this chapter, as lines rather than curves. This is purely for convenience, and none of the discussions that follow depend on this assumption.)

To arrive at the demand for domestic goods, we must first *subtract imports*. This is done in Figure 18-1(b) and it gives us the line *AA*, which represents the domestic demand for domestic goods. The distance between *DD* and *AA* equals the value of imports, IM/ε. Because the quantity of imports increases with income, the distance between the two lines increases with income. We can establish two facts about line AA, which will be useful later in the chapter:

For a given real exchange rate ε, IM/ε—the value of imports in terms of domestic goods—moves exactly with IM, the quantity of imports.

- AA is flatter than DD. As income increases, some of the additional domestic demand falls on foreign goods rather than on domestic goods. In other words, as income increases, the domestic demand for domestic goods increases less than total domestic demand.
- As long as some of the additional demand falls on domestic goods, AA has a positive slope: An increase in income leads to some increase in the demand for domestic goods.

Finally, we must *add exports.* This is done in Figure 18-1(c) and gives us the line ZZ, which is above AA. The line ZZ represents the demand for domestic goods. The distance between ZZ and AA equals exports, X. Because exports do not depend on domestic income (they depend on foreign income), the distance between ZZ and AA is constant, which is why the two lines are parallel. Because AA is flatter than DD, ZZ is also flatter than DD.

From the information in Figure 18-1(c) we can also characterize the behavior of net exports—the difference between exports and imports $(X - IM/\varepsilon)$—as a function of output. At output level Y, for example, exports are given by the distance AC and imports by the distance AB, so net exports are given by the distance BC.

Recall that *net exports* is synonymous with trade balance. Positive net exports correspond to a trade surplus; negative net exports correspond to a trade deficit.

This relation between net exports and output is represented as the line NX (for Net eXports) in Figure 18-1(d). Net exports are a decreasing function of output: As output increases, imports increase and exports are unaffected, so net exports decrease. Call Y_{TB} (TB for trade balance) the level of output at which the value of imports equals the value of exports, so that net exports are equal to zero. Levels of output above Y_{TB} lead to higher imports and to a trade deficit. Levels of output below Y_{TB} lead to lower imports and to a trade surplus.

18-2 EQUILIBRIUM OUTPUT AND THE TRADE BALANCE

The goods market is in equilibrium when domestic output equals the demand—both domestic and foreign—for domestic goods.

$$Y = Z$$

Collecting the relations we derived for the components of the demand for domestic goods, Z, we get

$$Y = C(Y - T) + I(Y, r) + G - IM(Y, \varepsilon)/\varepsilon + X(Y^*, \varepsilon) \qquad (18.4)$$

This equilibrium condition determines output as a function of all the variables we take as given, from taxes to the real exchange rate to foreign output. This is not a simple relation; Figure 18-2 represents it graphically in a more user-friendly way.

In Figure 18-2(a), demand is measured on the vertical axis, output (equivalently production or income) on the horizontal axis. The line ZZ plots demand as a function of output; this line just replicates the line ZZ in Figure 18-1(c). ZZ is upward sloping, but with a slope less than 1.

Equilibrium output is at the point where demand equals output, at the intersection of the line ZZ and the 45-degree line: point A in Figure 18-2(a), with associated output level Y.

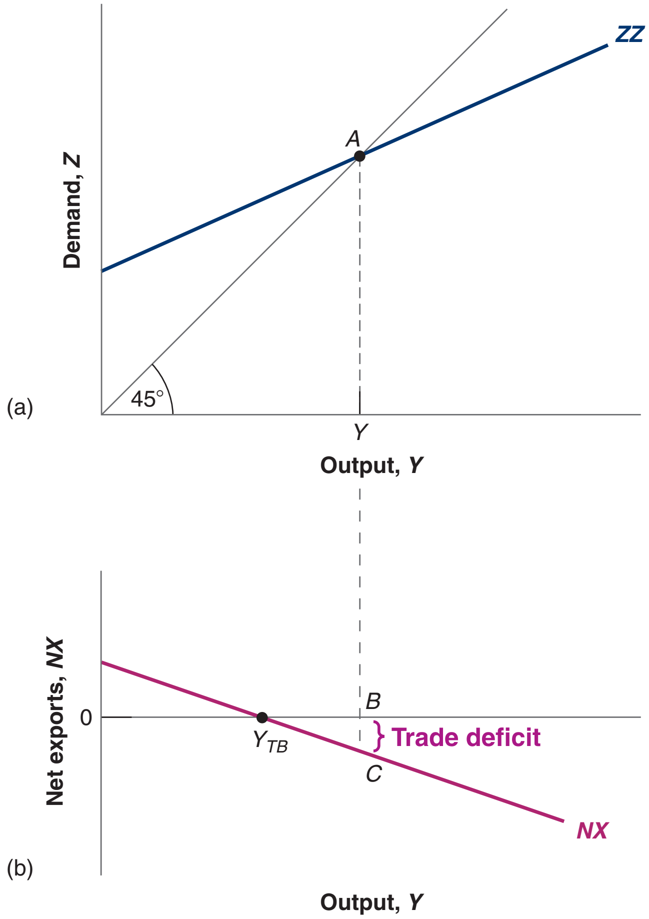
The equilibrium level of output is given by the condition $Y = Z$. The level of output at which there is trade balance is given by the condition $X = IM/\varepsilon$. These are two different conditions.

Figure 18-2(b) replicates Figure 18-1(d), drawing net exports as a decreasing function of output. There is in general no reason why the equilibrium level of output, Y, should be the same as the level of output at which trade is balanced, Y_{TB}. As shown in the figure, equilibrium output is associated with a trade deficit, equal to the

Figure 18-2

Equilibrium Output and Net Exports

The goods market is in equilibrium when domestic output is equal to the demand for domestic goods. At the equilibrium level of output, the trade balance may show a deficit or a surplus.



distance BC. The figure could be drawn differently, so that equilibrium output is associated instead with a trade surplus.

We now have the tools needed to answer the questions we asked at the beginning of this chapter.

18-3 INCREASES IN DEMAND—DOMESTIC OR FOREIGN

How do changes in demand affect output in an open economy? Let's start with what is by now an old favorite, an increase in government spending, then turn to a new exercise, on the effects of an increase in foreign demand.

Increases in Domestic Demand

As I did in the core, I start with just the goods market and introduce financial markets and labor markets later.

Suppose the economy is in a recession and the government decides to increase government spending in order to increase domestic demand and, in turn, output. What will be the effects on output and on the trade balance?

The answer is given in Figure 18-3. Before the increase in government spending, demand is given by ZZ in Figure 18-3(a) and the equilibrium is at point A, where output equals Y. Let's assume that trade is initially balanced—even though, as we have seen, there is no reason why this should be true in general. So, in Figure 18-3(b), $Y = Y_{TB}$.

What happens if the government increases spending by ΔG? At any level of output, demand is higher by ΔG, shifting the demand relation up by ΔG from ZZ to ZZ'. The

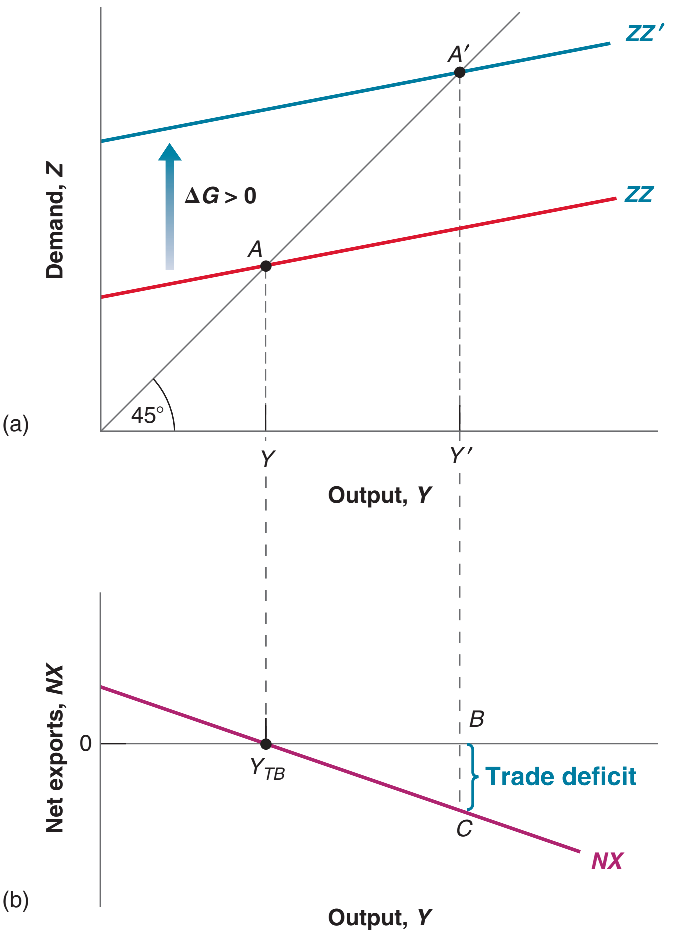


Figure 18-3

The Effects of an Increase in Government Spending

An increase in government spending leads to an increase in output and a trade deficit.

equilibrium point moves from A to A' and output increases from Y to Y'. The increase in output is larger than the increase in government spending: There is a multiplier effect.

So far, the story sounds the same as that for a closed economy in Chapter 3. However, there are two important differences:

- There is now an effect on the trade balance. Because government spending enters neither the exports relation nor the imports relation directly, the relation between net exports and output in Figure 18-3(b) does not shift. So the increase in output from Y to Y' leads to a *trade deficit* equal to BC: Imports go up, and exports do not change.
- Not only does government spending now generate a trade deficit, but the effect of government spending on output is smaller than it would be in a closed economy. Recall from Chapter 3 that the smaller the slope of the demand relation, the smaller the multiplier (for example, if ZZ were horizontal, the multiplier would be 1). And recall from Figure 18-1 that the demand relation, ZZ, is flatter than the demand relation in the closed economy, DD. This means the *multiplier is smaller in the open economy.*

Starting from trade balance, an increase in government spending leads to a trade deficit.

An increase in government spending increases output. The multiplier is smaller than in a closed economy.

The trade deficit and the smaller multiplier have the same origin. Because the economy is open, an increase in demand now falls not only on domestic goods but also on foreign goods. So when income increases, the effect on the demand for domestic goods is smaller than it would be in a closed economy, leading to a smaller multiplier. And because some of the increase in demand falls on imports—and exports are unchanged—the result is a trade deficit.

The smaller multiplier and the trade deficit have the same origin. Some domestic demand falls on foreign goods.

These two implications are important. In an open economy, an increase in domestic demand has a smaller effect on output than in a closed economy, and an adverse effect on the trade balance. Indeed, the more open the economy, the smaller the effect on output and the larger the adverse effect on the trade balance. Take the Netherlands, for example. As we saw in Chapter 17, the Netherlands' ratio of exports to GDP is very high. It is also true that the Netherlands' ratio of imports to GDP is very high. When domestic demand increases in the Netherlands, much of the increase in demand is likely to result in an increase in the demand for foreign goods rather than domestic goods. The effect of an increase in government spending is therefore likely to be a large increase in the Netherlands' trade deficit and only a small increase in its output, making domestic demand expansion a rather unattractive policy for the Netherlands. Even for the United States, which has a much lower import ratio, an increase in demand will be associated with a worsening of the trade balance.

Increases in Foreign Demand

Consider now an increase in foreign output, that is, an increase in Y^*. This could be due to an increase in foreign government spending, G^*—the policy change we just analyzed, but now taking place abroad. But we do not need to know where the increase in Y^* comes from to analyze its effects on the US economy.

Figure 18-4 shows the effects of an increase in foreign activity on domestic output and the trade balance. The initial demand for domestic goods is given by ZZ in

Figure 18-4

The Effects of an Increase in Foreign Demand

An increase in foreign demand leads to an increase in output and a trade surplus.

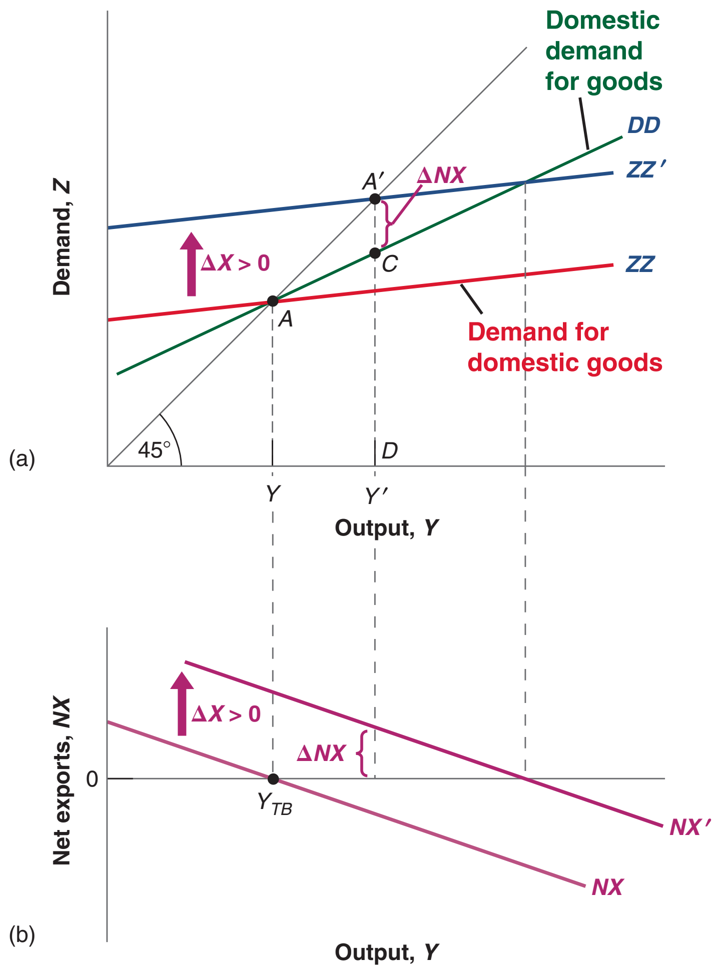


Figure 18-4(a). The equilibrium is at point A, with output level Y. Let's again assume trade is balanced, so that in Figure 18-4(b) the net exports associated with Y equal zero $(Y = Y_{TB})$.

It will be useful below to refer to the line that shows the *domestic demand for goods* $C + I + G$ as a function of income. This line is drawn as *DD*. Recall from Figure 18-1 that *DD* is steeper than *ZZ*. The difference between *ZZ* and *DD* equals net exports, so that if trade is balanced at point A, then ZZ and *DD* intersect at point A.

DD is the domestic demand for goods. *ZZ* is the demand for domestic goods. The difference between the two is equal to the trade deficit.

Now consider the effects of an increase in foreign output, ΔY^* (for the moment, ignore the line *DD*; we need it only later). Higher foreign output means higher foreign demand, including higher foreign demand for US goods. So the direct effect of the increase in foreign output is an increase in US exports by some amount, which we shall denote by ΔX.

- For a given level of output, this increase in exports leads to an increase in the demand for US goods by ΔX, so the line showing the demand for domestic goods as a function of output shifts up by ΔX, from *ZZ* to *ZZ'*.
- For a given level of output, net exports go up by ΔX. So the line showing net exports as a function of output in Figure 18-4(b) also shifts up by ΔX, from *NX* to *NX'*.

The new equilibrium is at point A' in Figure 18-4(a), with output level Y'. The increase in foreign output leads to an increase in domestic output: Higher foreign output leads to higher exports of domestic goods, which increases domestic output and the domestic demand for goods through the multiplier.

What happens to the trade balance? We know that exports go up. But could it be that the increase in domestic output leads to such a large increase in imports that the trade balance actually deteriorates? No: The trade balance must improve. To see why, note that, when foreign demand increases, the demand for domestic goods shifts up from *ZZ* to *ZZ'*; but the line *DD*, which gives the *domestic demand for goods* as a function of output, does not shift. At the new equilibrium level of output Y', domestic demand is given by the distance *DC*, and the demand for domestic goods is given by *DA'*. Net exports are therefore given by the distance *CA'*—which, because *DD* is necessarily below *ZZ'*, is necessarily positive. Thus, the increase in imports does not offset the increase in exports, and the trade balance improves.

*Y** directly affects exports and so enters the relation between the demand for domestic goods and output. An increase in *Y** shifts *ZZ* up. *Y** does not affect domestic consumption, domestic investment, or domestic government spending directly, and so it does not enter the relation between the domestic demand for goods and output. An increase in *Y** does not shift *DD*.

An increase in foreign output increases domestic output and improves the trade balance.

Fiscal Policy Revisited

We have derived two results so far:

- An increase in domestic demand leads to an increase in domestic output but also to a deterioration of the trade balance. (We looked at an increase in government spending, but the results would have been the same for a decrease in taxes, an increase in consumer spending, and so on.)
- An increase in foreign demand (which could come from the same types of changes taking place abroad) leads to an increase in domestic output and an improvement in the trade balance.

These results, in turn, have two important implications. Both were in evidence during the Great Financial Crisis.

First, and most obviously, they imply that shocks to demand in one country affect other countries. The stronger the trade links between countries, the stronger the interactions, and the more countries will move together. This is what we saw in Figure 17–1. Although the crisis started in the United States, it quickly affected the rest of the world. Trade links were not the only reason; financial links also played a central role. But the evidence points to a strong effect of trade, starting with a decrease in exports from other countries to the United States.

FOCUS

The G20 and the 2009 Fiscal Stimulus

In November 2008, the leaders of the **G20** countries gathered for an emergency meeting in Washington. The G20, a group of ministers of finance and central bank governors from 20 countries, including both the major advanced and the major emerging countries in the world, was created in 1999 but had not played a major role until the crisis. With mounting evidence that the crisis was going to be both deep and widespread, the group met to coordinate their responses in terms of both macroeconomic and financial policies.

On the macroeconomic front, it had become clear that monetary policy would not be enough, and so the focus turned to fiscal policy. The decrease in output was going to lead to a decrease in revenues, and thus an increase in budget deficits. Dominique Strauss-Kahn, then the managing director of the International Monetary Fund, argued that further fiscal actions were needed and suggested taking additional discretionary measures—either decreases in taxes or increases in spending—adding up to roughly 2% of GDP on average for each country. This is what he said:

"The fiscal stimulus is now essential to restore global growth. Each country's fiscal stimulus can be twice as effective in raising domestic output growth if its major trading partners also have a stimulus package."

He noted that some countries had more room for maneuver than others: "We believe that those countries—advanced and emerging economies—with the strongest fiscal policy frameworks, the best ability to finance fiscal expansion, and the most clearly sustainable debt should take the lead."

Over the next few months, most countries did indeed adopt measures aimed at increasing either private or public spending. For the G20 as a whole, discretionary measures added up to about 2.3% of GDP in 2009. Some countries with less fiscal room, such as Italy, did less; some countries, such as the United States or France, did more.

Was this fiscal stimulus successful? Some have argued that it was not. After all, the world economy had large negative growth in 2009. The issue here is one of counterfactuals. What would have happened in the absence of the stimulus? Many believe that, absent the fiscal stimulus, growth would have been even more negative, perhaps catastrophically so. Counterfactuals are hard to prove or disprove, and thus the controversy is likely to go on. (On the issue of counterfactuals and the difference between economists and politicians, there is a nice quote from former US congressman Barney Frank:

> "Not for the first time, as an elected official, I envy economists. Economists have available to them, in an analytical approach, the counterfactual. Economists can explain that a given decision was the best one that could be made, because they can show what would have happened in the counterfactual situation. They can contrast what happened to what would have happened. No one has ever gotten reelected where the bumper sticker said, 'It would have been worse without me.' You probably can get tenure with that. But you can't win office.")

Was this fiscal stimulus dangerous? Some have argued that it led to a large increase in public debt, which forced governments to adjust, leading to a fiscal contraction and making recovery more difficult (we discussed this in Chapter 5 and will return to it in Chapter 22). This argument is largely misplaced. Most of the increase in debt came not from the discretionary measures that were taken, but from the decrease in revenues due to the decrease in output during the crisis. Debt would have increased even in the absence of the discretionary measures. It remains true, however, that the large increase in debt led governments to adopt a contractionary fiscal policy and that it probably slowed down the recovery.

For more of the discussion at the time, see "Financial Crisis Response: IMF Spells Out Need for Global Fiscal Stimulus," in IMF Survey Magazine Online, December 29, 2008 (www.imf.org/external/pubs/ft/survey/so/2008/int122908a.htm).

Second, these interactions complicate the task of policymakers, especially in the case of fiscal policy. Let's explore this argument more closely.

Start with the following observation: Governments do not like to run trade deficits and for good reasons. The main reason: A country that consistently runs a trade deficit accumulates debt vis-à-vis the rest of the world, and therefore has to pay steadily higher interest payments to the rest of the world. Thus, it is no wonder that countries prefer increases in foreign demand (which improve the trade balance) to increases in domestic demand (which worsen the trade balance).

But these preferences can have disastrous implications. Consider a group of countries, all doing a large amount of trade with each other, so that an increase in demand in any one country falls largely on the goods produced in the other countries. Suppose all these countries are in recession and each has roughly balanced trade to start. In this case, each country might be reluctant to take measures to increase domestic demand. Were it to do so, the result might be a small increase in output but also a large trade deficit. Instead, each country might just wait for the other countries to increase their demand. This way, it gets the best of both worlds—higher output and an improvement in its trade balance. But if all the countries wait, nothing will happen, and the recession may last a long time.

Is there a way out? There is—at least in theory. If all countries coordinate their macroeconomic policies so as to increase domestic demand simultaneously, each can increase demand and output without increasing its trade deficit (vis-à-vis the others; their combined trade deficit with respect to the rest of the world will still increase). The reason is clear. The coordinated increase in demand leads to increases in both exports and imports in each country. It is still true that domestic demand expansion leads to larger imports; but this increase in imports is offset by the increase in exports, which comes from the foreign demand expansions.

In practice, however, **policy coordination** is not so easy to achieve.

Some countries might have to do more than others and may not want to do so. Suppose that only some countries are in recession. Countries that are not in a recession will be reluctant to increase their own demand; but if they do not, the countries that expand will run a trade deficit vis-à-vis countries that do not. Or suppose some countries are already running a large budget deficit. These countries might not want to cut taxes or further increase spending as this would further increase their deficits. They will ask other countries to take on more of the adjustment. Those other countries may be reluctant to do so.

Countries also have a strong incentive to promise to coordinate and then not deliver on their promise. Once all countries have agreed to, say, an increase in spending, each country has an incentive not to deliver, so as to benefit from the increase in demand elsewhere and thereby improve its trade position. But if each country cheats, or does not do everything it promised, there will be insufficient demand expansion to get out of the recession.

The result is that, despite declarations by governments at international meetings, coordination often fizzles. Only when things are really bad does coordination appear to take hold. This was the case in 2009 and is explored in the Focus Box "The G20 and the 2009 Fiscal Stimulus."

18-4 DEPRECIATION, THE TRADE BALANCE, AND OUTPUT

Suppose the US government takes policy measures that lead to a depreciation of the dollar—a decrease in the nominal exchange rate. (We shall see in Chapter 20 how this can be done by using monetary policy. For the moment I shall assume the government can simply set the exchange rate.)

Recall that the real exchange rate is given by

$$\varepsilon = \frac{E\,P}{P^*}$$

The real exchange rate, ε (the price of domestic goods in terms of foreign goods), is equal to the nominal exchange rate, E (the price of domestic currency in terms of foreign currency), times the domestic price level, P, divided by the foreign price level, P^*. In the short run, we can take the two price levels P and P^* as given. This implies that the nominal depreciation is reflected one-for-one in a real depreciation. More concretely, if the dollar depreciates vis-à-vis the yen by 10% (a 10% nominal depreciation) and if price levels in Japan and the United States do not change, US goods will be 10% cheaper compared to Japanese goods (a 10% real depreciation).

Let's now ask how this real depreciation will affect the US trade balance and US output.

Given P and P^*, E increases $\Rightarrow \varepsilon = \dfrac{EP}{P^*}$ increases.

A look ahead: In Chapter 20, we shall look at the effects of a nominal depreciation when we allow the price level to adjust over time. You will see that a nominal depreciation leads to a real depreciation in the short run but not in the medium run.

Depreciation and the Trade Balance: the Marshall-Lerner Condition

Return to the definition of net exports:

$$NX = X - IM/\varepsilon$$

More concretely, if the dollar depreciates vis-à-vis the yen by 10%:

US goods will be cheaper in Japan, leading to a larger quantity of US exports to Japan.

Japanese goods will be more expensive in the United States, leading to a smaller quantity of imports of Japanese goods to the United States and a higher import bill for a given quantity of imports of Japanese goods to the United States.

Replace X and IM by their expressions from equations (18.2) and (18.3):

$$NX = X(Y^*, \varepsilon) - IM(Y, \varepsilon)/\varepsilon$$

As the real exchange rate ε enters the right side of the equation in three places, this makes it clear that the real depreciation affects the trade balance through three separate channels:

- *Exports, X, increase.* The real depreciation makes US goods relatively less expensive abroad. This leads to an increase in foreign demand for US goods—an increase in US exports.
- *Imports, IM, decrease*. The real depreciation makes foreign goods relatively more expensive in the United States. This leads to a shift in domestic demand toward domestic goods and to a decrease in the quantity of imports.
- *The relative price of foreign goods in terms of domestic goods,* $1/\varepsilon$*, increases.* For given IM, this increases the import bill, IM/ε. The same quantity of imports now costs more to buy (in terms of domestic goods).

It is named after the two economists, Alfred Marshall and Abba Lerner, who first derived it.

For the trade balance to improve following a depreciation, exports must increase enough and imports must decrease enough to compensate for the increase in the price of imports. The condition under which a real depreciation leads to an increase in net exports is known as the **Marshall-Lerner condition**. (It is derived formally in the appendix, called "Derivation of the Marshall Lerner-Condition," at the end of this chapter.) It turns out—with a complication we will state when we introduce dynamics later in this chapter—that this condition is satisfied in reality. So, for the rest of this book, I shall assume that a real depreciation—a decrease in ε—leads to an increase in net exports—an increase in NX.

The Effects of a Real Depreciation

We have looked so far at the *direct* effects of a depreciation on the trade balance—that is, the effects *given US and foreign output*. But the effects do not end there. The change in net exports changes domestic output, which further affects net exports.

Because the effects of a real depreciation are much like those of an increase in foreign output, we can use Figure 18-4 to show the effects of an increase in foreign output.

Just like an increase in foreign output, a depreciation leads to an increase in net exports (assuming, as we do, that the Marshall-Lerner condition holds) at any level of output. Both the demand relation (ZZ in Figure 18-4(a)) and the net exports relation (NX in Figure 18-4(b)) shift up. The equilibrium moves from A to A', and output increases from Y to Y'. By the same argument we used previously, the trade balance improves. The increase in imports induced by the increase in output is smaller than the direct improvement in the trade balance induced by the depreciation.

To summarize: The depreciation leads to a shift in demand, both foreign and domestic, toward domestic goods. This shift in demand leads, in turn, to both an increase in domestic output and an improvement in the trade balance.

Although a depreciation and an increase in foreign output have the same effect on domestic output and the trade balance, there is a subtle but important difference between the two. A depreciation works by making foreign goods relatively more expensive. But this means that, for a given income, people—who now pay more to buy foreign goods because of the depreciation—are worse off. This mechanism is strongly felt in countries that go through a large depreciation. Governments trying to achieve a large depreciation often find themselves with strikes and riots in the streets as people react to the much higher prices of imported goods. This was the case in Mexico, for example, where the large depreciation of the peso in 1994–1995—from 29 cents per peso in November 1994 to 17 cents per peso in May 1995—led to a large decline in workers' living standards and to social unrest.

There is an alternative to rioting asking for and obtaining an increase in wages. But if wages increase, the prices of domestic goods will increase as well, leading to a smaller real depreciation. To discuss this mechanism, we need to look at the supply side in more detail than we have done so far. We return to the dynamics of depreciation, wage and price movements in Chapter 20.

Combining Exchange Rate and Fiscal Policies

Suppose output is at its natural level but the economy is running a large trade deficit. The government would like to reduce the trade deficit while leaving output unchanged so as to avoid overheating. What should it do?

A depreciation alone will not do. It will reduce the trade deficit, but it will also increase output. Nor will a fiscal contraction do. It will reduce the trade deficit, but it will decrease output. What should the government do? The answer: Use the right combination of depreciation and fiscal contraction. Figure 18-5 shows what this combination should be.

Suppose the initial equilibrium in Figure 18-5(a) is at *A*, associated with output *Y*. At this level of output, there is a trade deficit, given by the distance *BC* in Figure 18-5(b).

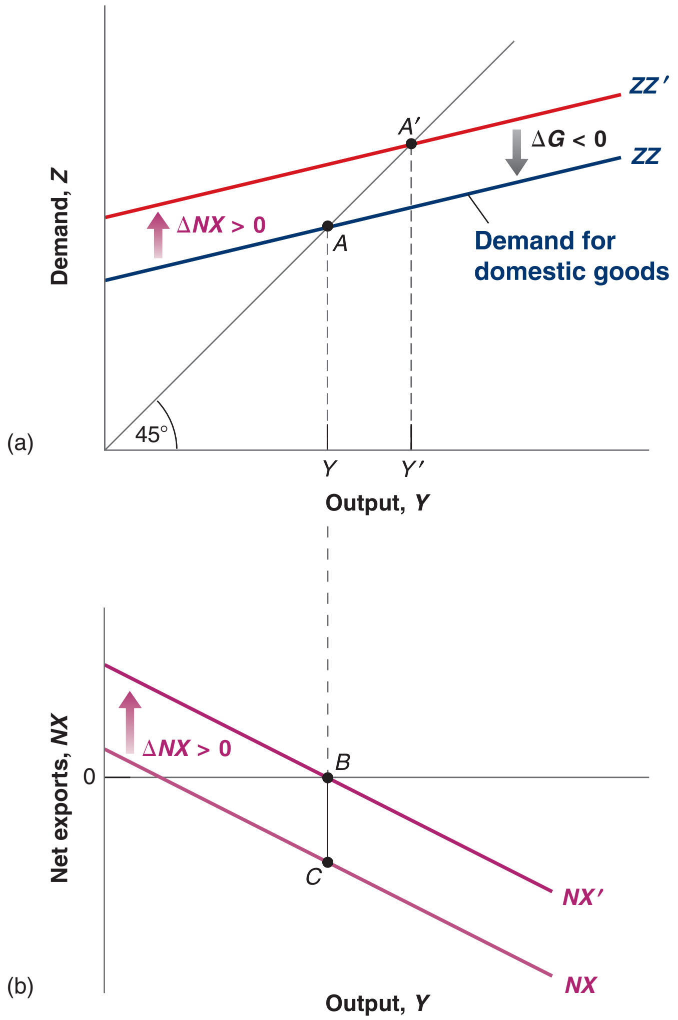


Figure 18-5

Reducing the Trade Deficit without Changing Output

To reduce the trade deficit without changing output, the government must both achieve a depreciation and decrease government spending.

Table 18-1	Exchange Rate and Fiscal Policy Combinations	
Initial Conditions	**Trade Surplus**	**Trade Deficit**
Low output	ε? G↑	ε↓ G?
High output	ε↑ G?	ε? G↓

If the government wants to eliminate the trade deficit without changing output, it must do two things:

- It must achieve a depreciation sufficient to eliminate the trade deficit at the initial level of output. So the depreciation must be such as to shift the net exports relation from NX to NX' in Figure 18-5(b). The problem is that this depreciation, and the associated increase in net exports, also shifts the demand relation in Figure 18-5(a) from ZZ to ZZ'. In the absence of other measures, the equilibrium would move from A to A' and output would increase from Y to Y'.
- To avoid the increase in output, it must therefore reduce government spending so as to shift ZZ' back to ZZ. This combination of a depreciation and a fiscal contraction leads to the same level of output and an improved trade balance.

If this combination seems quite different from the policy adopted by the Trump administration to reduce the US trade deficit, you are right. More on this in Chapter 19.

A general lesson: If you want to achieve two targets (here, output and trade balance), you better use two instruments (here, fiscal policy and the exchange rate).

There is a general point behind this example. To the extent that governments care about *both* the level of output and the trade balance, they have to use *both* fiscal policy and exchange rate policies. We just saw one such combination. Table 18-1 gives you others, depending on the initial output and trade situation. Take, for example, the box in the top right corner of the table: Initial output is too low (put another way, unemployment is too high) and the economy has a trade deficit. A depreciation will help on both the trade and the output fronts: It reduces the trade deficit and increases output. But there is no reason for the depreciation to achieve both the correct increase in output and the elimination of the trade deficit. Depending on the initial situation and the relative effects of the depreciation on output and the trade balance, the government may need to complement the depreciation with either an increase or a decrease in government spending. This ambiguity is captured by the question mark. Make sure that you understand the logic behind each of the other three boxes. (For another example of the role of the real exchange rate and output in affecting the current account balance, look at the Focus Box "The Disappearance of the Current Account Deficit in Greece: Good News or Bad News?")

FOCUS

The Disappearance of the Current Account Deficit in Greece: Good News or Bad News?

Starting in the early 2000s, several euro periphery countries ran larger and larger current account deficits. Particularly striking was the increase in the current account deficit of Greece. As shown in Figure 1, the current account went from an already large deficit of 6% of GDP in 2000 to more than 15% in 2008. When the Great Financial crisis started, Greece found it increasingly difficult to borrow abroad, forcing it to reduce borrowing and thus to reduce its current account deficit. But by 2018, the current account deficit was less than 1% of GDP.

It is an impressive turnaround. Is it unambiguously good news? Not necessarily. The discussion in the text suggests that there are two ways a current account may improve. The first is that the country becomes more competitive: The real exchange rate decreases, exports increase, imports decrease, and the current account balance improves. The second is that the country's output decreases: Exports, which depend on what happens in the rest of the world, may remain the same, but imports come down with output, and the current account balance improves.

Unfortunately, the evidence is that the second mechanism played a central role in the adjustment.

Given that Greece is a member of the euro area, it could not rely on an adjustment of the nominal exchange rate to become more competitive, at least vis-à-vis its euro

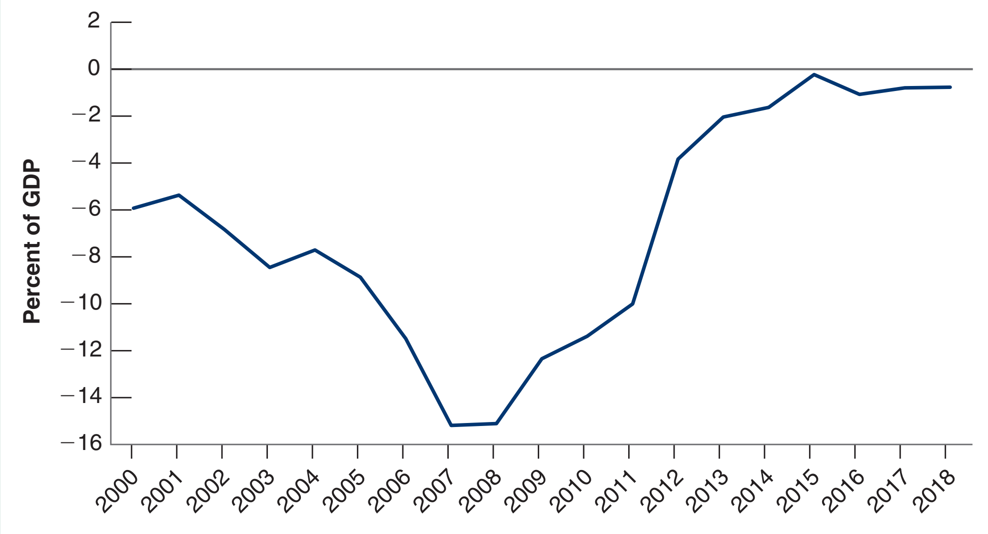


Figure 1

The Evolution of the Greek Current Account Deficit since 2000

Source: IMF Managing Director Dominique Strauss-Kahn Calls G-20 Action Plan Significant Step toward Stronger International Cooperation, Press Release No. 08/286, November 15, 2008.

partners. It had to rely on a decrease in wages and prices, and this has proven to be slow and difficult (more on this in Chapter 20).

Instead, much of the adjustment has taken place through a decrease in imports, triggered by a decrease in output, an adjustment known as **import compression**. Figure 2 shows the evolution of imports, exports, and GDP in Greece since 2000. All three series are normalized to equal 1.0 in 2000. Note first how much output has decreased from its peak, by roughly 30% since 2008. Note then how imports have moved in tandem with output, also decreasing by 36% since 2008. Exports have done better but not great. After sharply decreasing in 2009, reflecting the world crisis and the decrease in demand from the rest of the world, they are only 10% above their 2008 level (world output growth since 2008 has been a cumulative 34%).

In short, the disappearance of the current account deficit in Greece is, on net, largely bad news. Turning to the future, what happens to the current account next depends largely on what happens to output, and this in turn depends on where output is relative to potential output. If much of the decrease in actual output reflects a decrease in potential output, then output will remain low and the current account will remain in balance. If, as seems more likely, actual output is still far below potential output (if there is, in the terminology of Chapter 9, a large negative output gap), then unless further real depreciation takes place, the return of output to potential will come with higher imports, and thus a likely return to current account deficits.

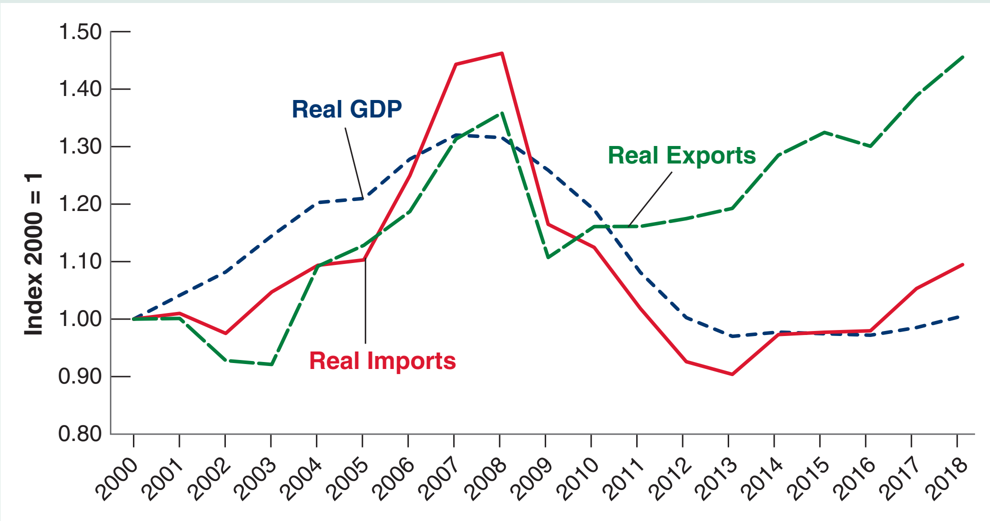


Figure 2

Imports, Exports, and GDP in Greece since 2000

Source: IMF Managing Director Dominique Strauss-Kahn Calls G-20 Action Plan Significant Step toward Stronger International Cooperation, Press Release No. 08/286, November 15, 2008.

18-5 SAVING, INVESTMENT, AND THE CURRENT ACCOUNT BALANCE

Getting there involves some manipulations, but do not worry—the end result is intuitive.

You saw in Chapter 3 that we could rewrite the condition for equilibrium in the goods market as the condition that investment was equal to saving—the sum of private and public saving. We can now derive the corresponding condition for the open economy, and you will see how useful this alternative way of looking at the equilibrium can be.

Start from our equilibrium condition

$$Y = C + I + G - IM/\varepsilon + X$$

Move consumption, C, from the right side to the left side of the equation, subtract taxes, T, from both sides, and denote net exports $(IM/\varepsilon + X)$ by NX to get

$$Y - T - C = I + (G - T) + NX$$

Recall that, in an open economy, the total income of domestic residents is equal to domestic income, which is equal to domestic output, Y, plus net income from abroad, NI. Add NI to both sides of the equation:

$$(Y + NI - T) - C = I + (G - T) + (NX + NI)$$

Note that the term in parentheses on the left side is equal to disposable income, so the left side is equal to disposable income minus consumption (i.e., private saving, S). The sum of net exports and net income from abroad on the right side is equal to the current account. Denote the current account by CA and rewrite the previous equation as:

$$S = I + (G - T) + CA$$

Reorganize the equation to read:

$$CA = S + (T - G) - I \qquad (18.5)$$

Commentators often do not make a distinction between the trade balance and the current account balance. This is not necessarily a major crime: Because net income typically moves slowly over time, the trade and current account balances typically move closely together. In the United States, however, net income is large, so the current account balance typically looks much better than the trade balance.

The current account balance is equal to saving (private plus public) minus investment. A current account surplus implies that the country is saving more than it invests. A current account deficit implies that the country is saving less than it invests.

One way of getting more intuition for this relation is to go back to the discussion of the current account and the financial account in Chapter 17. There we saw that a current account surplus implies net lending from the country to the rest of the world, and a current account deficit implies net borrowing by the country from the rest of the world.

Consider a country that invests more than it saves, so that $S + (T - G) - I$ is negative. That country must be borrowing the difference from the rest of the world, and must therefore be running a current account deficit. Symmetrically, a country that lends to the rest of the world is a country that saves more than it invests.

Note some of the things that equation (18.5) says:

- An increase in investment must be reflected in either an increase in private or public saving or a deterioration of the current account balance—a smaller current account surplus, or a larger current account deficit, depending on whether the current account is initially in surplus or in deficit.
- A deterioration in the government budget balance—either a smaller budget surplus or a larger budget deficit—must be reflected either in an increase in private saving or a decrease in investment, or in a deterioration of the current account balance.
- A country with a high saving rate (private plus public) must have either a high investment rate or a large current account surplus.

Note also, however, what equation (18.5) does *not* say.

It does not say, for example, whether a government budget deficit will lead to a current account deficit, or instead to an increase in private saving or a decrease in investment. To find out what happens in response to a budget deficit, we must explicitly solve for what happens to output and its components using the assumptions that we have made about consumption, investment, exports, and imports. That is, we need to do the complete analysis laid out in this chapter. Using only equation (18.5) can, if you are not careful, be very misleading. To see how misleading, consider, for example, the following argument (which is so common that you may have read something similar in newspapers):

> "It is clear that the United States cannot reduce its large current account deficit through a depreciation. Look at equation (18.5). It shows that the current account deficit is equal to investment minus saving. Why should a depreciation affect either saving or investment? So how can a depreciation affect the current account deficit?"

The argument might sound convincing, but we know it is wrong. We showed earlier that a depreciation leads to an improvement in a country's trade position and, by implication—given net income—an improvement in the current account. What is wrong with the argument? A depreciation actually does affect saving and investment, by affecting the demand for domestic goods and thereby increasing output. Higher output leads to an increase in saving over investment, or equivalently, to a decrease in the current account deficit.

A good way to make sure you understand the material in this section is to go back and look at the various cases we have considered—changes in government spending, changes in foreign output, combinations of depreciation and fiscal contraction, and so on. Trace what happens in each case to each of the four components of equation (18.5): private saving, public saving (equivalently, the budget surplus), investment, and the current account balance. Make sure, as always, that you can tell the story in words.

◀ Suppose, for example, that the US government wants to reduce the current account deficit without changing the level of output, so it uses a combination of depreciation and fiscal contraction. What happens to private saving, public saving, and investment?

Let me end the chapter with a challenge. Assess the following three statements and decide which one(s) is (are) right:

- The large US current account deficit (which you saw in Chapter 17) shows that the United States is no longer competitive. It is a sign of weakness. Forget saving or investment. The United States must urgently improve its competitiveness.
- The large US current account deficit shows that the United States just does not save enough to finance its investment. It is a sign of weakness. Forget competitiveness. The United States must urgently increase its saving rate.
- The large US current account deficit is just a mirror image of the US financial account surplus. What is happening is that the rest of the world wants to put its funds in the United States. The US financial account surplus, and by implication, the US current account deficit, is in fact a sign of strength, and there is no need to take policy measures to reduce it.

SUMMARY

- In an open economy, the demand for domestic goods is equal to the domestic demand for goods (consumption plus investment plus government spending) minus the value of imports (in terms of domestic goods) plus exports.
- In an open economy, an increase in domestic demand leads to a smaller increase in output than it would in a closed economy because some of the additional demand falls on imports. For the same reason, an increase in domestic demand also leads to a deterioration of the trade balance.
- An increase in foreign demand leads, as a result of increased exports, to both an increase in domestic output and an improvement of the trade balance.
- Because increases in foreign demand improve the trade balance and increases in domestic demand worsen the trade balance, countries might be tempted to wait for increases in foreign demand to move them out of a recession. When a group of countries is in recession, coordination can, in principle, help their recovery.
- If the Marshall-Lerner condition is satisfied—and the empirical evidence indicates that it is—a real depreciation leads to an improvement in net exports.
- The condition for equilibrium in the goods market can be rewritten as the condition that saving (public and private) minus investment must be equal to the current account balance. A current account surplus corresponds to an excess of saving over investment. A current account deficit corresponds to an excess of investment over saving.

KEY TERMS

demand for domestic goods, 376
domestic demand for goods, 376
G20, 384
policy coordination, 385
Marshall-Lerner condition, 386
import compression, 389

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The reduction in the current account deficit in Greece from 2008 to 2018 means that citizens in Greece are better off.
b. The national income identity implies that budget deficits cause trade deficits.
c. Opening the economy to trade tends to increase the multiplier because an increase in expenditure leads to more exports.
d. If the trade deficit is equal to zero, then the domestic demand for goods and the demand for domestic goods are equal.
e. A small open economy can reduce its trade deficit through fiscal contraction at a smaller cost in output than can a large open economy.
f. A decline in real income can lead to a decline in imports and thus a trade surplus.
g. In joining the Euro, Greece made real exchange rate adjustments with its trading partners easier.

2. Real and nominal exchange rates and inflation

Using the definition of the real exchange rate (and Propositions 7 and 8 in Appendix 2 at the end of the book), you can show that

$$\frac{(\varepsilon_t - \varepsilon_{t-1})}{\varepsilon_{t-1}} = \frac{(E_t - E_{t-1})}{E_{t-1}} + \pi_t - \pi_t^*$$

In words, the percentage real appreciation equals the percentage nominal appreciation plus the difference between domestic and foreign inflation.

a. If domestic inflation is higher than foreign inflation, and the domestic country has a fixed exchange rate, what happens to the real exchange rate over time? Assume that the Marshall-Lerner condition holds. What happens to the trade balance over time? Explain in words.
b. Suppose the real exchange rate is currently at the level required to have the trade balance equal to zero. In this case, if domestic inflation is higher than foreign inflation, what must happen over time to maintain a trade balance of zero?

3. A European recession and the US economy

a. In 2017, European Union spending on US goods accounted for 19% of US exports (see Table 17-2), and US exports amounted to 12.3% of US GDP (see Table 17-1). What was the share of European Union spending on US goods relative to US GDP?
b. Assume that the multiplier in the United States is 2 and that a major slump in Europe would reduce output and imports from the United States by 5% (relative to its normal level). Given your answer to part a, what is the impact of the European slump on US GDP?
c. If the European slump also leads to a slowdown of the other economies that import goods from the United States, the effect could be larger. To put a bound to the size of the effect, assume that US exports decrease by 5% (as a result

of changes in foreign output) in one year. What is the effect of a 5% drop in exports on US GDP?

d. Comment on this statement. "If Europe has a major slump, US growth will also grind to a halt."

4. A further look at Table 18-1

Table 18-1 has four entries. Using Figure 18-5 as a guide, draw the situations illustrated in each of the four entries in Table 18-1. Be sure you understand why the direction of change in government spending and the real exchange rate is labeled as ambiguous in each entry.

DIG DEEPER

5. Net exports and foreign demand

a. Suppose there is an increase in foreign output. Show the effect on the domestic economy (i.e., replicate Figure 18-4). What is the effect on domestic output? On domestic net exports?

b. If the interest rate remains constant, what will happen to domestic investment? If taxes are fixed, what will happen to the domestic budget deficit?

c. Using equation (18.5), what must happen to private saving? Explain.

d. Foreign output does not appear in equation (18.5), yet it evidently affects net exports. Explain how this is possible.

6. Eliminating a trade deficit

a. Consider an economy with a trade deficit ($NX < 0$) and with output equal to its natural level. Suppose that, even though output may deviate from its natural level in the short run, it returns to its natural level in the medium run. Assume that the natural level is unaffected by the real exchange rate. What must happen to the real exchange rate over the medium run to eliminate the trade deficit (i.e., to increase NX to 0)?

b. Now write down the national income identity. Assume again that output returns to its natural level in the medium run. If NX increases to 0, what must happen to domestic demand ($C + I + G$) in the medium run? What government policies are available to reduce domestic demand in the medium run? Identify which components of domestic demand each of these policies affects.

7. Multipliers, openness, and fiscal policy

Consider an open economy characterized by the following equations:

$$C = c_0 + c_1(Y - T)$$
$$I = d_0 + d_1 Y$$
$$IM = m_1 Y$$
$$X = x_1 Y^*$$

The parameters m_1 and x_1 are the propensities to import and export. Assume that the real exchange rate is fixed at a value of 1 and treat foreign income, Y^*, as fixed. Also assume that taxes are fixed and that government purchases are exogenous (i.e., decided by the government). We explore the effectiveness of changes in G under alternative assumptions about the propensity to import.

a. Write the equilibrium condition in the market for domestic goods and solve for Y.

b. Suppose government purchases increase by one unit. What is the effect on output? (Assume that $0 < m_1 < c_1 + d_1 < 1$. Explain why.)

c. How do net exports change when government purchases increase by one unit?

Now consider two economies, one with $m_1 = 0.5$ and the other with $m_1 = 0.1$. Each economy is characterized by $(c_1 + d_1) = 0.6$.

d. Suppose one of the economies is much larger than the other. Which economy do you expect to have the larger value of m_1? Explain.

e. Calculate your answers to parts b and c for each economy by substituting the appropriate parameter values.

f. In which economy will fiscal policy have a larger effect on output? In which economy will fiscal policy have a larger effect on net exports?

8. Policy coordination and the world economy

Consider an open economy in which the real exchange rate is fixed and equal to one. Consumption, investment, government spending, and taxes are given by

$$C = 10 + 0.8(Y - T),\ I = 10,\ G = 10,\ \text{and}\ T = 10$$

Imports and exports are given by

$$IM = 0.3\,Y \quad \text{and} \quad X = 0.3\,Y^*$$

where Y^* denotes foreign output.

a. Solve for equilibrium output in the domestic economy, given Y^*. What is the multiplier in this economy? If we were to close the economy—so that exports and imports were identically equal to zero—what would the multiplier be? Why would the multiplier be different in a closed economy?

b. Assume that the foreign economy is characterized by the same equations as the domestic economy (with asterisks reversed). Use the two sets of equations to solve for the equilibrium output of each country. (*Hint:* Use the equations for the foreign economy to solve for Y^* as a function of Y and substitute this solution for Y^* in part a.) What is the multiplier for each country now? Why is it different from the open economy multiplier in part a?

c. Assume that the domestic government, G, has a target level of output of 125. Assuming that the foreign government does not change G^*, what is the increase in G necessary to achieve the target output in the domestic economy? Solve for net exports and the budget deficit in each country.

d. Suppose each government has a target level of output of 125 and that each government increases government spending by the same amount. What is the common increase in G and G^* necessary to achieve the target output in both countries? Solve for net exports and the budget deficit in each country.

e. Why is fiscal coordination, such as the common increase in G and G^* in part d, difficult to achieve in practice?

EXPLORE FURTHER

9. The US trade deficit, current account deficit, and investment

a. Define national saving as private saving plus the government surplus—that is, as $S + T - G$. Now, using equation (18.5), describe the relation among the current account deficit, net investment income, and the difference between national saving and domestic investment.

b. Using the FRED economic database, retrieve annual data for nominal GDP (series GDPA), gross domestic investment (series W170RC1A027NBEA), and net exports (series A019RC1A027NBEA) from 1980 to the most recent year available. Divide gross domestic investment and net exports by GDP in each year to express their values as a percentage of GDP. Which years have the largest trade deficits as a percentage of GDP?

c. The trade surplus in 1980 was roughly zero. Compute the average ratio of investment to GDP and the average value of the trade balance as a percent of GDP in three periods: 1980–1989, 1990–1999, 2000–2009 and 2010 to the latest point. Would it appear that trade deficits have been used to finance investment?

d. Is a trade deficit more worrisome when not accompanied by a corresponding increase in investment? Explain your answer.

e. The previous question focuses on the trade deficit rather than the current account deficit. How does net investment income (NI) relate to the difference between the trade deficit and the current account deficit in the United States? You can download GDP (series GDP) and GNP (series GNP) from the FRED database at the Federal Reserve Bank of St. Louis. This difference is a measure of NI. Is this value rising or falling over time? What is the implication of such changes?

FURTHER READINGS

- A good discussion of the relation among trade deficits, current account deficits, budget deficits, private saving, and investment is given in Barry Bosworth's *Saving and Investment in a Global Economy* (Brookings Institution, 1993).
- For more on the relation between the exchange rate and the trade balance, read "Exchange Rates and Trade Flows: Disconnected?" Chapter 3, World Economic Outlook, International Monetary Fund, October 2015.

APPENDIX: Derivation of the Marshall-Lerner Condition

Start from the definition of net exports

$$NX = X - IM/\varepsilon$$

Assume trade to be initially balanced, so that $NX = 0$ and $X = IM/\varepsilon$ or, equivalently, $\varepsilon X = IM$.

The Marshall-Lerner condition is the condition under which a real depreciation, a decrease in ε, leads to an increase in net exports.

To derive this condition, first multiply both sides of the equation above by ε to get

$$\varepsilon NX = \varepsilon X - IM$$

Now consider a change in the real exchange rate of $\Delta\varepsilon$. The effect of this change on the left side of the equation is given by $(\Delta\varepsilon)\, NX + \varepsilon\Delta(NX)$.

Note that, if trade is initially balanced, $NX = 0$, so the first term in this expression is equal to zero and the effect of the change on the left side is simply given by $\varepsilon\Delta(NX)$.

The effect of the change in the real exchange rate on the right side of the equation is given by $(\Delta\varepsilon)\, X + \varepsilon(\Delta X) - (\Delta IM)$. Putting the two sides together gives

$$\varepsilon(\Delta NX) = (\Delta\varepsilon)\, X + \varepsilon(\Delta X) - (\Delta IM)$$

Divide both sides by εX to get:

$$[\varepsilon(\Delta NX)]/\varepsilon X = [(\Delta\varepsilon)X]/\varepsilon X + [\varepsilon(\Delta X)]/\varepsilon X - [\Delta IM]/\varepsilon X$$

Simplify, and use the fact that, if trade is initially balanced, $\varepsilon X = IM$ to replace εX by IM in the last term on the right. This gives

$$(\Delta NX)/X = (\Delta\varepsilon)/\varepsilon + (\Delta X)/X - \Delta IM/IM$$

The result is intuitive. The change in the trade balance (as a ratio to exports) in response to a real depreciation is the sum of three terms:

- The first term is equal to the proportional change in the real exchange rate. It is negative if there is a real depreciation.
- The second term is equal to the proportional change in exports. It is positive if there is a real depreciation.
- The third term is equal to minus the proportional change in imports. It is positive if there is a real depreciation.

The Marshall-Lerner condition is the condition that the sum of these three terms be positive. If it is satisfied, a real depreciation leads to an improvement in the trade balance.

A numerical example will help here. Suppose that a 1% depreciation leads to a proportional increase in exports of 0.9% and a proportional decrease in imports of 0.8%. (Econometric evidence on the relation of exports and imports to the real exchange rate suggests that these are reasonable numbers.) In this case, the right-hand side of the equation is equal to $-1\% + 0.9\% - (-0.8\%) = 0.7\%$. Thus, the trade balance improves, and the Marshall-Lerner condition is satisfied.

This page is intentionally left blank

19

Output, the Interest Rate, and the Exchange Rate

In Chapter 18, we treated the exchange rate as one of the policy instruments available to the government. But the exchange rate is not a policy instrument. Rather, it is determined in the foreign exchange market, where, as you saw in Chapter 17, there is an enormous amount of trading. This fact raises two obvious questions: What determines the exchange rate? How can policymakers affect it?

These questions motivate this chapter. To answer them, we reintroduce financial markets, which we had left aside in Chapter 18. We examine the implications of equilibrium in both the goods market and financial markets, including the foreign exchange market. This allows us to characterize the joint movements of output, the interest rate, and the exchange rate in an open economy. The model we develop is an extension to the open economy of the IS-LM model you first saw in Chapter 5 and is known as the **Mundell-Fleming model**—after the two economists, Robert Mundell and Marcus Fleming, who put it together in the 1960s. (The model presented here retains the spirit of the original Mundell-Fleming model but differs in its details.)

Section 19-1 looks at equilibrium in the goods market.

Section 19-2 looks at equilibrium in financial markets, including the foreign exchange market.

Section 19-3 puts the two equilibrium conditions together and looks at the determination of output, the interest rate, and the exchange rate.

Section 19-4 looks at the role of policy under flexible exchange rates.

Section 19-5 looks at the role of policy under fixed exchange rates.

If you remember one basic message from this chapter, it should be: The effects of monetary and fiscal policies differ very much depending on the exchange rate regime. ▶▶▶

19-1 EQUILIBRIUM IN THE GOODS MARKET

Equilibrium in the goods market was the focus of Chapter 18, where we derived the equilibrium condition equation (18.4):

$$Y = \underset{(+)}{C(Y - T)} + \underset{(+,-)}{I(Y, r)} + G - \underset{(+,+)}{IM(Y, \varepsilon)}/\varepsilon + \underset{(+,-)}{X(Y^*, \varepsilon)}$$

For the goods market to be in equilibrium, output (the left side of the equation) must be equal to the demand for domestic goods (the right side of the equation). The demand for domestic goods is equal to consumption, C, plus investment, I, plus government spending, G, minus the value of imports, IM/ε, plus exports, X.

Goods market equilibrium (IS): Output = Demand for domestic goods.

- Consumption, C, depends positively on disposable income $Y-T$.
- Investment, I, depends positively on output, Y, and negatively on the real interest rate, r.
- Government spending, G, is taken as given.
- The quantity of imports, IM, depends positively on both output, Y, and the real exchange rate, ε. The value of imports in terms of domestic goods is equal to the quantity of imports divided by the real exchange rate.
- Exports, X, depend positively on foreign output, Y^*, and negatively on the real exchange rate, ε.

It will be convenient in what follows to regroup the last two terms under net exports (NX), defined as exports minus the value of imports:

$$NX\ (Y, Y^*, \varepsilon) \equiv X\ (Y^*, \varepsilon) - IM\ (Y, \varepsilon)/\varepsilon$$

It follows from our assumptions about imports and exports that net exports, NX, depend on domestic output, Y, foreign output, Y^*, and the real exchange rate, ε. An increase in domestic output increases imports, thus decreasing net exports. An increase in foreign output increases exports, thus increasing net exports. An increase in the real exchange rate leads to a decrease in net exports.

We shall assume, throughout the chapter, that the Marshall-Lerner condition holds. Under this condition, an increase in the real exchange rate—a real appreciation—leads to a decrease in net exports (see Chapter 18).

Using this definition of net exports, we can rewrite the equilibrium condition as

$$Y = \underset{(+)}{C(Y - T)} + \underset{(+,-)}{I(Y, r)} + G + \underset{(-,+,-)}{NX\,(Y, Y^*, \varepsilon)} \tag{19.1}$$

For our purposes, the main implication of equation (19.1) is that both the real interest rate and the real exchange rate affect demand and in turn equilibrium output.

- An increase in the real interest rate leads to a decrease in investment spending and, as a result, to a decrease in the demand for domestic goods. This leads, through the multiplier, to a decrease in output.
- An increase in the real exchange rate leads to a shift in demand toward foreign goods and, as a result, to a decrease in net exports. The decrease in net exports decreases the demand for domestic goods. This leads, through the multiplier, to a decrease in output.

For the remainder of the chapter, we shall simplify equation (19.1) in two ways:

- Given our focus on the short run, we assumed in our previous treatment of the IS-LM model that the (domestic) price level was given. We shall make the same assumption here and extend this assumption to the foreign price level, so the real exchange rate, $\varepsilon = EP/P^*$, and the nominal exchange rate, E, move together. A decrease in the nominal exchange rate—a nominal depreciation—leads, one-for-one, to a decrease in the real exchange rate—a real depreciation. Conversely, an increase in the nominal exchange rate—a nominal appreciation—leads, one-for-one, to an increase in the real exchange rate—a real appreciation. If, for notational convenience, we choose P and P^* so that $P/P^* = 1$ (we can do so because both are index numbers), then $\varepsilon = E$ and we can replace ε with E in equation (19.1).

- Because we take the domestic price level as given, there is no inflation, either actual or expected. Therefore, the nominal interest rate and the real interest rate are the same, and we can replace the real interest rate, r, in equation (19.1) with the nominal interest rate, i.

First simplification: $P = P^* = 1$, so $\varepsilon = E$.

Second simplification: $\pi^e = 0$, so $r = i$.

With these two simplifications, equation (19.1) becomes

$$Y = \underset{(+)}{C(Y - T)} + \underset{(+,-)}{I(Y, i)} + G + \underset{(-,+,-)}{NX(Y, Y^*, E)} \tag{19.2}$$

In words: Goods market equilibrium implies that output depends negatively on both the nominal interest rate and the nominal exchange rate.

By now you realize that the way to understand various macroeconomic mechanisms is to refine the basic model in one direction and simplify it in others (here, opening the economy but ignoring risk). Keeping all the refinements would lead to a rich model (and this is what macroeconometric models do) but would make for a terrible textbook. Things would become far too complicated.

19-2 EQUILIBRIUM IN FINANCIAL MARKETS

When we looked at financial markets in the IS-LM model, we assumed that people chose only between two financial assets, money and bonds. Now that we look at a financially open economy, we must also take into account the fact that people have a choice between domestic bonds and foreign bonds.

Remember that we have assumed that people only hold domestic currency and do not hold foreign currency, so we do not have to look at that choice.

Domestic Bonds versus Foreign Bonds

As we look at the choice between domestic and foreign bonds, we shall rely on the assumption we introduced in Chapter 17: Financial investors, domestic or foreign, go for the highest expected rate of return, ignoring risk. This implies that, in equilibrium, both domestic and foreign bonds must have the same expected rate of return; otherwise, investors would be willing to hold only one or the other, but not both, and this could not be an equilibrium. (Like all economic relations, this relation is only an approximation to reality and does not always hold. For more on this, see the Focus Box “Capital Flows, Sudden Stops, and the Limits of the Interest Parity Condition.”)

As we saw in Chapter 17 (equation (17.2)), this assumption implies that the following arbitrage relation—the *interest parity condition*—must hold:

$$(1 + i_t) = (1 + i_t^*)\left(\frac{E_t}{E_{t+1}^e}\right) \tag{19.3}$$

where i_t is the domestic interest rate, i_t^* is the foreign interest rate, E_t is the current exchange rate, and E_{t+1}^e is the expected future exchange rate. The left side of the equation gives the return, in terms of domestic currency, from holding domestic bonds. The right side of the equation gives the expected return, also in terms of domestic currency, from holding foreign bonds. In equilibrium, the two expected returns must be equal.

The presence of E_t comes from the fact that to buy the foreign bond, you must first exchange domestic currency for foreign currency. The presence of E_{t+1}^e comes from the fact that to bring the funds back next period, you will have to exchange foreign currency for domestic currency.

Multiply both sides by E_{t+1}^e and reorganize to get

$$E_t = \frac{1 + i_t}{1 + i_t^*} E_{t+1}^e \tag{19.4}$$

For now, we shall take the expected future exchange rate as given and denote it as $\bar{E}^e$ (we shall relax this assumption in Chapter 20). Under this assumption, and dropping time indexes, the interest parity condition becomes

$$E = \frac{1 + i}{1 + i^*} \bar{E}^e \tag{19.5}$$

FOCUS

Capital Flows, Sudden Stops, and the Limits to the Interest Parity Condition

The interest parity condition assumes that financial investors care only about expected returns. But as we discussed in Chapter 14, investors also care about risk and liquidity. Much of the time, one can leave aside these other factors. Sometimes, however, they play a big role in investors' decisions and in determining exchange rate movements. The perception that risk has increased leads investors to want to sell most or all the assets they have in a country, no matter what the interest rate. These selling episodes, which have affected many Latin American and Asian emerging economies in the past, are known as **sudden stops**. During these episodes, the interest parity condition fails, and the exchange rate of these emerging market countries may fall a lot without much change in domestic or foreign interest rates.

The start of the Great Recession was associated with large capital movements, which also had little to do with expected returns. Worried about uncertainty, many investors from advanced countries decided to take their funds home, where they felt safer. (Ironically, even though the crisis originated from the United States, the United States was still seen as a **safe haven**, leading many investors to sell emerging market countries' assets and buy US assets.) The result was large capital outflows from several emerging countries. An example is given in Figure 1, which shows the net purchases of Brazilian equities by foreign investors from 2000 on. Note that the flows turned sharply negative in the second half of 2008, going from nearly 30% of annual GDP to−25%, only to rebound in 2009. (Negative net purchases indicate that foreign investors sold more stocks than they bought during the quarter.)

The sharp negative flows had major effects on Brazil and other emerging market countries, leading to strong downward pressure on their exchange rates and serious problems in their financial systems. For example, domestic banks that had relied on foreign investors for funds found themselves short of funds, which forced them to cut lending to domestic firms and households. This was an important channel of transmission of the crisis from the United States to the rest of the world.

Further reading: Among the countries affected by large capital outflows in 2008 and 2009 were several small advanced economies, notably Ireland and Iceland. A number of these countries had built up the same financial vulnerabilities as the United States (those we studied in Chapter 6) and suffered badly. A good and easy read is Michael Lewis's chapters on Ireland and Iceland in *Boomerang: Travels in a New Third World* (2011, Norton Books).

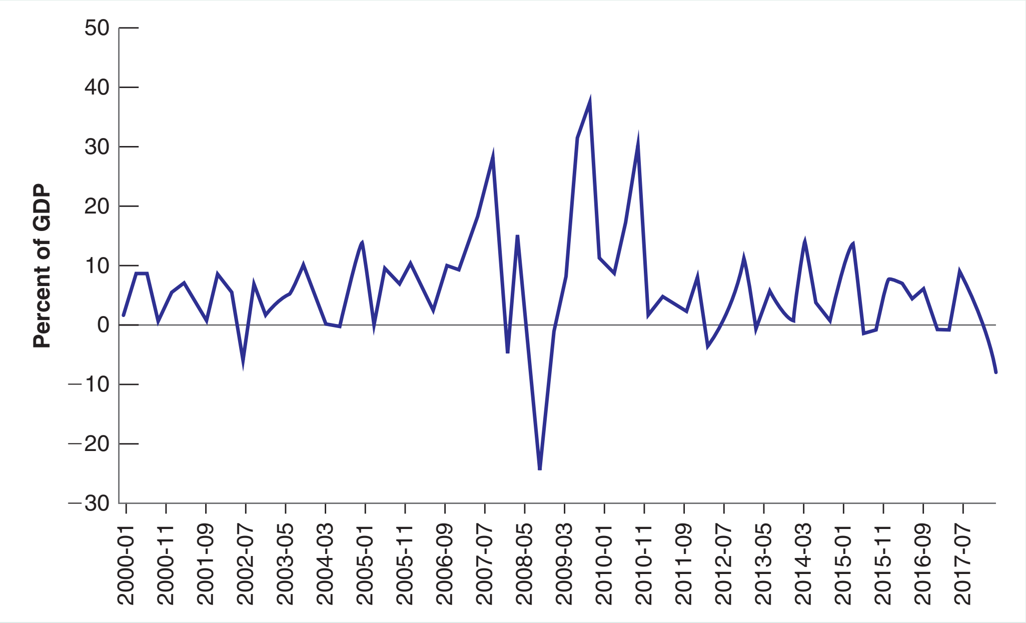


Figure 1

Net Purchases of Brazilian Equities since 2000

Source: IMF BOP statistics.

This relation tells us that the current exchange rate depends on the domestic interest rate, the foreign interest rate, and the expected future exchange rate.

- An increase in the domestic interest rate leads to an increase in the exchange rate.
- An increase in the foreign interest rate leads to a decrease in the exchange rate.
- An increase in the expected future exchange rate leads to an increase in the current exchange rate.

This relation plays a central role in the real world and will play a central role in this chapter. Consider the following example:

Think of financial investors—investors, for short—choosing between US bonds and Japanese bonds. Suppose that the one-year interest rate on both is 2%. The current exchange rate is 100 (one dollar is worth 100 yen), and the expected exchange rate a year from now is also 100. Under these assumptions, both US and Japanese bonds have the same expected return in dollars, and the interest parity condition holds.

Suppose that investors now expect the exchange rate, for whatever reason, to be 10% higher a year from now, so $\bar{E}^e$ is now equal to 110. At an unchanged current exchange rate, US bonds are now much more attractive than Japanese bonds. Both offer an interest rate of 2%, in dollars or yen, but a year from today the yen are now expected to be worth 10% less in terms of dollars. In terms of dollars, the return on Japanese bonds is therefore 2% (the interest rate) -10% (the expected depreciation of the yen relative to the dollar), or -8%.

An increase in $\bar{E}^e$ is an expected appreciation of the dollar relative to the yen. Equivalently, it is an expected depreciation of the yen relative to the dollar.

What, then, will happen to the current exchange rate? At the initial exchange rate of 100, investors will want to shift out of Japanese bonds into US bonds. To do so, they must first sell Japanese bonds for yen, then sell yen for dollars, and then use the dollars to buy US bonds. As investors sell yen and buy dollars, the dollar will appreciate relative to the yen. By how much? Equation (19.5) gives the answer: $E = (1.02/1.02)\ 110 = 110$. The current exchange rate must increase in the same proportion as the expected future exchange rate. Put another way, the dollar must appreciate today by 10%. When it has appreciated by 10%, so $E = \bar{E}^e = 110$, the expected returns on US and Japanese bonds are again equal, and there is equilibrium in the foreign exchange market.

Suppose instead that the Fed raises the domestic interest rate from 2% to 5%. Assume that the Japanese interest rate remains unchanged at 2%, and the expected future exchange rate remains unchanged at 100. At an unchanged current exchange rate, US bonds are again more attractive than Japanese bonds. US bonds yield a return of 5% in dollars. Japanese bonds give a return of 2% in yen, and—because the exchange rate is expected to be the same next year as it is today—an expected return of 2% in dollars as well.

Now what will happen to the current exchange rate? Again, at the initial exchange rate of 100, investors will want to shift out of Japanese bonds into US bonds. As they do so, they sell yen for dollars and the dollar will appreciate. By how much? Equation (19.5) gives the answer: $E = (1.05/1.02)\ 100 \approx 103$. The current exchange rate will increase by approximately 3%.

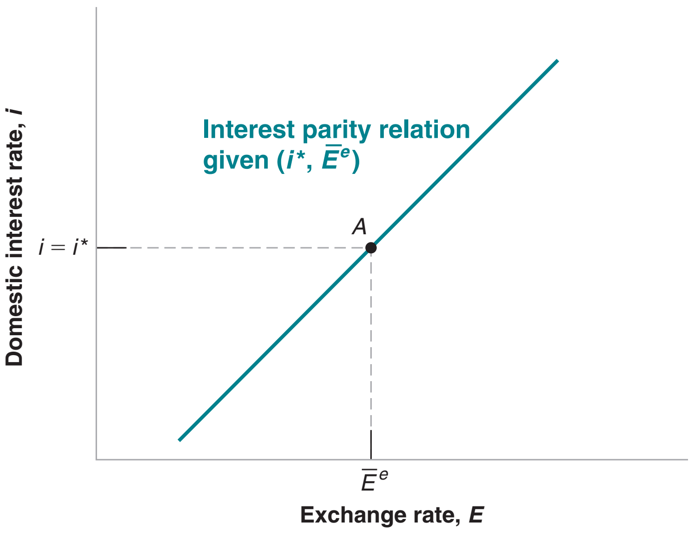
Why 3%? Think of what happens when the dollar appreciates. If, as we have assumed, investors do not change their expectation of the future exchange rate, then the more the dollar appreciates today, the more investors expect it to depreciate in the future (as it is expected to return to the same value in the future). When the dollar has appreciated by 3% today, investors expect it to depreciate by 3% during the coming year. Equivalently, they expect the yen to appreciate relative to the dollar by 3% over the coming year. The expected rate of return in dollars from holding Japanese bonds is therefore 2% (the interest rate in yen) + 3% (the expected yen appreciation), or 5%. This expected rate of return is the same as the rate of return on holding US bonds, so there is equilibrium in the foreign exchange market.

Make sure you understand the argument. Why doesn't the dollar appreciate by, say, 20%?

Figure 19-1

The Relation between the Interest Rate and the Exchange Rate Implied by Interest Parity

A higher domestic interest rate leads to a higher exchange rate—an appreciation.



Note that our argument relies heavily on the assumption that, when the interest rate changes, the expected exchange rate remains unchanged. This implies that an appreciation today leads to an expected depreciation in the future because the exchange rate is expected to return to the same, unchanged, value. We shall relax the assumption that the expected future exchange rate is fixed in Chapter 20. But the basic conclusion will remain: *An increase in the domestic interest rate relative to the foreign interest rate leads to an appreciation of the domestic currency.*

Figure 19-1 plots the relation between the domestic interest rate, i, and the exchange rate, E, implied by equation (19.5)—the interest parity relation. The relation is drawn for a given expected future exchange rate, $\bar{E}^e$, and a given foreign interest rate, i^*, and is represented by an upward-sloping line. The higher the domestic interest rate, the higher the exchange rate. Equation (19.5) also implies that when the domestic interest rate is equal to the foreign interest rate ($i = i^*$), the exchange rate is equal to the expected future exchange rate ($E = \bar{E}^e$). This implies that the line corresponding to the interest parity condition goes through point A (where $i = i^*$) in the figure.

What happens to the line if (1) i^* increases? (2) $\bar{E}^e$ increases?

19-3 PUTTING GOODS AND FINANCIAL MARKETS TOGETHER

We now have the elements we need to understand the movements of output, the interest rate, and the exchange rate.

Goods market equilibrium implies that output depends, among other factors, on the interest rate and the exchange rate:

$$Y = C(Y - T) + I(Y, i) + G + NX(Y, Y^*, E)$$

Let's think of the interest rate, i, as the policy rate set by the central bank:

$$i = \bar{i}$$

And the interest parity condition implies a positive relation between the domestic interest rate and the exchange rate:

$$E = \frac{1 + i}{1 + i^*} \bar{E}^e$$

Together, these three relations determine output, the interest rate, and the exchange rate. Working with three equations and three variables is not easy. But we can easily reduce them to two by using the interest parity condition to eliminate the exchange rate in the goods market equilibrium relation. Doing this gives us the following two equations, the open economy versions of the IS and LM relations:

$$IS: \quad Y = C(Y - T) + I(Y, i) + G + NX\left(Y, Y^*, \frac{1 + i}{1 + i^*}\bar{E}^e\right)$$

$$LM: \quad i = \bar{i}$$

Together, the two equations determine the interest rate and equilibrium output. Using equation (19.5) then gives us the implied exchange rate. Take the IS relation first and consider the effects of an increase in the interest rate on output. An increase in the interest rate now has two effects:

- The first effect, which was already present in a closed economy, is the direct effect on investment. A higher interest rate leads to a decrease in investment, a decrease in the demand for domestic goods, and a decrease in output.
- The second effect, which is present in the open economy, is the effect through the exchange rate. A higher interest rate leads to an increase in the exchange rate—an appreciation. The appreciation, which makes domestic goods more expensive relative to foreign goods, leads to a decrease in net exports, and therefore a decrease in the demand for domestic goods and a decrease in output.

Both effects work in the same direction. An increase in the interest rate decreases demand directly and indirectly—through the adverse effect of the appreciation on demand.

The IS relation between the interest rate and output is drawn in Figure 19-2(a), for given values of all the other variables in the relation—T, G, Y^*, i^*,and $\bar{E}^e$. The IS curve is downward sloping: An increase in the interest rate leads to lower output. The curve looks much the same as in the closed economy, but it hides a more complex relation than before. The interest rate affects output not only directly but also indirectly through the exchange rate.

An increase in the interest rate leads, both directly and indirectly (through the exchange rate), to a decrease in output.

The LM relation is the same as in the closed economy: it is a horizontal line, at the level of the interest rate $\bar{i}$ set by the central bank.

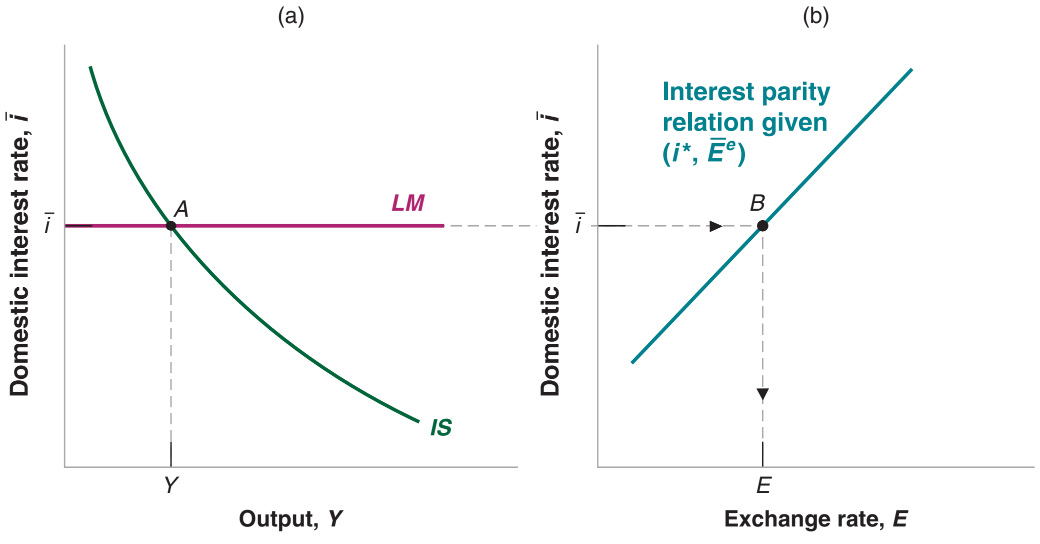


Figure 19-2

The IS-LM Model in an Open Economy

An increase in the interest rate reduces output both directly and indirectly (through the exchange rate). The IS curve is downward sloping, for both reasons. The LM curve is horizontal, as in Chapter 6.

Equilibrium in the goods and financial markets is attained at point A in Figure 19-2(a), with output level Y and interest rate $\bar{i}$. The equilibrium value of the exchange rate cannot be read directly from the graph. But it is easily obtained from Figure 19-2(b), which replicates Figure 19-1 and gives the exchange rate associated with a given interest rate found at point B, given the foreign interest rate i^* and the expected exchange rate. The exchange rate associated with the equilibrium interest rate $\bar{i}$ is equal to E.

Let's summarize. We have derived the IS and LM relations for an open economy:

The IS curve is downward sloping. An increase in the interest rate leads both directly and indirectly (through the exchange rate) to a decrease in demand and a decrease in output.

The LM curve is horizontal at the interest rate set by the central bank.

Equilibrium output and the equilibrium interest rate are given by the intersection of the IS and LM curves. Given the foreign interest rate and the expected future exchange rate, the equilibrium interest rate determines the equilibrium exchange rate.

19-4 THE EFFECTS OF POLICY IN AN OPEN ECONOMY

Having derived the IS-LM model for the open economy, we can now put it to use and look at the effects of monetary and fiscal policy.

The Effects of Monetary Policy in an Open Economy

Let's start from the effects of the central bank's decision to increase the domestic interest rate. Look at Figure 19-3(a). At a given level of output, with a higher interest rate, the LM curve shifts up, from LM to LM'. The IS curve does not shift (remember that the IS curve shifts only if G or T or Y^* or i^* change). The equilibrium moves from point A to point A'. In Figure 19-3(b), the increase in the interest rate leads to an appreciation.

A monetary contraction shifts the LM curve up. It shifts neither the IS curve nor the interest parity curve.

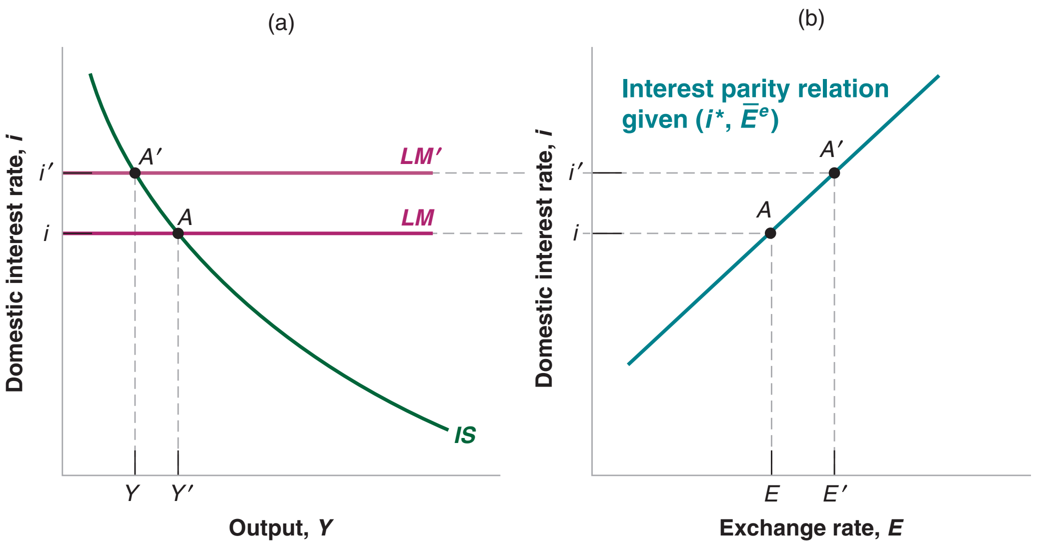
So, in an open economy, monetary policy works through two channels: First, as in the closed economy, it works through the effect of the interest rate on spending; second, it works through the effect of the interest rate on the exchange rate and the effect of the exchange rate on exports and imports. Both effects work in the same direction. In the case of a monetary contraction, the higher interest rate and the appreciation decrease both demand and output.

Can you tell what happens to net exports? (The answer is no: They can go either way. Make sure you understand why.)

Figure 19-3

The Effects of an Increase in the Interest Rate

An increase in the interest rate leads to a decrease in output and an appreciation.



The Effects of Fiscal Policy in an Open Economy

Let's look now at a change in government spending. Suppose that, starting from a balanced budget, the government decides to increase defense spending without raising taxes and so runs a budget deficit. What happens to the level of output? To the composition of output? To the interest rate? To the exchange rate?

Let us first assume that before the increase in government spending, the level of out put, Y, was below potential. If the increase in G moves output toward potential, but not above potential, the central bank will not be worried that inflation might increase (remember our discussion in Chapter 9, particularly Figure 9-2 and will keep the interest rate unchanged. What happens to the economy is described in Figure 19-4. The economy is initially at point A. The increase in government spending, $\Delta G > 0$,increases output at a given interest rate, shifting the IS curve to the right, from *IS* to *IS′* in Figure 19-4(a). Because the central bank does not change the policy rate, the LM curve does not shift. The new equilibrium is at point A', with a higher level of output, Y'. In panel (b), because the interest rate has not changed, neither has the exchange rate. So an increase in government spending, when the central bank keeps the interest rate unchanged, leads to an increase in output with no change in the exchange rate.

An increase in government spending shifts the IS curve to the right. It shifts neither the LM curve nor the interest parity line.

Can we tell what happens to the various components of demand?

- Clearly, both consumption and government spending increase: Consumption goes up because of the increase in income; government spending goes up by assumption.
- Investment also rises because it depends on both output and the interest rate: $I = I(Y, i)$. Here output rises and the interest rate does not change, thus investment rises.
- What about net exports? Recall that net exports depend on domestic output, foreign output, and the exchange rate: $NX = NX(Y, Y^*, E)$. Foreign output is unchanged, as we are assuming that the rest of the world does not respond to the increase in domestic government spending. The exchange rate is also unchanged, because the interest rate does not change. We are left with the effect of higher domestic output; as the increase in output increases imports at an unchanged exchange rate, net exports decrease. As a result, the budget deficit leads to a deterioration of the trade balance. If trade was balanced to start, then the budget deficit leads to a trade deficit. Note that, although an increase in the budget deficit increases the trade deficit, the effect is far from mechanical. It works through the effect of the budget deficit on output and, in turn, on the trade deficit.

Now assume instead that the increase in G happens in an economy where output is close to potential output, Yn. The government could decide to increase government spending even if the economy is already at potential output because, for example, it

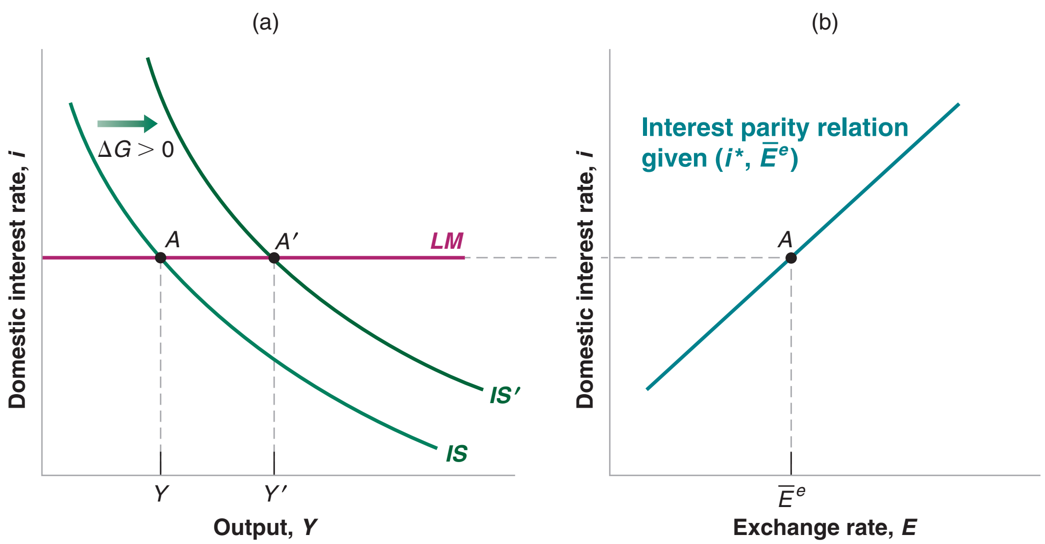


Figure 19-4

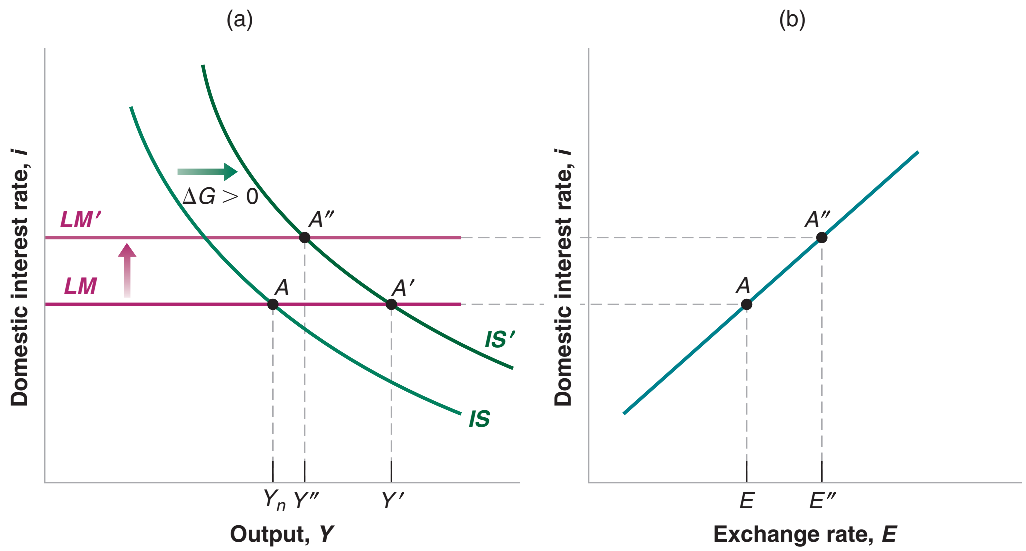
The Effects of an Increase in Government Spending with an Unchanged Interest Rate

An increase in government spending leads to an increase in output. If the central bank keeps the interest rate unchanged, the exchange rate also remains unchanged.

Figure 19-5

The Effects of an Increase in Government Spending When the Central Bank Responds by Raising the Interest Rate

An increase in government spending leads to an increase in output. If the central bank responds by raising the interest rate, the exchange rate will appreciate.



needs to pay for an exceptional event, such as a big flood, and wants to postpone tax increases. Or it could do it for political reasons, because it wants to increase spending but does not want to increase taxes (more on this in Chapter 22). In this case the central bank will worry that the increase in G, by moving the economy above potential output, might push inflation up. It is likely to respond by raising the interest rate. What happens then is described in Figure 19-5. At an unchanged interest rate, output would increase from Y_n to Y' and the exchange rate would not change. But if the central bank accompanies the increase in government spending with an increase in the interest rate, output will increase by less, from Y_n to Y'', and the exchange rate will appreciate, from E to E''.

Again, can we tell what happens to the various components of demand?

- As before, both consumption and government spending increase: consumption goes up because of the increase in income, and government spending goes up by assumption.
- What happens to investment is now ambiguous. Investment depends on both output and the interest rate: $I = I(Y, i)$. Here output rises, but so does the interest rate.
- Net exports decrease for two reasons: Output goes up, increasing imports; the exchange rate appreciates, increasing imports and decreasing exports. The budget deficit leads to a trade deficit. (Whether, however, the trade deficit is larger than if the policy rate remained constant is ambiguous. The appreciation makes it worse, but the higher interest rate leads to a smaller increase in output and thus a smaller increase in imports.)

This version of the IS-LM model for the open economy was first put together in the 1960s by the two economists I mentioned at the outset of the chapter, Robert Mundell, at Columbia University, and Marcus Fleming, at the International Monetary Fund—although their model reflected the economies of the 1960s, when central banks used to set the supply of money, M, rather than the interest rate as they do today (remember our discussions in Chapters 4 and 6), so their model was slightly different from the model presented here.

Robert Mundell was awarded the Nobel Prize in Economics in 1999. ►

How well does the Mundell-Fleming model fit the facts? Typically, quite well, and this is why it is still in use. Like all simple models, it often needs to be extended. One can incorporate, for example, the role of risk in affecting portfolio decisions, or the implications of the zero lower bound, two important aspects of the financial crisis. But the simple exercises we worked through in Figures 19-3, 19-4, and 19-5 are a good starting point to organize thoughts. How the model can be used to interpret events or think about policy is shown in two Focus Boxes. The first looks at the effects of the combination of monetary contraction and fiscal expansion that took place in the United States in the early 1980s. The second looks at whether we can expect the trade tariffs introduced by the Trump administration to reduce the trade deficit.

FOCUS

Monetary Contraction and Fiscal Expansion: The United States in the Early 1980s

The early 1980s in the United States were dominated by sharp changes in both monetary policy and in fiscal policy.

In the late 1970s, the Chairman of the Fed, Paul Volcker, concluded that US inflation was too high and had to be reduced. Starting in late 1979, he embarked on a path of sharp increases in interest rates, realizing this might lead to a recession in the short run but lower inflation in the medium run.

The change in fiscal policy was triggered by the election of Ronald Reagan in 1980. Reagan was elected on the promise of more conservative policies, namely lower taxation and a scaling back of the government's role in economic activity. This commitment led to the Economic Recovery Tax Act of August 1981. Personal income taxes were cut by a total of 23% in three installments from 1981 to 1983. Corporate taxes were also reduced. These tax cuts were not, however, accompanied by corresponding decreases in government spending, and the result was a steady increase in budget deficits

What were the Reagan administration's motivations for cutting taxes without implementing corresponding cuts in spending? These are still being debated, but there is agreement that there were two main motivations:

One motivation came from the beliefs of a fringe, but influential, group of economists called the **supply siders**, who argued that a cut in tax rates would cause people and firms to work much harder and more productively, and that the resulting increase in activity would actually lead to an increase, not a decrease, in tax revenues. Using the terminology of this textbook: They thought that potential output would substantially increase, actual output would follow, and tax revenues would increase. Whatever the merits of the argument appeared to be then, it proved wrong. Even if some people did work harder and more productively after the tax cuts, tax revenues decreased and the fiscal deficit increased.

The other motivation was more cynical. It was a bet that the cut in taxes, and the resulting increase in deficits, would scare Congress into cutting spending or, at the least, into not increasing spending—a strategy known as "starve the beast." This motivation turned out to be partly right: Congress found itself under enormous pressure not to increase spending, and the growth of spending in the 1980s was surely lower than it would have been otherwise. Nonetheless, the adjustment of spending was not enough to offset the shortfall in tax revenues and avoid the rapid increase in deficits.

Whatever the reason for the deficits, the combined effects of higher interest rates and a fiscal expansion were very much in line with what the Mundell-Fleming model predicts. Table 1 shows the evolution of the main macroeconomic variables from 1980 to 1984.

From 1980 to 1982, the evolution of the economy was dominated by the effects of the increase in interest rates, both nominal and real. They increased sharply, leading to both a large dollar appreciation and a recession. The goal of lowering inflation was achieved; by 1982, inflation was down to about 4%, down from 12.5% in 1980. Lower output and the dollar appreciation had opposing effects on the trade balance (lower output leading to lower imports and an improvement in the trade balance; the appreciation of the dollar leading to a deterioration in the trade balance), resulting in little change in the trade deficit until 1983.

From 1983 on, the evolution of the economy was dominated by the effects of the fiscal expansion. As our model predicts, these effects were strong output growth, high interest rates, and further dollar appreciation. High output growth and the dollar appreciation led to an increase in the trade deficit to 2.7% of GDP by 1984. By the mid-1980s, the main macroeconomic policy issue had become that of the **twin deficits**: the budget deficit and the trade deficit. The twin deficits were to remain a central macroeconomic concern throughout the 1980s and early 1990s.

Table 1 Major US Macroeconomic Variables, 1980–1984

	1980	1981	1982	1983	1984
GDP growth (%)	−0.5	1.8	−2.2	3.9	6.2
Unemployment rate (%)	7.1	7.6	9.7	9.6	7.5
Inflation (CPI) (%)	12.5	8.9	3.8	3.8	3.9
Nominal interest rate (%)	11.5	14.0	10.6	8.6	9.6
Real interest rate (%)	2.5	4.9	6.0	5.1	5.9
Real exchange rate	85	101	111	117	129
Trade surplus (% of GDP)	−0.5	−0.4	−0.6	−1.5	−2.7

Inflation: rate of change of the CPI. The nominal interest rate is the three-month T-bill rate. The real interest rate is equal to the nominal rate minus the forecast of inflation by DRI, a private forecasting firm. The real exchange rate is the trade-weighted real exchange rate, normalized so that 1973 = 100. A negative trade surplus is a trade deficit.

FOCUS

US Trade Deficits and Trump Administration Trade Tariffs

One of the themes of President Trump's 2016 campaign was that the US trade deficit, which stood at 2.7% of GDP at the time, had to be reduced. He saw the trade deficit as an indication that foreign countries, China in particular, were taking advantage of the United States. In 2018, following up on this promise, the Trump administration increased tariffs on solar panels, then on steel and aluminum imports, and then on a range of Chinese goods.

Most economists disagreed with both the diagnosis and the method.

About the diagnosis: Most economists saw the trade deficits as reflecting chronically low US saving relative to investment, together with the willingness of foreigners to lend to the United States, allowing for the deficits to be easily financed through foreign borrowing. Thus, they argued, trade deficits reflected a US problem and had to be solved by increasing domestic saving, private or public. Some economists went further and argued that the origin of the trade deficit was actually the attractiveness of US assets. Foreigners were eager to hold US assets, and this foreign demand led to dollar appreciation, decreasing in turn the relative competitiveness of US goods and leading to the trade deficit. (You may want to return to the discussion of the relation between trade deficits, saving, and investment in Section 18-5.) Thus, these economists argued, the trade deficit reflected the financial strength of the United States and was not particularly worrisome.

About the method: Economists argued that, if the origin of the trade deficit lay in low US saving or in the attractiveness of US assets, tariffs were unlikely to work. Let's see why.

At a given exchange rate, an increase in tariffs increases the price of imports for US consumers. How much depends on what foreign firms do in reaction to the tariff:

They can decrease their pre-tariff price, in which case the price facing US consumers increases by less than the tariff. They can even decrease the pre-tariff price enough that the price facing US consumers does not change. In this case, the volume of imports may not change, but how much the United States pays for imports (the pre-tariff price) goes down, so the value of imports in terms of domestic goods decreases.

Or they can instead keep the same pre-tariff price, in which case the price facing US consumers increases by the increase in the tariff, and the demand for imports is then likely to decrease. In this case, the price that the United States pays for imports does not change, but the volume of imports decreases, leading again to a decrease in the value of imports.

In both cases, at a given nominal exchange rate, the value of imports in terms of domestic goods decreases. Given exports, this implies a smaller trade deficit. (There is a difference, however, in terms of who pays for the tariffs. In the first case, the foreign firms pay; in the second case, the US consumers pay.) The smaller trade deficit leads to an increase in demand and an increase in output. It all seems to work. So why were the economists skeptical that the tariffs would not reduce the trade deficit?

For four reasons:

- To the extent that it triggered a tariff war, and other countries responded by raising tariffs on US goods, US exports might decrease in line with the decrease in US imports, leading to no change in the trade deficit. And, indeed, in response to US tariffs, China increased tariffs on US goods in late 2018.
- To the extent that the US economy was close to potential, as was indeed the case in 2018, an increase in output might lead to overheating and thus force the Fed to increase the interest rate. This in turn would lead to a dollar appreciation (remember the interest parity condition), partially cancelling the effects of the tariffs on net exports.
- Even if the Fed did not increase the interest rate, expectations of a lower trade deficit and thus a smaller need for foreign borrowing, now and in the future, may lead to an appreciation of the dollar, again partially offsetting the effects of the tariffs on net exports.
- Finally, while it was increasing tariffs, the Trump administration also implemented a tax reform passed in 2017, which led to a large increase in the fiscal deficit in 2018. As we saw in this chapter, a larger fiscal deficit is likely to lead to a larger trade deficit: If the Fed did not intervene, the fiscal deficit would lead to larger output and higher imports. If the Fed intervened to limit the increase in output, the increase in interest rates would lead to a dollar appreciation, and, again, to a larger trade deficit. In either case, fiscal policy was likely to offset or even dominate the effects of tariffs on the trade balance.

What happened? At the time of writing, it is too early to draw strong conclusions. Trade negotiations are still going on. It takes time for exporters and importers to react to tariffs and exchange rate movements. So far, however, the evidence is on the side of the economists. This is shown in Figure 1, which plots the evolution of the ratio of net exports to GDP and of the real exchange rate (normalized to 100 in 2010), quarterly, since 2010.

The shaded area corresponds to the four quarters of 2018. Since the beginning of 2018, the real exchange rate (left-hand vertical axis) has increased by close to 9%.

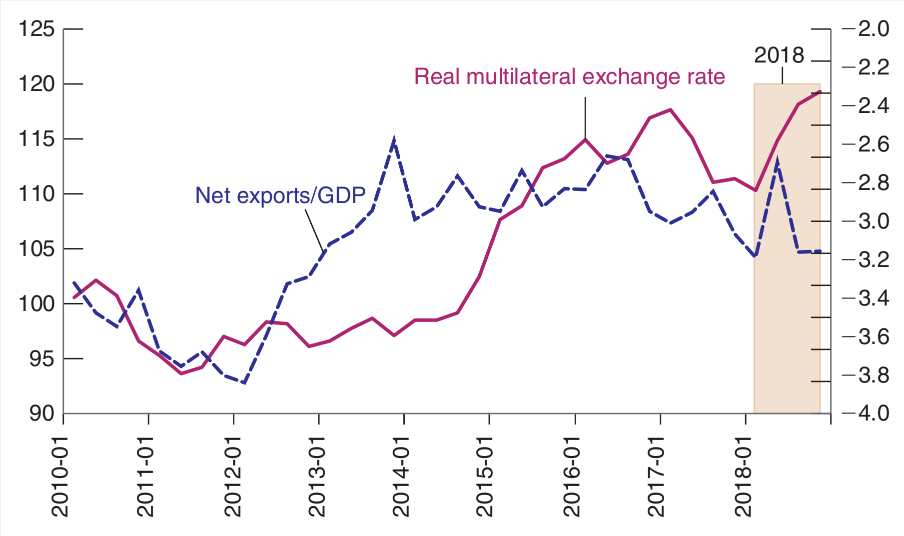


Figure 1

Net exports/GDP and the real multilateral exchange rate since 2010

Source: FRED: NETEXP, GDP, RBUSBIS

The trade deficit (right-hand vertical axis) is roughly the same as it was at the beginning of 2018, around 3% of GDP. For the moment, things have not moved in the direction the Trump administration hoped for. (What would you do if you wanted to reduce the US trade deficit while keeping output at potential?)

19-5 FIXED EXCHANGE RATES

We have assumed so far that the central bank chose the interest rate and let the exchange rate adjust freely in whatever manner was implied by equilibrium in the foreign exchange market. In many countries, this assumption does not reflect reality. Central banks act under implicit or explicit exchange rate targets and use monetary policy to achieve those targets. The targets are sometimes implicit, sometimes explicit; they are sometimes specific values, sometimes bands or ranges. These exchange rate arrangements (or *regimes*, as they are called) come under many names. Let's first see what the names mean.

Pegs, Crawling Pegs, Bands, the EMS, and the Euro

At one end of the spectrum are countries with flexible exchange rates such as the United States, the United Kingdom, Japan, and Canada. These countries have no explicit exchange rate targets. Although their central banks do not ignore movements in the exchange rate, they have shown themselves quite willing to let their exchange rates fluctuate considerably.

At the other end are countries that operate under *fixed exchange rates.* They maintain a fixed exchange rate in terms of some foreign currency. Some **peg** their currency to the dollar. For example, from 1991 to 2001, Argentina pegged its currency, the peso, at the highly symbolic exchange rate of one dollar for one peso (more on this in Chapter 20). Other countries used to peg their currency to the French franc (most of these are former French colonies in Africa); as the French franc has been replaced by the euro, they are now pegged to the euro. Still other countries peg their currency to a basket of foreign currencies, with the weights reflecting the composition of their trade.

The label *fixed* is a bit misleading. It is not the case that the exchange rate in countries with a fixed exchange rate never actually changes. But changes are rare. An extreme case is that of the African countries pegged to the French franc. When their exchange rates were readjusted in January 1994, it was the first adjustment in 45 years! Because these changes are rare, economists use specific words to distinguish them from the daily changes that occur under flexible exchange rates. A decrease in the exchange rate under a regime of fixed exchange rates is called a *devaluation* rather than a depreciation, and an increase in the exchange rate under a regime of fixed exchange rates is called a *revaluation* rather than an appreciation.

These terms were introduced in Chapter 17.

Between these extremes are countries with various degrees of commitment to an exchange rate target. For example, some countries operate under a **crawling peg**. The name describes it well. These countries typically have inflation rates that exceed the US inflation rate. If they were to peg their nominal exchange rate against the dollar, the more rapid increase in their domestic price level above the US price level would lead to a steady real appreciation and rapidly make their goods uncompetitive. To avoid this effect, these countries choose a predetermined rate of depreciation against the dollar: They choose to "crawl" (move slowly) vis-à-vis the dollar.

Recall the definition of the real exchange rate $\varepsilon = EP/P^*$. If domestic inflation is higher than foreign inflation: P increases faster than P^*. If E is fixed, EP/P^* steadily increases.

Yet another arrangement is for a group of countries to maintain their bilateral exchange rates (the exchange rate between each pair of countries) within some bands. Perhaps the most prominent example was the **European Monetary System (EMS)**, which determined the movements of exchange rates in the European Union from 1978 to 1998. Under the EMS rules, member countries agreed to maintain their exchange rate relative to the other currencies in the system within narrow limits or **bands** around a **central parity**—a given value for the exchange rate. Changes in the central parity and devaluations or revaluations of specific currencies could occur, but only by common agreement among member countries. After a major crisis in 1992, which led several countries to drop out of the EMS, exchange rate adjustments became more and more infrequent, leading a number of countries to move one step further and adopt a common currency, the **euro**. The conversion from domestic currencies to the euro began on January 1, 1999, and was completed in early 2002. We shall return to the implications of the move to the euro in Chapter 20.

We shall look at the 1992 crisis in Chapter 20.

You can think of countries adopting a common currency as adopting an extreme form of fixed exchange rates. Their "exchange rate" is fixed at one-to-one between any pair of countries.

We shall discuss the pros and cons of different exchange regimes in the next chapter. But first, we must understand how pegging (also called *fixing*) the exchange rate affects monetary policy and fiscal policy. This is what we do in the rest of this section.

Monetary Policy When the Exchange Rate Is Fixed

Suppose a country decides to peg its exchange rate at some chosen value; call it $\bar{E}$. How does it achieve this? The government cannot just announce the value of the exchange rate and remain idle. Rather, it must take measures so that its chosen exchange rate will prevail in the foreign exchange market. Let's look at the implications and mechanics of pegging.

Pegging or no pegging, the exchange rate and the nominal interest rate must satisfy the interest parity condition

$$(1 + i_t) = (1 + i_t^*)\left(\frac{E_t}{E_{t+1}^e}\right)$$

When the country pegs its exchange rate at $\bar{E}$, the current exchange rate $E_t = \bar{E}$. If financial and foreign exchange markets believe that the exchange rate will remain pegged at this value, then their expectation of the future exchange rate, E_{t+1}^e, is also equal to $\bar{E}$, and the interest parity relation becomes

$$(1 + i_t) = (1 + i_t^*) \Rightarrow i_t = i_t^*$$

In words: If financial investors expect the exchange rate to remain unchanged, they will require the same nominal interest rate in both countries. *Under a fixed exchange rate and perfect capital mobility, the domestic interest rate must be equal to the interest rate of the foreign country the country is pegging to.*

Let's summarize. *Under fixed exchange rates, the central bank gives up monetary policy as a policy instrument.* With a fixed exchange rate, the domestic interest rate must be equal to the foreign interest rate.

These results depend on the interest rate parity condition, which in turn depends on the assumption of perfect capital mobility—that financial investors go for the highest expected rate of return. The case of fixed exchange rates with imperfect capital mobility, which is more relevant for middle-income countries such as those in Latin America or Asia, is treated in the appendix to this chapter.

Fiscal Policy When the Exchange Rate Is Fixed

If monetary policy can no longer be used under fixed exchange rates, what about fiscal policy?

The effects of an increase in government spending when the central bank pegs the exchange rate are identical to those we saw in Figure 19-4 for the case of flexible exchange rates and an unchanged monetary policy. This is because, under flexible exchange rates, if the increase in spending, is not accompanied by a change in the interest rate, the exchange rate doesn't move. Thus, when government spending increases, it makes no difference whether the country pegs its exchange rate. The difference between fixed and flexible exchange is the ability of the central bank to respond. We saw in Figure 19-5 that if the increase in government spending pushed the economy above potential output, thus raising the possibility that inflation might increase, the central bank could respond by raising the interest rate. This option is no longer available under fixed exchange rates because the interest rate must be equal to the foreign rate.

Under flexible exchange rates the central bank can respond to an increase in government spending by raising the interest rate, as in Figure 19-5. This option is not available under fixed exchange rates because the interest rate must be equal to the foreign rate.

As this chapter comes to an end, a question should have started to form in your mind: Why would a country choose to fix its exchange rate? You have seen several reasons why this appears to be a bad idea:

- By fixing the exchange rate, a country gives up a powerful tool for correcting trade imbalances or changing the level of economic activity.
- By committing to a given exchange rate, a country also gives up control of its interest rate. Not only that, but the country must match movements in the foreign interest rate, at the risk of unwanted effects on its own activity. This is what happened in the early 1990s in Europe. Because of the increase in demand as a result of the reunification of West and East Germany, Germany felt it had to increase its interest rate. To maintain their parity with the Deutsche Mark, other countries in the European Monetary System (EMS) were forced to increase their interest rates, something they would rather have avoided. (This is the topic of the Focus Box "German Reunification, Interest Rates, and the EMS.")
- Although the country retains control of fiscal policy, one policy instrument may not be enough. As you saw in Chapter 18, a fiscal expansion can help the economy get out of a recession, but only at the cost of a larger trade deficit. And a country that wants, for example, to decrease its budget deficit cannot, under a fixed exchange rate, use monetary policy to offset the contractionary effect of its fiscal policy on output.

Why, then, do some countries fix their exchange rate? Why have 19 European countries—with perhaps more to come—adopted a common currency? To answer these questions, we must do some more work. We must look at what happens not only in the short run—which is what we did in this chapter—but also in the medium run, when the price level can adjust. And we must look at the nature of exchange rate crises. Once we have done this, we shall be able to assess the pros and cons of different exchange rate regimes. These are the topics we take up in Chapter 20.

FOCUS

German Reunification, Interest Rates, and the EMS

Under a fixed exchange rate regime such as the European Monetary System (EMS)—the system that prevailed before the introduction of the euro—no individual country can change its interest rate if the other countries do not change theirs as well. How, then, do interest rates actually change? Two arrangements are possible. One is for all the member countries to coordinate changes in their interest rates. The other is for one of the countries to take the lead and for the other countries to follow—this is in effect what happened in the EMS, with Germany as the leader.

During the 1980s, most European central banks shared similar goals and were happy to let the Bundesbank (the German central bank) take the lead. But in 1990, German unification led to a sharp divergence in goals between the Bundesbank and the central banks of the other EMS countries. Large budget deficits triggered by transfers to people and firms in Eastern Germany, together with an investment boom, led to a large increase in demand in Germany. The Bundesbank's fear that this shift would generate too strong an increase in activity led it to adopt a restrictive monetary policy. The result was strong growth in Germany together with a large increase in interest rates.

This may have been the right policy mix for Germany, but for the other European countries, it was much less appealing. They were not experiencing the same increase in demand, but to stay in the EMS they had to match German interest rates. The net result was a sharp decrease in demand and output in the other countries. These results are presented in Table 1, which gives nominal interest rates, real interest rates, inflation rates, and GDP growth from 1990 to 1992 for Germany and two of its EMS partners, France and Belgium.

Note first that the high German nominal interest rates were matched by both France and Belgium. In fact, nominal interest rates were actually higher in France than in Germany in all three years! This is because France needed higher interest rates than Germany to maintain the Deutsche Mark (DM)/franc parity: Worried about a possible devaluation of the franc, financial investors required a higher interest rate on French bonds than on German bonds (more on this in the Focus Box on the EMS crisis in Chapter 20).

Although France and Belgium had to match—or, as we have just seen, more than match—German nominal rates, both countries had less inflation than Germany. The result was very high real interest rates, much higher than the rate in Germany: In both France and Belgium, average real interest rates from 1990 to 1992 were close to 7%. And in both countries, the period 1990–1992 was characterized by slow growth and rising unemployment. In France unemployment in 1992 was 10.4%, up from 8.9% in 1990; the corresponding numbers for Belgium were 12.1% and 8.7%.

A similar story was unfolding in the other EMS countries. By 1992, average unemployment in the European Union, which had been 8.7% in 1990, had increased to 10.3%. The effects of high real interest rates on spending were not the only source of this slowdown, but they were the main one.

By 1992, an increasing number of countries were wondering whether to keep defending their EMS parity or to give it up and lower their interest rates. Worried about the risk of devaluations, financial markets started to ask for higher interest rates in countries where they thought devaluations were more likely. The result was two major exchange rate crises, one in the fall of 1992 and the other in the summer of 1993. By the end of these two crises, two countries, Italy and the United Kingdom, had left the EMS. We shall look at these crises, their origins, and their implications in Chapter 20.

Table 1 Interest Rates, Inflation, and Output Growth after German Reunification: Germany, France, and Belgium, 1990–1992

	Nominal Interest Rates (%)			*Inflation (%)*		
	1990	1991	1992	1990	1991	1992
Germany	8.5	9.2	9.5	2.7	3.7	4.7
France	10.3	9.6	10.3	2.9	3.0	2.4
Belgium	9.6	9.4	9.4	2.9	2.7	2.4
	Real Interest Rates (%)			*GDP Growth (%)*		
	1990	1991	1992	1990	1991	1992
Germany	5.8	5.5	4.8	5.7	4.5	2.1
France	7.4	6.6	7.9	2.5	0.7	1.4
Belgium	6.7	6.7	7.0	3.3	2.1	0.8

The nominal interest rate is the short-term nominal interest rate. The real interest rate is the realized real interest rate over the year—that is, the nominal interest rate minus actual inflation over the year. All rates are annual.

Source: OECD Economic Outlook.

SUMMARY

- In an open economy, the demand for domestic goods, and in turn output, depends on both the interest rate and the exchange rate. An increase in the interest rate decreases the demand for domestic goods. An increase in the exchange rate—an appreciation—also decreases the demand for domestic goods.
- The exchange rate is determined by the interest parity condition, which states that domestic and foreign bonds must have the same expected rate of return in terms of domestic currency.
- Given the expected future exchange rate and the foreign interest rate, increases in the domestic interest rate lead to an increase in the exchange rate—an appreciation. Decreases in the domestic interest rate lead to a decrease in the exchange rate—a depreciation.
- Under flexible exchange rates, an expansionary fiscal policy leads to an increase in output. If the fiscal expansion is partially offset by tighter monetary policy, it leads to an increase in the interest rate and an appreciation.
- Under flexible exchange rates, a contractionary monetary policy leads to a decrease in output, an increase in the interest rate, and an appreciation.
- There are many types of exchange rate arrangements: fully flexible exchange rates, crawling pegs, fixed exchange rates (or pegs), or the adoption of a common currency. Under fixed exchange rates, a country maintains a fixed exchange rate in terms of a foreign currency or a basket of currencies.
- Under fixed exchange rates and the interest parity condition, a country must maintain an interest rate equal to the foreign interest rate. The central bank loses the use of monetary policy as a policy instrument. Fiscal policy becomes potentially more powerful than under flexible exchange rates, because fiscal policy requires monetary accommodation, and so does not lead to offsetting changes in the domestic interest rate and exchange rate.

KEY TERMS

Mundell-Fleming model, 397
sudden stops, 400
safe haven, 400
supply siders, 407
twin deficits, 407
peg, 409
crawling peg, 410
European Monetary System (EMS), 410
bands, 410
central parity, 410
euro, 410

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The interest rate parity condition means that interest rates are equal across countries.
b. Other things being equal, the interest parity condition implies that the domestic currency will appreciate in response to an increase in the expected exchange rate.
c. If financial investors expect the dollar to depreciate against the yen over the coming year, one-year interest rates will be higher in the United States than in Japan.
d. If the expected exchange rate appreciates, the current exchange rate immediately appreciates.
e. The central bank influences the value of the exchange rate by changing the domestic interest rate relative to the foreign interest rate.
f. An increase in domestic interest rates, all other factors equal, increases exports.
g. A fiscal expansion, all other factors equal, tends to increase net exports.
h. Fiscal policy has a greater effect on output in an economy with fixed exchange rates than in an economy with flexible exchange rates.
i. Under a fixed exchange rate, the central bank must keep the domestic interest rate equal to the foreign interest rates.
j. One important issue with increasing tariffs on imports into your country as a way of reducing your trade deficit is that other countries may retaliate by increasing their tariffs on your exports (their imports).

2. Consider an open economy with flexible exchange rates. Suppose output is at the natural level, but there is a trade deficit. The goal of policy is to reduce the trade deficit and leave the level of output at its natural level.

What is the appropriate fiscal and monetary policy mix?

3. In this chapter, we showed that a reduction in the interest rate in an economy operating under flexible exchange rates leads to an increase in output and a depreciation of the domestic currency.

a. How does the reduction in interest rates in an economy with flexible exchange rates affect consumption and investment?

b. How does the reduction in interest rates in an economy with flexible exchange rates affect net exports?

4. Flexible exchange rates and foreign macroeconomic events
Consider an open economy with flexible exchange rates. Let UIP stand for the uncovered interest parity condition.

a. In an *IS-LM–UIP* diagram, such as Figure 19-2, show the effect of an increase in foreign output, Y^*, on domestic output (Y) and the exchange rate (E), when the domestic central bank leaves the policy interest rate unchanged. Explain in words.
b. In an *IS-LM–UIP* diagram, show the effect of an increase in the foreign interest rate, i^*, on domestic output (Y) and the exchange rate (E), when the domestic central bank leaves the policy interest rate unchanged. Explain in words.

5. Flexible exchange rates and the responses to changes in foreign macroeconomic policy

Suppose there is an expansionary fiscal policy in the foreign country that increases Y^* and i^* at the same time.

a. In an *IS-LM–UIP* diagram, such as Figure 19-2, show the effect of the increase in foreign output, Y^*, and the increase in the foreign interest rate, i^*, on domestic output (Y) and the exchange rate (E), when the domestic central bank leaves the policy interest rate unchanged. Explain in words.
b. In an *IS-LM–UIP* diagram, show the effect of the increase in foreign output, Y^*, and the increase in the foreign interest rate, i^*, on domestic output (Y) and the exchange rate (E), when the domestic central bank matches the increase in the foreign interest rate with an equal increase in the domestic interest rate. Explain in words.
c. In an *IS-LM–UIP* diagram, show the required domestic monetary policy following the increase in foreign output, Y^*, and the increase in the foreign interest rate, i^*, if the goal of domestic monetary policy is to leave domestic output (Y) unchanged. Explain in words. When might such a policy be necessary?

DIG DEEPER

6. Fixed exchange rates and foreign macroeconomic policy

Consider a fixed exchange rate system, in which a group of countries (called follower countries) peg their currencies to the currency of one country (called the leader country). Because the currency of the leader country is not fixed against the currencies of countries outside the fixed exchange rate system, the leader country can conduct monetary policy as it wishes. For this problem, consider the domestic country to be a follower country and the foreign country to be the leader country.

a. How does an increase in interest rates in the leader country affect the interest rate and output in the follower country?
b. How does the increase in leader country interest rates change the composition of output in the follower country? Assume the follower country does not change fiscal policy.
c. Can the follower country use fiscal policy to offset the effects of the leader country's reduction in interest rates and leave domestic output unchanged? When might such a fiscal policy be desirable?
d. Fiscal policy involves changing government spending or changing taxes. Design a fiscal policy mix that leaves consumption and domestic output unchanged when the leader country increases interest rates. What component of output is changed?

7. The exchange rate as a policy tool

A flexible exchange rate combined with a willingness to change the domestic interest rate can increase the effectiveness of monetary policy in an open economy. Consider an economy that suffers a fall in business confidence (which tends to reduce investment).

a. In an *IS-LM–UIP* diagram, such as Figure 19-2, show the short-run effect of the fall in business confidence on output and the exchange rate when the central bank leaves the interest rate unchanged. How does the composition of output change?
b. Suppose the central bank is willing to cut the interest rate to restore the level of output to its original value. How does this change the composition of output?
c. If the exchange rate was fixed and the central bank could not change the interest rate (remember, it is fixed at the foreign value i^*), what policy options are left for the central bank?
d. Central banks generally favor flexible exchange rates. Explain why.

EXPLORE FURTHER

8. Demand for US assets, the dollar, and the trade deficit

This question explores how an increase in demand for US assets may have slowed the depreciation of the dollar that many economists believe is warranted by the large US trade deficit and the need to stimulate the demand for domestic goods after the crisis. Here, we modify the IS-LM–UIP framework to analyze the effects of an increase in demand for US assets. Write the modified uncovered interest parity condition as

$$(1 + i_t) = (1 + i_t^*)(E_t/E_{t+1}^e) - x$$

where the parameter x represents factors affecting the relative demand for domestic assets. An increase in x means that investors are willing to hold domestic assets at a lower interest rate (given the foreign interest rate and the current and expected exchange rates).

a. Solve the *UIP* condition for the current exchange rate, E_t.
b. Substitute the result from part a in the *IS* curve and construct the *UIP* diagram. As in the text, you may assume that P and P^* are constant and equal to one.
c. Suppose that as a result of a large trade deficit in the domestic economy, financial market participants believe that the domestic currency must depreciate in the future. Therefore, the expected exchange rate, E_{t+1}^e, decreases. Show the effect of the decrease in the expected exchange rate in the *IS-LM–UIP* diagram. What are the effects on the exchange rate and the trade balance? (*Hint*: In analyzing the effect on the trade balance, remember why the *IS* curve shifted in the first place.)
d. Now suppose that the relative demand for domestic assets, x, increases. As a benchmark, suppose that the increase in x is exactly enough to return the *IS* curve to its original position, before the decrease in the expected exchange rate. Show the combined effects of the decrease in E_{t+1}^e and the increase in x in your *IS-LM–UIP* diagram. What

are the ultimate effects on the exchange rate and the trade balance?

e. Based on your analysis, is it possible that an increase in demand for US assets could prevent the dollar from depreciating? Is it possible that an increase in demand for US assets could worsen the US trade balance? Explain your answers.

9. Bond yields and long-run currency movements

a. Go the Web site of *The Economist* (www.economist.com) and find data on 10-year interest rates. Look in the section "Markets & Data" and then the subsection "Economic and Financial Indicators." Look at the 10-year interest rates for the United States, Japan, China, Britain, Canada, Mexico, and the Euro area. For each country (treating the Euro area as a country), calculate the spreads as that country's interest rate minus the US interest rate.

b. From the uncovered interest parity condition, the spreads from part a are the annualized expected appreciation rates of the dollar against other currencies. To calculate the 10-year expected appreciation, you must compound. (So, if x is the spread, the 10-year expected appreciation is $[(1 + x)^{10} - 1]$. Be careful about decimal points.) Is the dollar expected to depreciate or appreciate by much against the currency of any of its six major trading partners?

c. Given your answer to part b, for which country(ies) is a significant appreciation or depreciation of the dollar expected over the next decade? Does your answer seem plausible?

10. The Trump administration tariffs

Use a search engine to find material on the Trump administration tariffs' impact on the economy. What do economists say about them? You will learn something even by just reading the heading and the first paragraph of three to five articles.

APPENDIX: Fixed Exchange Rates, Interest Rates, and Capital Mobility

The assumption of perfect capital mobility is a good approximation of what happens in countries with highly developed financial markets and few capital controls, such as the United States, the United Kingdom, Japan, and countries in the euro area. But this assumption is more questionable in countries that have less developed financial markets or have capital controls in place. In these countries, domestic financial investors may have neither the savvy nor the legal right to buy foreign bonds when domestic interest rates are low. The central bank may thus be able to decrease the interest rate while maintaining a given exchange rate.

To examine these issues, we need to have another look at the balance sheet of the central bank. In Chapter 4, we assumed the only asset held by the central bank was domestic bonds. In an open economy, the central bank actually holds two types of assets: (1) domestic bonds and (2) **foreign exchange reserves**, which we shall think of as foreign currency—although they also take the form of foreign bonds or foreign interest-paying assets. Think of the balance sheet of the central bank as represented in Figure 1.

On the asset side are bonds and foreign exchange reserves, and on the liability side is central bank money. There are now two ways in which the central bank can change central bank money: by purchases or sales of bonds in the bond market or by purchases or sales of foreign currency in the foreign exchange market. (If you did not read Section 4-3 in Chapter 4, replace *central bank money* with *money supply* and you will still get the basic argument.)

Perfect Capital Mobility and Fixed Exchange Rates

Consider first the effects of an open market operation under the joint assumptions of perfect capital mobility and fixed exchange rates (the assumptions we made in the last section of this chapter).

- Assume that the domestic interest rate and the foreign interest rate are initially equal, so $i = i^*$. Now suppose the central bank embarks on an expansionary open market operation, buying bonds in the bond market in amount ΔB, and creating money—increasing central bank money—in exchange. This purchase of bonds leads to a decrease in the domestic interest rate, i. This is, however, only the beginning of the story.
- Now that the domestic interest rate is lower than the foreign interest rate, financial investors prefer to hold foreign bonds. To buy foreign bonds, they must first buy foreign currency. They go to the foreign exchange market and sell domestic currency for foreign currency.
- If the central bank did nothing, the price of domestic currency would fall, and the result would be a depreciation. Under its commitment to a fixed exchange rate, however, the central bank cannot allow the currency to depreciate. So it must intervene in the foreign exchange market and sell foreign currency for domestic currency. As it sells foreign currency and buys domestic money, central bank money decreases.
- How much foreign currency must the central bank sell? It must keep selling until central bank money is back to its pre–open market operation level, so the domestic interest rate is again equal to the foreign interest rate. Only then are financial investors willing to hold domestic bonds.

Assets	Liabilities
Bonds Foreign exchange reserves	Central bank money

Figure 1

Balance Sheet of the Central Bank

How long do all these steps take? Under perfect capital mobility, all this may happen within minutes of the original open market operation. After these steps, the balance sheet of the central bank looks as represented in Figure 2. Bond holdings are up by ΔB, reserves of foreign currency are down by ΔB, and central bank money is unchanged, having gone up by ΔB in the open market operation and down by ΔB as a result of the sale of foreign currency in the foreign exchange market.

Let's summarize. Under fixed exchange rates and perfect capital mobility, the only effect of the open market operation is to change the *composition* of the central bank's balance sheet but not central bank money or the interest rate.

Imperfect Capital Mobility and Fixed Exchange Rates

Let's now move away from the assumption of perfect capital mobility. Suppose it takes some time for financial investors to shift between domestic bonds and foreign bonds.

Now an expansionary open market operation can initially bring the domestic interest rate below the foreign interest

Assets	Liabilities
Bonds: ΔB Reserves: $-\Delta B$	Central bank money: $\Delta B - \Delta B = 0$

Figure 2

Balance Sheet of the Central Bank after an Open Market Operation and Induced Intervention in the Foreign Exchange Market

rate. But over time, investors shift to foreign bonds, leading to an increase in the demand for foreign currency in the foreign exchange market. To avoid a depreciation of the domestic currency, the central bank must again stand ready to sell foreign currency and buy domestic currency. Eventually, the central bank buys enough domestic currency to offset the effects of the initial open market operation. Central bank money is back to its pre–open market operation level and so is the interest rate. The central bank holds more domestic bonds and smaller reserves of foreign currency.

The difference between this case and the case of perfect capital mobility is that, by accepting a loss in foreign exchange reserves, the central bank is now able to decrease interest rates *for some time.* If it takes just a few days for financial investors to adjust, the trade-off can be very unattractive—as many countries that have suffered large losses in reserves without much effect on the interest rate have discovered. But if the central bank can affect the domestic interest rate for a few weeks or months, it may, in some circumstances, be willing to do so.

Now let's deviate further from perfect capital mobility. Suppose, in response to a decrease in the domestic interest rate, financial investors are either unwilling or unable to move much of their portfolio into foreign bonds. For example, there are administrative and legal controls on financial transactions, making it illegal or very expensive for domestic residents to invest outside the country. This is the case for a number of emerging economies, from Latin America to China.

After an expansionary open market operation, the domestic interest rate decreases, making domestic bonds less attractive. Some domestic investors may still be able to move into foreign bonds, selling domestic currency for foreign currency. To maintain the exchange rate, the central bank must buy domestic currency and supply foreign currency. However, the foreign exchange intervention by the central bank may now be small compared to the initial open market operation. And if capital controls truly prevent investors from moving into foreign bonds at all, there may be no need for such a foreign exchange intervention.

Even leaving this extreme case aside, the net effects of the initial open market operation and the following foreign exchange interventions are likely to be *an increase in central bank money; a decrease in the domestic interest rate; an increase in the central bank's bond holdings; and some—limited—loss in reserves of foreign currency.*

With imperfect capital mobility, a country has some freedom to move the domestic interest rate while maintaining its exchange rate. This freedom depends primarily on three factors:

- The degree of development of its financial markets, and the willingness of domestic and foreign investors to shift between domestic and foreign assets.
- The degree of capital controls it is able to impose on both domestic and foreign investors.
- The amount of foreign exchange reserves it holds. The higher the reserves it has, the more it can afford the loss in reserves it is likely to sustain if it decreases the interest rate at a given exchange rate.

With the large movements in capital flows documented in the chapter, all of these issues are hot topics. Many countries are considering a more active use of capital controls than in the past. Many are also accumulating large reserves as a precaution against large capital outflows.

Key Term

foreign exchange reserves, 416

This page is intentionally left blank

Exchange Rate Regimes

20

In July 1944, representatives of 44 countries met in Bretton Woods, New Hampshire, to design a new international monetary and exchange rate system. The system they adopted was based on fixed exchange rates, with all member countries other than the United States fixing the price of their currency in terms of dollars. In 1973, a series of exchange rate crises brought an abrupt end to the system—and an end to what is now called "the Bretton Woods period." Since then, the world has been characterized by many exchange rate arrangements. Many countries operate under flexible exchange rates; some operate under fixed exchange rates; some go back and forth between regimes. Which exchange rate regime to choose is one of the most debated issues in macroeconomics and, as the cartoon suggests, a decision facing every country in the world. This chapter discusses this issue.

"Then it's agreed. Until the dollar firms up, we let the clamshell float."

Section 20-1 looks at the medium run. It shows that, in contrast to the results we derived for the short run in Chapter 19, an economy ends up with the same real exchange rate and output level in the medium run, whether it operates under fixed or flexible exchange rates. This does not make the exchange rate regime irrelevant—the short run matters very much—but it is an important qualification to our previous analysis.

Section 20-2 takes another look at fixed exchange rates, with a focus on exchange rate crises. During a typical exchange rate crisis, a country operating under a fixed exchange rate is forced, often under dramatic conditions, to abandon its parity and to devalue. Such crises were behind the breakdown of the Bretton Woods system. They rocked the European Monetary System in the early 1990s, and they were a major element of the Asian Crisis of the late 1990s. It is important to understand why they happen and what they imply.

Section 20-3 takes another look at flexible exchange rates. It shows that the behavior of exchange rates and the relation of the exchange rate to monetary policy are more complex than we assumed in Chapter 19. Large fluctuations in the exchange rate, and the difficulties in using monetary policy to affect the exchange rate, make a flexible exchange rate regime less attractive than it appeared to be in Chapter 19.

Section 20-4 puts all the analysis together and reviews the case for flexible or fixed rates. It discusses two important developments: the use of a common currency in much of Europe, and the move toward strong forms of fixed exchange rate regimes, from currency boards to dollarization.

If you remember one basic message from this chapter, it should be: There is no perfect exchange rate regime. Both fixed and flexible exchange rates have advantages and shortcomings. ▶▶▶

20-1 THE MEDIUM RUN

When we focused on the short run in Chapter 19, we drew a sharp contrast between the behavior of an economy with a flexible exchange rate and one with a fixed exchange rate:

- Under a flexible exchange rate, a country that needed to achieve a real depreciation (for example, to reduce its trade deficit, or to get out of a recession, or both) could do so by relying on an expansionary monetary policy to achieve both a lower interest rate and a decrease in the exchange rate—a depreciation.
- Under a fixed exchange rate, a country lost both instruments. By definition, its nominal exchange rate was fixed and thus could not be adjusted. Moreover, the fixed exchange rate and the interest parity condition implied that the country could not adjust its interest rate either; the domestic interest rate had to remain equal to the foreign interest rate.

This appeared to make a flexible exchange rate regime definitely more attractive than a fixed exchange rate regime. Why should a country give up two macroeconomic instruments—the exchange rate and the interest rate? As we now shift focus from the short run to the medium run, you will see that this previous conclusion needs to be qualified. Although our conclusions about the short run were valid, in the medium run the difference between the two regimes fades away. More specifically, in the medium run, the economy reaches the same real exchange rate and the same level of output whether it operates under a fixed or flexible exchange rate.

The intuition for this result is easy to give. Recall the definition of the real exchange rate:

$$\varepsilon = \frac{EP}{P^*}$$

The real exchange rate, ε, is equal to the nominal exchange rate, E (the price of domestic currency in terms of foreign currency) times the domestic price level, P, divided by the foreign price level, P^*. There are, therefore, two ways in which the real exchange rate can adjust:

- Through a change in the nominal exchange rate, E: By definition, this can be done only under flexible exchange rates. And if we assume that the domestic price level, P, and the foreign price level, P^*, do not change in the short run, it is the only way to adjust the real exchange rate in the short run.
- Through a change in the domestic price level, P, relative to the foreign price level, P^*. In the medium run, as prices adjust, this option is open even to a country operating under a fixed (nominal) exchange rate. And this is indeed what happens under a fixed exchange rate: The adjustment takes place through the price level rather than through the nominal exchange rate.

There are three ways in which a US car can become cheaper relative to a Japanese car: first, through a decrease in the dollar price of the US car; second, through an increase in the yen price of the Japanese car; third, through a decrease in the nominal exchange rate—a decrease in the value of the dollar in terms of the yen.

The IS Relation under Fixed Exchange Rates

In an open economy with fixed exchange rates, we can write the IS relation as

$$Y = Y\left(\frac{\bar{E}P}{P^*},\ G,\ T,\ i^* - \pi^e,\ Y^*\right) \qquad (20.1)$$
$$(-,\ +, -,\qquad -,\ +)$$

The derivation of equation (20.1) is left to Appendix 1 at the end of this chapter, "Deriving the IS Relation under Fixed Exchange Rates." The intuition behind the equation is straightforward, however. Demand, and in turn, output, depend:

- negatively on the real exchange rate, $\bar{E}P/P^*$. $\bar{E}$ denotes the fixed nominal exchange rate; P and P^* denote the domestic and foreign price levels. A higher real exchange rate implies a lower demand for domestic goods and in turn lower output.
- positively on government spending, G, and negatively on taxes, T.
- negatively on the domestic real interest rate, which itself is equal to the domestic nominal interest rate minus expected inflation. Under the interest parity condition and fixed exchange rates, the domestic nominal interest rate is equal to the foreign nominal interest rate, i^*, so the domestic real interest rate is given by $i^* - \pi^e$.
- positively on foreign output, Y^*, through the effect on exports.

Equilibrium in the Short and the Medium Run

Consider an economy where the real exchange rate is too high. As a result, the trade balance is in deficit and output is below potential.

As we saw in Chapter 19, under a flexible exchange rate regime, the central bank could in principle solve the problem. It could, by decreasing the interest rate, induce a nominal depreciation. Given the domestic and the foreign price levels, which we assumed were fixed in the short run, the nominal depreciation implied a real depreciation, an improvement in the trade balance, and an increase in output.

Under a fixed exchange rate regime, however, the central bank cannot move the domestic interest rate. Thus, in the short run, the trade deficit remains, and the country remains in recession.

In the medium run, however, prices can adjust. We saw in the core that the behavior of prices is well described by the Phillips curve relation (Chapter 9, equation (9.3)):

$$\pi - \pi^e = (\alpha/L)(Y - Y_n)$$

When output is above potential, the inflation rate (i.e., the rate of change of prices) is higher than expected. When output is below potential, as is the case we are considering here, the inflation rate is lower than expected. Following the discussion in Chapter 9, I shall assume here that expected inflation is constant so that the Phillips curve relation is given by:

$$\pi - \bar{\pi} = (\alpha/L)(Y - Y_n) \tag{20.2}$$

We are now nearly ready to think about dynamics and the medium run. We need to make an assumption about the initial domestic and foreign inflation rates. Denote the foreign inflation rate by π^*. Suppose that if output were equal to potential output, domestic and foreign inflation were equal to each other and both equal to $\bar{\pi}$, so $\pi = \pi^* = \bar{\pi}$. That is, if both economies are operating at potential, inflation rates were the same, relative price levels would remain constant, and so would the real exchange rate.

As we are assuming that we start from a situation where output is below potential, equation (20.2) implies that domestic inflation is lower than it would be if output were at potential, and thus lower than foreign inflation. Put another way, the domestic price level increases more slowly than the foreign price level. This implies that, given the nominal exchange rate, which is fixed, the real exchange rate decreases. As a result, net exports increase over time, and so does output. In the medium run, output is back at potential, and domestic inflation is back to $\bar{\pi}$, thus equal to foreign inflation. With domestic and foreign inflation being equal, the real exchange rate is constant.

$\pi < \pi^* \Rightarrow \bar{E}P/P^* \downarrow$

To summarize:

- In the short run, a fixed nominal exchange rate implies a fixed real exchange rate.
- In the medium run, the real exchange rate can adjust even if the nominal exchange rate is fixed. This adjustment is achieved through movements in relative price levels over time.

The Case for and against a Devaluation

The result that, even under fixed exchange rates, the economy can return to potential output in the medium run is important. But it does not eliminate the fact that the process of adjustment can be long and painful. Output may remain too low and unemployment too high for a long time.

Are there faster and better ways to return output to potential? The answer, within the model we have just developed, is a clear yes. Suppose that the government decides, while keeping the fixed exchange rate regime, to allow for a *one-time* devaluation. Given the price level, the devaluation (a decrease in the nominal exchange rate) leads, in the short run, to a real depreciation (a decrease in the real exchange rate) and therefore to an increase in output. In principle, the right size devaluation can thus achieve in the short run what was achieved above only in the medium run, and thus avoid much of the pain. So whenever a country under fixed exchange rates faces either a large trade deficit or a severe recession, there is heavy political pressure either to give up the fixed exchange rate regime altogether or, at least, to have a one-time devaluation. Perhaps

FOCUS

The Return of Britain to the Gold Standard: Keynes versus Churchill

In 1925, Britain decided to return to the gold standard. The **gold standard** was a system in which each country fixed the price of its currency in terms of gold and stood ready to exchange gold for currency at the stated parity. This system implied fixed exchange rates between countries. (If, for example, one unit of currency in country A was worth two units of gold, and one unit of currency in country B was worth one unit of gold, the exchange rate between the two was 2 (or ½, depending on which you take as a domestic country).

The gold standard had been in place from 1870 until World War I. Because of the need to finance the war, and to do so in part by money creation, Britain suspended the gold standard in 1914. In 1925, Winston Churchill, then Britain's Chancellor of the Exchequer (the British equivalent of Secretary of the Treasury in the United States), decided to return to the gold standard and to return to it at the pre-war parity—that is, at the pre-war value of the pound in terms of gold. But because prices had increased faster in Britain than in many of its trading partners, returning to the pre-war parity implied a large real appreciation: At the same nominal exchange rate as before the war, British goods were now relatively more expensive relative to foreign goods. (Go back to the definition of the real exchange rate, $\varepsilon = EP/P^*$: The price level in Britain, P, had increased more than the foreign price level, P^*. At a given nominal exchange rate, E , this implied that ε was higher, that Britain suffered from a real appreciation.)

Keynes severely criticized the decision to return to the pre-war parity. In *The Economic Consequences of Mr. Churchill*, a book he published in 1925, Keynes argued as follows: If Britain were going to return to the gold standard, it should have done so at a lower price of currency in terms of gold; that is, at a lower nominal exchange rate than the pre-war nominal exchange rate. In a newspaper article, he articulated his views as follows:

> *There remains, however, the objection to which I have never ceased to attach importance, against the return to gold in actual present conditions, in view of the possible consequences on the state of trade and employment. I believe that our price level is too high, if it is converted to gold at the par of exchange, in relation to gold prices elsewhere; and if we consider the prices of those articles only which are not the subject of international trade, and of services, i.e. wages, we shall find that these are materially too high—not less than 5 per cent, and probably 10 per cent. Thus, unless the situation is saved by a rise of prices elsewhere, the Chancellor is committing us to a policy of forcing down money wages by perhaps 2 shillings in the Pound.*
>
> *I do not believe that this can be achieved without the gravest danger to industrial profits and industrial peace. I would much rather leave the gold value of our currency where it was some months ago than embark on a struggle with every trade union in the country to reduce money wages. It seems wiser and simpler and saner to leave the currency to find its own level for some time longer rather than force a situation where employers are faced with the alternative of closing down or of lowering wages, cost what the struggle may.*
>
> *For this reason, I remain of the opinion that the Chancellor of the Exchequer has done an ill-judged thing—ill judged because we are running the risk for no adequate reward if all goes well.*

Keynes's prediction turned out to be right. While other countries were growing, Britain remained in recession for the rest of the decade. Most economic historians attribute a good part of the blame to the initial overvaluation.[1]

the most forceful presentation of this view was made 90 years ago by John Maynard Keynes, who argued against Winston Churchill's decision to return the British pound in 1925 to its pre–World War I parity with gold. His arguments are presented in the Focus Box "The Return of Britain to the Gold Standard: Keynes versus Churchill." Most economic historians believe that history proved Keynes right, and that overvaluation of the pound was one of the main reasons for Britain's poor economic performance after World War I.

Those who oppose a shift to flexible exchange rates or who oppose a devaluation argue that there are good reasons to choose fixed exchange rates, and that too much willingness to devalue defeats the purpose of adopting a fixed exchange rate regime in the first place. They argue that too much willingness on the part of governments to consider devaluations actually leads to an increased likelihood of exchange rate crises. To understand their arguments, we now turn to these crises: what triggers them and what their implications might be.

[1] *"An American Study of Shares versus Bonds as Permanent Investment,"* The Nation and the Athenaeum, *Supplement, May 2, 1925, pp. 157-158.*

20-2 EXCHANGE RATE CRISES UNDER FIXED EXCHANGE RATES

Suppose a country has chosen to operate under a fixed exchange rate. Suppose also that financial investors start believing there may soon be an exchange rate adjustment—either a devaluation or a shift to a flexible exchange rate regime accompanied by a depreciation. We just saw why this might be the case:

- The real exchange rate may be too high. Or put another way, the domestic currency may be overvalued, leading to too large a current account deficit. In this case, a real depreciation is called for. Although this could be achieved in the medium run without a devaluation, financial investors conclude that the government will take the quickest way out—and devalue.

  Such an overvaluation often happens in countries that peg their nominal exchange rate to the currency of a country with lower inflation. Higher relative inflation implies a steadily increasing price of domestic goods relative to foreign goods, a steady real appreciation, and so a steady worsening of the trade position. As time passes, the need for an adjustment of the real exchange rate increases, and financial investors become more and more nervous. They start thinking that a devaluation might be coming.
- Internal conditions may call for a decrease in the domestic interest rate. As we have seen, a decrease in the domestic interest rate cannot be achieved under fixed exchange rates. But it can be achieved if the country is willing to shift to a flexible exchange rate regime. If a country lets the exchange rate **float** and then decreases its domestic interest rate, we know from Chapter 19 that this will trigger a decrease in the nominal exchange rate—a nominal depreciation.

To let a currency "float" is to allow a move from a fixed to a flexible exchange rate regime. A floating exchange rate regime is the same as a flexible exchange rate regime.

As soon as financial markets believe a devaluation may be coming, then maintaining the exchange rate requires an increase—often a large one—in the domestic interest rate.

Because it is more convenient, we use the approximation, equation (17.4), rather than the original interest parity condition, equation (17.2).

To see this, return to the interest parity condition we derived in Chapter 17:

$$i_t = i_t^* - \frac{(E_{t+1}^e - E_t)}{E_t} \tag{20.3}$$

In Chapter 17, we interpreted this equation as a relation among the *one-year* domestic and foreign nominal interest rates, the current exchange rate, and the expected exchange rate a year hence. But the choice of one year as the period was arbitrary. The relation holds over a day, a week, a month. If financial markets expect the exchange rate to be 2% lower a month from now, they will hold domestic bonds only if the one-month domestic interest rate exceeds the one-month foreign interest rate by 2% (or, if we express interest rates at an annual rate, if the annual domestic interest rate exceeds the annual foreign interest rate by 2% $\times$ 12 = 24%).

Under fixed exchange rates, the current exchange rate, E_t, is set at some level, say $E_t = \bar{E}$. If markets expect that the parity will be maintained over the period, then $E_{t+1}^e = \bar{E}$, and the interest parity condition simply states that the domestic and foreign interest rates must be equal.

Suppose, however, that participants in financial markets start anticipating a devaluation—a decision by the central bank to give up the parity and decrease the exchange rate in the future. Suppose they believe that, in the coming month, there is a 75% chance the parity will be maintained and a 25% chance there will be a 20% devaluation. The term $(E_{t+1}^e - E_t)/E_t$ in the interest parity equation (20.3), which we assumed to be equal to zero earlier, now equals $0.75 \times 0\% + 0.25 \times (-20\%) = -5\%$ (a 75% chance of no change plus a 25% chance of a devaluation of 20%).

They may require more than that, given that there is clearly a lot of risk involved. Our computation ignores the risk premium.

This implies that, if the central bank wants to maintain the existing parity, it must now set a monthly interest rate 5% higher than before—60% higher at an annual rate (12 months × 5% per month); 60% is the interest differential needed to convince investors to hold domestic bonds rather than foreign bonds! Any smaller interest differential and investors will not want to hold domestic bonds.

What, then, are the choices confronting the government and the central bank?

In most countries, the government is formally in charge of choosing the parity and the central bank is formally in charge of maintaining it. In practice, choosing and maintaining the parity are joint responsibilities of the government and the central bank.

- First, the government and the central bank can try to convince markets that they have no intention of devaluing. This is always the first line of defense: Communiqués are issued, and prime ministers go on TV to reiterate their absolute commitment to the existing parity. But words are cheap, and they rarely convince financial investors.
- Second, the central bank can increase the interest rate, but by less than would be needed to satisfy equation (20.3)—in our example, by less than 60%. Although domestic interest rates are high, they are not high enough to fully compensate for the perceived risk of devaluation. This action typically leads to a large capital outflow because, under the circumstances, financial investors still prefer to get out of domestic bonds and into foreign bonds, which offer higher expected returns in terms of domestic currency. Thus investors sell domestic bonds, getting the proceeds in domestic currency. They then go to the foreign exchange market to sell domestic currency for foreign currency in order to buy foreign bonds. If the central bank did not intervene in the foreign exchange market, the large sales of domestic currency for foreign currency would lead to a depreciation. If the central bank wants to maintain the fixed exchange rate, it must therefore stand ready to buy domestic currency and sell foreign currency at the current exchange rate. In doing so, it often loses most of its reserves of foreign currency. (The mechanics of central bank intervention were described in the appendix to Chapter 19.)
- Eventually—after a few hours or a few weeks—the choice for the central bank becomes either to increase the interest rate enough to satisfy equation (20.3) or to validate the market's expectations and devalue. Setting a very high short-term domestic interest rate can have a devastating effect on demand and output—no firm wants to invest, and no consumer wants to borrow when interest rates are very high. This course of action makes sense only if (1) the perceived probability of a devaluation is small, so the interest rate does not have to be too high; and (2) the government believes markets will soon become convinced that no devaluation is coming, allowing domestic interest rates to decrease. Otherwise, the only option is to devalue. (All these steps were at the center of the exchange rate crisis that affected much of Western Europe in 1992. See the Focus Box "The 1992 EMS Crisis.")

In the summer of 1998, Boris Yeltsin announced that the Russian government had no intention of devaluing the ruble. Two weeks later, the ruble collapsed.

To summarize: Expectations that a devaluation may be coming can trigger an exchange rate crisis. Faced with such expectations, the government has two options:

- give in and devalue, or
- fight and maintain the parity, at the cost of very high interest rates and a potential recession. Fighting may not work anyway; the recession may force the government to change policy later on—or force the government out of office.

An interesting twist here is that a devaluation can occur even if the belief that a devaluation was coming was initially groundless. In other words, even if the government initially has no intention of devaluing, it might be forced to do so if financial markets believe that it will devalue. The cost of maintaining the parity would be a long period of high interest rates and a recession; the government might prefer to devalue instead.

This should remind you of our discussion of bank runs in Chapter 6. The rumor that a bank is in trouble may trigger a run on the bank and force it to close, whether or not there was truth to the rumor.

FOCUS

The 1992 EMS Crisis

An example of the problems we discussed in this section is the exchange rate crisis that shook the European Monetary System (EMS) in the early 1990s.

At the start of the 1990s, the EMS appeared to work well. Established in 1979, the EMS was an exchange rate system based on fixed parities with bands. Each member country (among them, France, Germany, Italy, and, beginning in 1990, the United Kingdom) had to maintain its exchange rate vis-à-vis all other member countries within narrow bands. The first few years had been rocky, with many realignments—adjustment of parities—among member countries. But from 1987 to 1992 there were only two realignments, and there was increasing talk about narrowing the bands further and even moving to the next stage—adoption of a common currency.

In 1992, however, financial markets became increasingly convinced that more realignments were soon to come. The reason was one we saw in Chapter 19: the macroeconomic implications of Germany's reunification. Because of the pressure on demand after reunification, the Bundesbank (the German central bank) was maintaining high interest rates to avoid too large an increase in output and an increase in inflation in Germany. While Germany's EMS partners needed lower interest rates to reduce a growing unemployment problem, they had to match the German interest rates to maintain their EMS parities. To financial markets, the position of Germany's EMS partners looked increasingly untenable. Lower interest rates outside Germany and thus devaluations of many currencies relative to the Deutsche Mark (DM) appeared increasingly likely.

Throughout 1992, the perceived probability of a devaluation forced a number of EMS countries to maintain higher nominal interest rates than even those in Germany. Still, the first major crisis did not come until September 1992.

In early September 1992, the belief that a number of countries were soon going to devalue led to speculative attacks on a number of currencies, with financial investors selling in anticipation of a devaluation. All the lines of defense described earlier were used by the central banks and the governments of the countries under attack. First, solemn communiqués were issued, but with no discernible effect. Then interest rates were increased. For example, Sweden's overnight interest rate (the rate for lending and borrowing overnight) increased to 500% (expressed at an annual rate)! But even such extremely high interest rates were not enough to prevent capital outflows and large losses of foreign exchange reserves by the central banks under pressure.

At that point, different countries took different courses of action. Spain devalued its exchange rate. Italy and the United Kingdom suspended their participation in the EMS. France decided to tough it out through higher interest rates until the storm was over. Figure 1 shows the evolution of the exchange rates relative to the DM for a number of European countries from January 1992 to December 1993. You can clearly see the effects of the September 1992 crisis, highlighted in the figure, and the ensuing depreciations/devaluations.

By the end of September, investors, by and large, believed that no further devaluations were imminent. Some countries were no longer in the EMS. Others had devalued but remained in the EMS, and those that had maintained their parity had shown their determination to stay in the EMS, even if this

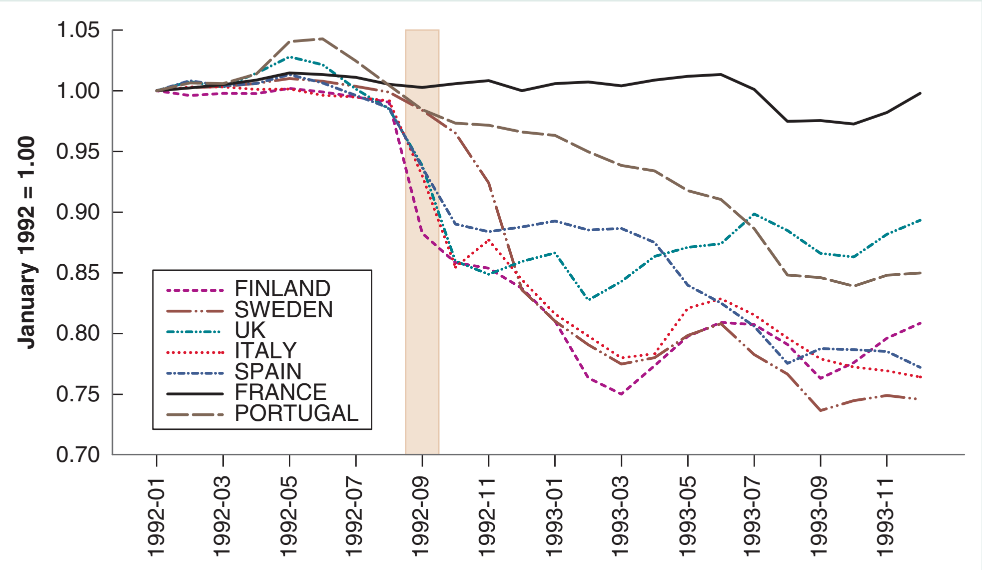


Figure 1

Exchange Rates of Selected European Countries Relative to the Deutsche Mark, January 1992 to December 1993

Source: IMF database.

meant very high interest rates. But the underlying problem—the high German interest rates—was still present, and it was only a matter of time before the next crisis. In November 1992, new speculation forced a devaluation of the Spanish peseta, the Portuguese escudo, and the Swedish krona. The peseta and the escudo were further devalued in May 1993. In July 1993, after yet another large speculative attack, EMS countries decided to adopt large fluctuation bands (plus or minus 15%) around central parities, in effect moving to a system that allowed for large exchange rate fluctuations.

This system with wider bands was kept until the adoption of a common currency, the euro, in January 1999.

To summarize: The 1992 EMS crisis came from the perception by financial markets that the high interest rates forced by Germany on its partners under the rules of the EMS were becoming very costly.

The belief that some countries might want to devalue or get out of the EMS led investors to ask for even higher interest rates, making it even more costly for those countries to maintain their parity.

In the end, some countries could not bear the cost; some devalued, some dropped out. Others remained in the system, but at a substantial cost in terms of output. (For example, average growth in France from 1990 to 1996 was 1.2%, compared to 2.3% for Germany over the same period.)

20-3 EXCHANGE RATE MOVEMENTS UNDER FLEXIBLE EXCHANGE RATES

In the model we developed in Chapter 19, there was a simple relation between the interest rate and the exchange rate: The lower the interest rate, the lower the exchange rate. This implied that a country that wanted to maintain a stable exchange rate just had to maintain its interest rate close to the foreign interest rate. A country that wanted to achieve a given depreciation just had to decrease its interest rate by the right amount.

In reality, the relation between the interest rate and the exchange rate is not so simple. Exchange rates often move even in the absence of movements in interest rates. Furthermore, the size of the effect of a given change in the interest rate on the exchange rate is hard to predict. This makes it much harder for monetary policy to achieve its desired outcome.

To see why things are more complicated, we must return once again to the interest parity condition we derived in Chapter 17, equation (17.2):

$$(1 + i_t) = (1 + i_t^*)\left(\frac{E_t}{E_{t+1}^e}\right)$$

As we did in Chapter 19 (equation (19.4)), multiply both sides by E_{t+1}^e and reorganize to get

$$E_t = \frac{1 + i_t}{1 + i_t^*} E_{t+1}^e \tag{20.4}$$

Think of the time period (from t to $t + 1$) as one year. The exchange rate this year depends on the one-year domestic interest rate, the one-year foreign interest rate, and the exchange rate expected for next year.

We assumed in Chapter 19 that the expected exchange rate next year, E_{t+1}^e, was constant. But this was a simplification. The exchange rate expected one year hence is not constant. Using equation (20.4), but now for next year, it is clear that the exchange rate next year will depend on next year's one-year domestic interest rate, the one-year foreign interest rate, the exchange rate expected for the year after, and so on. So any change in expectations of *current and future* domestic and foreign interest rates, as well as changes in the expected exchange rate in the far future, will affect the exchange rate today.

Let's explore this more closely. Write equation (20.4) for year $t + 1$ rather than for year t:

$$E_{t+1} = \frac{1 + i_{t+1}}{1 + i_{t+1}^*} E_{t+2}^e$$

The exchange rate in year $t + 1$ depends on the domestic and foreign interest rates for year $t + 1$, as well as on the expected future exchange rate in year $t + 2$. So the expectation of the exchange rate in year $t + 1$ held as of year t is given by:

$$E^e_{t+1} = \frac{1 + i^e_{t+1}}{1 + i^{*e}_{t+1}} E^e_{t+2}$$

Replacing E^e_{t+1} in equation (20.4) with the expression above gives:

$$E_t = \frac{(1 + i_t)(1 + i^e_{t+1})}{(1 + i^*_t)(1 + i^{*e}_{t+1})} E^e_{t+2}$$

The current exchange rate depends on this year's domestic and foreign interest rates, on next year's expected domestic and foreign interest rates, and on the expected exchange rate two years from now. Continuing to solve forward in time in the same way (by replacing E^e_{t+2}, E^e_{t+3}, and so on until year $t + n$), we get:

$$E_t = \frac{(1 + i_t)(1 + i^e_{t+1}) \cdots (1 + i^e_{t+n})}{(1 + i^*_t)(1 + i^{*e}_{t+1}) \cdots (1 + i^{*e}_{t+n})} E^e_{t+n+1} \qquad (20.5)$$

Suppose we take n to be large, say 10 years (equation (20.5) holds for any value of n). This relation tells us that the current exchange rate depends on two sets of factors:

- Current and expected domestic and foreign interest rates for each year over the next 10 years.
- The expected exchange rate 10 years from now.

The basic lesson from Appendix 2: For all the statements, you can put "real" in front of exchange rates and interest rates, and the statements will also hold.

For some purposes, it is useful to go further and derive a relation among current and expected future domestic and foreign *real* interest rates, the current *real* exchange rate, and the expected future *real* exchange rate. This is done in Appendix 2 at the end of this chapter. (The derivation is not much fun, but it is a useful way of brushing up on the relation between real interest rates and nominal interest rates, and real exchange rates and nominal exchange rates.) Equation (20.5) is sufficient to make three important points, each outlined in more detail below:

- The level of today's exchange rate will move one-for-one with the future expected exchange rate.
- Today's exchange rate will move when future expected interest rates move in either country.
- Because today's exchange rate moves with any change in expectations, the exchange rate will be volatile, that is, move frequently and perhaps by large amounts.

Exchange Rates and the Current Account

Any factor that moves the expected future exchange rate, E^e_{t+n}, also moves the current exchange rate, E_t. Indeed, if the domestic interest rate and the foreign interest rate are expected to be the same in both countries from t to $t + n$, the fraction on the right in equation (20.5) is equal to 1, so the relation reduces to $E_t = E^e_{t+n}$. In words: The effect of any change in the expected future exchange rate on the current exchange rate is one-for-one.

If we think of n as large (say, 10 years or more), we can think of E^e_{t+n} as the exchange rate required to achieve current account balance in the medium or long run. Countries cannot borrow—run a current account deficit—forever, and will not want to lend—run a current account surplus—forever either. Thus, any news that affects forecasts of the current account balance in the future is likely to have an effect on the expected future exchange rate and in turn on the exchange rate today. For example, the announcement of a larger-than-expected current account deficit may lead investors to conclude that a depreciation will eventually be needed to repay the increased debt. Thus, E^e_{t+n} will decrease, leading in turn to a decrease in E_t today.

News about future current accounts is likely to affect the exchange rate today. What effect would you expect, for example, from the announcement of a major oil discovery?

Exchange Rates and Current and Future Interest Rates

Any factor that moves current or expected future domestic or foreign interest rates between years t and $t + n$ also moves the current exchange rate. For example, given foreign interest rates, an increase in current or expected future domestic interest rates leads to an increase in E_t—an appreciation.

News about current and future domestic and foreign interest rates is likely to affect the exchange rate.

This implies that any variable that causes investors to change their expectations of future interest rates will lead to a change in the exchange rate today. For example, the "dance of the dollar" in the 1980s that we discussed in Chapter 17—the sharp appreciation of the dollar in the first half of the decade, followed by an equally sharp depreciation later—can be largely explained by the movement in current and expected future US interest rates relative to interest rates in the rest of the world during that period. During the first half of the 1980s, tight monetary policy and expansionary fiscal policy combined to increase both short- and long-term interest rates, with the increase in long-term rates reflecting anticipations of high short-term interest rates in the future. This increase in both current and expected future interest rates was, in turn, the main cause of the dollar appreciation. Both fiscal and monetary policy were reversed in the second half of the decade, leading to lower US interest rates and a depreciation of the dollar.

For more on the relation between long-term interest rates and current and expected future short-term interest rates, go back to Chapter 14.

Exchange Rate Volatility

The third implication follows from the first two. In reality, and in contrast to our analysis in Chapter 19, the relation between the interest rate and the exchange rate is anything but mechanical. When the central bank cuts the policy rate, financial markets have to assess whether this action signals a major shift in monetary policy and the cut in the interest rate is just the first of many such cuts, or whether this cut is just a temporary movement in interest rates. Announcements by the central bank may not be useful. The central bank itself may not know what it will do in the future; typically, it will be reacting to early signals, which may be reversed later. Investors also have to assess how foreign central banks will react: whether they will stay put or follow suit and cut their own interest rates. All this makes it much harder to predict the effect of the change in the interest rate be on the exchange rate.

We leave aside here other factors that also move the exchange rate, such as changing perceptions of risk, which we discussed in a Focus Box titled "Capital Flows, Sudden Stops, and the Limits to the Interest Parity Condition" in Chapter 19.

Let's be more concrete. Go back to equation (20.5). Assume that $E^e_{t+n} = 1$. Assume that current and expected future domestic interest rates, and current and expected future foreign interest rates, are all equal to 5%. The current exchange rate is then given by:

$$E_t = \frac{(1.05)^n}{(1.05)^n} 1 = 1$$

Now consider a reduction in the current domestic interest rate, i_t, from 5% to 3%. Will this lead to a decrease in E_t—to a depreciation—and if so by how much? The answer: It depends.

If this reminds you of our discussion in Chapter 14 of how monetary policy affects stock prices, you are right. This is more than a coincidence. Like stock prices, the exchange rate depends very much on expectations of variables far into the future. How expectations change in response to a change in a current variable (here, the interest rate) determines the outcome.

Suppose the interest rate is expected to be lower for just one year, so the $n - 1$ expected future interest rates remain unchanged. The current exchange rate then decreases to:

$$E_t = \frac{(1.03)(1.05)^{n-1}}{(1.05)^n} = \frac{1.03}{1.05} = 0.98$$

The lower interest rate leads to a decrease in the exchange rate—a depreciation—of only 2%.

Suppose instead that, when the current interest rate declines from 5% to 3%, investors expect the decline to last for five years (so $i^e_{t+4} = \cdots = i^e_{t+1} = i_t = 3\%$). The exchange rate then decreases to:

$$E_t = \frac{(1.03)^5(1.05)^{n-5}}{(1.05)^n} = \frac{(1.03)^5}{(1.05)^5} = 0.90$$

The lower interest rate now leads to a decrease in the exchange rate—a depreciation—of 10%, a much larger effect.

You can surely think of still more outcomes. Suppose investors had anticipated that the central bank was going to decrease interest rates, and the actual decrease turns out to be smaller than they anticipated. In this case, the investors will revise their expectations of future nominal interest rates *upward*, leading to an appreciation rather than a depreciation of the currency.

When, at the end of the Bretton Woods period, countries moved from fixed exchange rates to flexible exchange rates, most economists expected that exchange rates, while flexible, would be stable. The large fluctuations in exchange rates that followed—and have continued to this day—came as a surprise. For some time, these fluctuations were thought to be the result of irrational speculation in foreign exchange markets. It was not until the mid-1970s that economists realized that these large movements could be explained, as we have explained here, by the rational reaction of financial markets to news about future interest rates and the future exchange rate. This has an important implication:

A country that decides to operate under flexible exchange rates must accept the fact that it will be exposed to substantial exchange rate fluctuations over time.

20-4 CHOOSING BETWEEN EXCHANGE RATE REGIMES

Let us now return to the question that motivates this chapter. Should countries choose flexible exchange rates or fixed exchange rates? Are there circumstances under which flexible rates dominate, and others under which fixed rates dominate?

Much of what we have seen in this and the previous chapter would seem to favor flexible exchange rates:

- Section 20-1 argued that the exchange rate regime may not matter in the medium run, but it is still the case that it matters in the short run. In the short run, countries that operate under fixed exchange rates and perfect capital mobility give up two macroeconomic instruments: the interest rate and the exchange rate. This not only reduces their ability to respond to shocks but can lead to exchange rate crises.
- Section 20-2 argued that, in a country with fixed exchange rates, the anticipation of a devaluation leads investors to ask for high interest rates. This in turn makes the economic situation worse and puts more pressure on the country to devalue. This is another argument against fixed exchange rates.
- Section 20-3 introduced one argument against flexible exchange rates: they are likely to fluctuate a lot and be difficult to control through monetary policy.

On balance, it appears that, from a macroeconomic viewpoint, flexible exchange rates are preferable to fixed exchange rates. This is indeed the consensus that has emerged among economists and policymakers. The consensus is as follows:

In general, flexible exchange rates are preferable. There are, however, two exceptions. First, when a group of countries is already tightly integrated, a common currency may be the right solution. Second, when the central bank cannot be trusted to follow a responsible monetary policy under flexible exchange rates, a strong form of fixed exchange rates, such as a currency board or dollarization, may be the right solution.

Let us discuss in turn these two exceptions.

Common Currency Areas

Countries that operate under a fixed exchange rate regime are constrained to have the same interest rate. But how costly is that constraint? If the countries face roughly the same macroeconomic problems and the same shocks, they would likely have chosen similar policies in the first place. Forcing them to have the same monetary policy may not be much of a constraint.

This argument was first explored by Robert Mundell, who looked at the conditions under which a set of countries might want to operate under fixed exchange rates or even adopt a common currency. For countries to constitute an **optimal currency area**, he argued, they need to satisfy one of three conditions:

This is the same Mundell who put together the Mundell-Fleming model you saw in Chapter 19.

- The countries must experience similar shocks. We just saw the rationale for this. If they experience similar shocks, they would have chosen roughly the same monetary policy anyway.
- Or, if the countries experience different shocks, prices and wages must be very flexible, so that, if a country needs to regain competitiveness, it can do so by decreasing prices relative to other members, despite the fact that it cannot use the exchange rate.
- Or, if the countries experience different shocks, they must have high factor mobility. For example, if workers are willing to move from countries that are doing poorly to countries that are doing well, factor mobility rather than macroeconomic policy can allow countries to adjust to shocks. When the unemployment rate is high in a country, workers leave that country to take jobs elsewhere and the unemployment rate in that country decreases back to normal. If the unemployment rate is low, workers come to the country and the unemployment rate in the country increases back to normal. The exchange rate is not needed.

Following Mundell's analysis, most economists believe, for example, that the common currency area composed of the 50 states of the United States is close to an optimal currency area. True, the first condition is not satisfied; individual states suffer from different shocks. California is more affected by shifts in demand from Asia than the rest of the United States. Texas is more affected by what happens to the price of oil, and so on. But the third condition is largely satisfied. There is considerable labor mobility across states. When a state does poorly, workers leave that state; when it does well, workers come to that state. State unemployment rates quickly return to normal, not because of state-level macroeconomic policy but because of labor mobility.

Each US state could have its own currency that freely floated against other state currencies. But this is not the way things are. The United States is a common currency area with one currency, the US dollar.

Clearly, there are also many advantages of using a common currency. For firms and consumers in the United States, the benefits of having a common currency are obvious; imagine how complicated life would be if you had to change currency every time you crossed a state line. The benefits go beyond these lower transaction costs. When prices are quoted in the same currency, it becomes much easier for buyers to compare prices and competition between firms increases, benefiting consumers. Given these benefits and the limited macroeconomic costs, it makes good sense for the United States to have a single currency.

In adopting the euro, Europe made the same choice as the United States. When the process of conversion from national currencies to the euro ended in early 2002, the euro became the common currency for 11 European countries. (See the Focus Box "The Euro: A Short History.") The count of countries using the euro at time of writing is now 19. Is the economic argument for this new common currency area as compelling as it is for the United States?

There is little question that a common currency yields for Europe many of the same benefits that it has for the United States. A report by the European Commission estimates that the elimination of foreign exchange transactions in the euro area leads to a reduction in costs of 0.5% of the combined GDP of these countries. There are also clear signs that the use of a common currency increases competition. When shopping for cars, for example, European consumers search for the lowest euro price anywhere in the euro area.

There is, however, much less agreement on whether Europe constitutes an optimal common currency area. This is because none of the three Mundell conditions appears to be satisfied. European countries experienced very different shocks in the past. Recall our discussion of Germany's reunification and how differently it affected Germany and the other European countries in the 1990s. Prices and wages adjust very slowly. And labor mobility is very low in Europe and likely to remain low. Workers move much less *within* European countries than their counterparts do in a given US state. And because of language and cultural differences among European countries, mobility *between* countries is even lower.

The worry that this might lead to long slumps in member countries if they were hit by a country-specific adverse shock was present even before the crisis. The crisis showed that the worry was well justified. Portugal, Greece, and Ireland, which had seen strong demand growth and large increases in current account deficits (see the Focus Box on current account deficits in Chapter 18), suddenly suffered a sharp decrease in spending, a sharp decrease in output, and increasing difficulty in financing their current account deficits. A large depreciation would have helped them increase demand and improve their current account, but with a common currency this could be done only through a decrease in prices relative to their euro partners. The result was a long and painful adjustment process, which, at the time of writing, is far from over.

Figure 20-1 shows the evolution of unemployment and the real multilateral exchange rate for Spain. A boom from 1994 to 2007 led to a steady decrease in unemployment, but

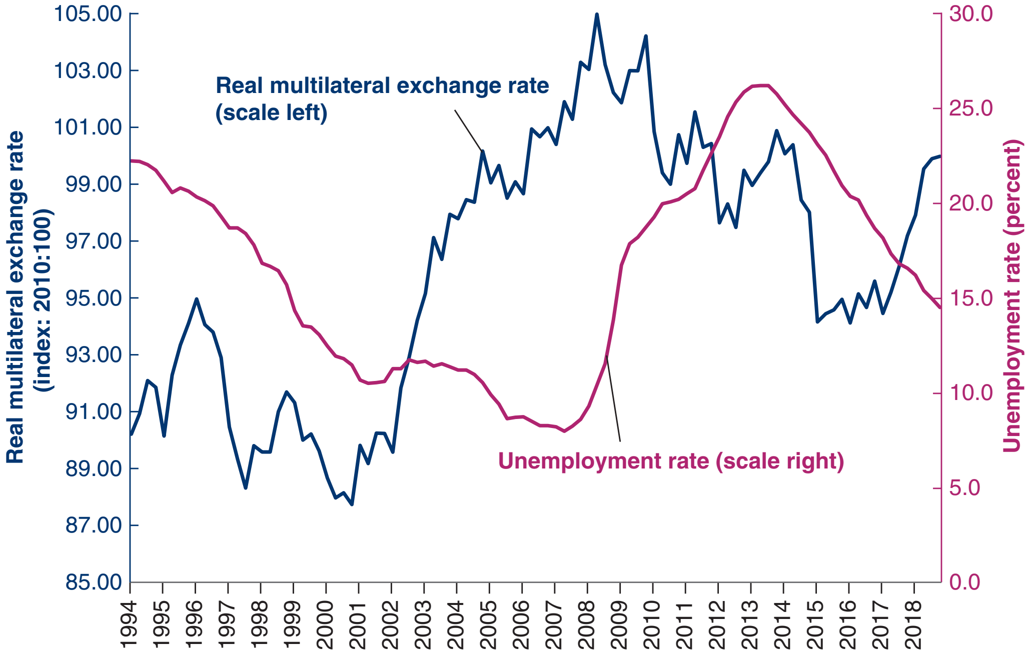


Figure 20-1

Evolution of the Unemployment Rate and the Real Exchange Rate in Spain since 1994

A boom from 1994 to 2007 led to lower unemployment and a real appreciation. In 2008, it turned into a bust, leading to a sharp increase in unemployment and a real depreciation.

Source: FRED: BESBIS, LRHUTTTTESM156S

The Euro: A Short History

FOCUS

- As the European Union (EU) celebrated its 30th anniversary in 1988, a number of governments decided the time had come to plan a move to a common currency. They asked Jacques Delors, president of the EU, to prepare a report, which he presented in June 1989.

  The Delors report suggested moving to a European Monetary Union (EMU) in three stages: Stage I was the abolition of capital controls. Stage II was the choice of fixed parities, to be maintained except for "exceptional circumstances." Stage III was the adoption of a single currency.
- Stage I was implemented in July 1990.
- Stage II began in 1994, after the exchange rate crises of 1992–1993 had subsided (see the Focus Box "The 1992 EMS Crisis"). A minor but symbolic decision involved choosing the name of the new common currency. The French liked *Ecu* (European currency unit), which is also the name of an old French currency. But its partners preferred *euro*, and the name was adopted in 1995.
- In parallel, EU countries held referendums on whether they should adopt the **Maastricht treaty**. The treaty, negotiated in 1991, set three main conditions for joining the EMU: low inflation, a budget deficit below 3%, and a public debt below 60%. The Maastricht treaty was not popular, and in many countries the outcome of the popular vote was close. In France, the treaty passed with only 51% of the votes. In Denmark, the treaty was rejected. The United Kingdom negotiated an "opt-out" clause that allowed Britain not to join the new currency union.
- In the mid-1990s, it looked as if few European countries would satisfy the Maastricht conditions. But several countries took drastic measures to reduce their budget deficit. When the time came to decide, in May 1998, which countries would be members of the euro area, 11 countries made the cut: Austria, Belgium, Finland, France, Germany, Italy, Ireland, Luxembourg, the Netherlands, Portugal, and Spain. The United Kingdom, Denmark, and Sweden decided to stay out. Greece did not qualify initially and didn't join until 2001. (After it joined, it was revealed that it had "cooked the books" and understated the size of its budget deficit in order to qualify.) Since then, Cyprus, Estonia, Malta, Slovakia, and Slovenia have joined.
- Stage III began in January 1999. Parities between the 11 currencies and the euro were "irrevocably" fixed. The new **European Central Bank (ECB)** based in Frankfurt became responsible for monetary policy for the euro area.

From 1999 to 2002, the euro existed as a unit of account, but euro coins and bank notes did not exist. In effect, the euro area was still functioning as an area with fixed exchange rates. The next and final step was the introduction of euro coins and bank notes in January 2002. For the first few months of 2002, national currencies and the euro circulated side by side. Later in the year, national currencies were taken out of circulation.

Today, the euro is the only currency used in the euro area. The numbers of countries adopting the euro reached 19 when Latvia and Lithuania joined.

For more on the euro, go to www.euro.ecb.int/. The Wikipedia page on the euro is also very good.

also a steady real appreciation. In 2008, the boom turned into a bust, with a sharp increase in unemployment until 2013, and a real depreciation. (You might wonder why the unemployment rate continued to decline after real depreciation ended in 2015. This is due in part to an improvement in nonprice competitiveness and in part to stronger domestic demand.)

The challenge for the euro, looking forward, is whether such long slumps can be avoided in the future. Reforms are being explored to eliminate some of the factors that made the slump worse in those countries. A number of reforms are being put in place, from a banking union to a fiscal union, and they should allow countries to better resist adverse shocks. Whether these measures will be enough to be avoid crises in the future remains to be seen.

Hard Pegs, Currency Boards, and Dollarization

The second case for fixed exchange rates is different from the first. It is based on the argument that there may be times when a country may want to limit its ability to use monetary policy.

◀ More on this in Chapter 21.

Look at a country that has had very high inflation in the recent past—perhaps because it was unable to finance its budget deficit by any means other than money creation, resulting in high money growth and high inflation. Suppose the country decides to reduce money growth and inflation. One way to convince financial markets that it is serious about doing this is to fix its exchange rate. The need to use monetary policy to maintain the parity then ties the hands of the monetary authority.

FOCUS

Lessons from Argentina's Currency Board

When Carlos Menem became President of Argentina in 1989, he inherited an economic mess. Inflation was running at more than 30% per month. Output growth was negative.

Menem and his economy minister, Domingo Cavallo, quickly concluded that under these circumstances, the only way to bring money growth—and by implication, inflation—under control was to peg the peso (Argentina's currency) to the US dollar and to do this through a hard peg. So in 1991, Cavallo announced that Argentina would adopt a currency board. The central bank would stand ready to exchange pesos for dollars on demand. Furthermore, it would do so at the highly symbolic rate of one dollar for one peso.

Both the creation of a currency board and the choice of a symbolic exchange rate had the same objectives: to convince investors that the government was serious about the peg, to make it more difficult for future governments to give up the parity and devalue, and, by thus making the fixed exchange rate more credible, to decrease the risk of a foreign exchange crisis.

For a while, the currency board appeared to work extremely well. Inflation, which had exceeded 2,300% in 1990, was down to 4% by 1994! This was clearly the result of the tight constraints the currency board put on money growth. Even more impressive, the collapse in inflation was accompanied by strong output growth: it averaged 5% per year from 1991 to 1999.

Beginning in 1999, however, growth turned negative and Argentina went into a long and deep recession. Was the recession the result of the currency board? Yes and no:

- Throughout the second half of the 1990s, the dollar steadily appreciated relative to the other major world currencies. Because the peso was pegged to the dollar, the peso also appreciated. By the late 1990s, it was clear that the peso was overvalued, leading to a decrease in demand for goods from Argentina, a decline in output, and an increase in the trade deficit.
- Was the currency board fully responsible for the recession? No; there were other causes. But the currency board made it much harder to fight it. Lower interest rates and a depreciation of the peso would have helped the economy recover, but under the currency board, this was not an option.

In 2001, the economic crisis turned into both a financial and an exchange rate crisis, along the lines we described in Section 20-2:

- Because of the recession, Argentina's fiscal deficit had increased, leading to an increase in government debt. Worried that the government might default on its debt, financial investors started asking for very high interest rates on government bonds, making the fiscal deficit even larger and further increasing the risk of default.
- Worried that Argentina would abandon the currency board and devalue to fight the recession, investors started asking for very high interest rates in pesos, making it more costly for the government to sustain the parity with the dollar, and so making it more likely that the currency board would indeed be abandoned.

In December 2001, the government defaulted on part of its debt. In early 2002, it gave up the currency board and let the peso float. The peso sharply depreciated, reaching 3.75 pesos for 1 dollar by June 2002! People and firms that, given their earlier confidence in the peg, had borrowed in dollars found themselves with a large increase in the value of their dollar debts in terms of pesos. Many firms went bankrupt. The banking system collapsed. Despite the sharp real depreciation, which should have helped exports, GDP in Argentina fell by 11% in 2002, and unemployment increased to nearly 20%. Output growth turned positive in 2003, but it took until 2005 for GDP to reach its 1998 level again.

Does this mean that the currency board was a bad idea? Economists still disagree:

- Some economists argue that it was a good idea but that it did not go far enough. They contend that Argentina should have simply dollarized (i.e., adopted the dollar outright as its currency and eliminated the peso altogether). Eliminating the domestic currency would have eliminated the risk of a devaluation. The lesson, they say, is that even a currency board does not provide a sufficiently hard peg for the exchange rate. Only dollarization will do.
- Other (indeed, most) economists argue that the currency board might have been a good idea at the start, but that it should not have been kept in place for so long. Once inflation was under control, Argentina should have moved from a currency board to a floating exchange rate regime. The problem is that Argentina kept the fixed parity with the dollar for too long, to the point that the peso became overvalued and an exchange rate crisis was inevitable.

The debate about "fix versus flex," soft pegs, hard pegs, currency boards, and common currencies is unlikely to be settled any time soon.

For a fascinating, fun, and opinionated book about Argentina's crisis, read Paul Blustein's *And the Money Kept Rolling In (and Out): Wall Street, the IMF, and the Bankrupting of Argentina* (Perseus Books Group, 2005).

To the extent that financial markets expect the parity to be maintained, they will stop worrying about money growth being used to finance the budget deficit.

Note the qualifier: "To the extent that financial markets expect the parity to be maintained." Fixing the exchange rate is not a magic solution. The country also needs to convince financial investors that not only is the exchange rate fixed today, it will remain fixed in the future. There are two ways in which it can do so:

- Make the fixed exchange rate part of a more general macroeconomic package. Fixing the exchange rate while continuing to run a large budget deficit will only convince financial markets that money growth will start again and that a devaluation is soon to come.
- Make it symbolically or technically harder to change the parity, an approach known as a **hard peg**.

An extreme form of a hard peg is simply to replace the domestic currency with a foreign currency. Because the foreign currency chosen is typically the US dollar, this is known as **dollarization**. Few countries are willing, however, to give up their currency and adopt the currency of another country. A less extreme way is to use a **currency board**. Under a currency board, a central bank stands ready to exchange foreign currency for domestic currency at the official exchange rate set by the government. Furthermore—and this is the difference with a standard fixed exchange rate regime—the central bank cannot engage in open market operations (that is, buy or sell government bonds).

When Israel was suffering from high inflation in the 1980s, an Israeli finance minister proposed such a measure as part of a stabilization program. His proposal was perceived as an attack on the sovereignty of Israel and he was fired.

Perhaps the best-known example of a currency board is that adopted by Argentina in 1991 but abandoned in a crisis at the end of 2001. The story is told in the Focus Box "Lessons from Argentina's Currency Board." Economists differ on what conclusions one should draw from what happened in Argentina. Some conclude that currency boards are not *hard* enough: They do not prevent exchange rate crises. So if a country decides to adopt a fixed exchange rate, it should go all the way and dollarize. Others conclude that adopting a fixed exchange rate is a bad idea. If currency boards are used at all, they should be used only for a short period of time, until the central bank has reestablished its credibility and the country returns to a floating exchange rate regime.

SUMMARY

- Even under a fixed exchange rate regime, countries can adjust their *real* exchange rate in the medium run. They can do so by relying on adjustments in the price level. Nevertheless, the adjustment can be long and painful. Exchange rate adjustments can allow the economy to adjust faster and thus reduce the pain that comes from a long adjustment.
- Under fixed exchange rates, exchange rate crises typically start when participants in financial markets believe the currency may soon be devalued. Defending the parity then requires high interest rates, with potentially large adverse macroeconomic effects. These adverse effects may force the country to devalue, even if there were no initial plans for such a devaluation.
- The exchange rate today depends on both (1) the difference between current and expected future domestic interest rates, and current and expected future foreign interest rates; and (2) the expected future exchange rate.

  Any factor that increases current or expected future domestic interest rates leads to an increase in the exchange rate today.

  Any factor that increases current or expected future foreign interest rates leads to a decrease in the exchange rate today. Any factor that increases the expected future exchange rate leads to an increase in the exchange rate today.
- There is wide agreement among economists that flexible exchange regimes generally dominate fixed exchange rate regimes, except in two cases:

  When a group of countries is highly integrated and forms an optimal currency area. (You can think of a common currency for a group of countries as an extreme form of fixed exchange rates among those countries.) For countries to form an optimal currency area, they must either face largely similar shocks or, if they experience different shocks, there must be either high wage and price flexibility or high labor mobility among them.

  When a central bank cannot be trusted to follow a responsible monetary policy under flexible exchange rates. In this case, a strong form of fixed exchange rates, such as dollarization or a currency board, provides a way of tying the hands of the central bank.

KEY TERMS

gold standard, 423
float, 424
optimal currency area, 431
Maastricht treaty, 433
European Central Bank (ECB), 433
hard peg, 435
dollarization, 435
currency board, 435

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. If the nominal exchange rate is fixed, the real exchange rate is fixed.
b. When domestic inflation equals foreign inflation, the real exchange rate is fixed.
c. A devaluation is an increase in the nominal exchange rate.
d. Britain's return to the gold standard caused years of high unemployment.
e. A sudden fear that a country is going to devalue leads to an increase in the domestic interest rate.
f. A change in the expected future exchange rate changes the current exchange rate.
g. The effect of a reduction in domestic interest rates on the exchange rate depends on the length of time domestic interest rates are expected to be below foreign interest rates.
h. Because economies tend to return to their natural level of output in the medium run, it makes no difference whether a country chooses a fixed or flexible exchange rate.
i. High labor mobility within Europe makes the Euro area a good candidate for a common currency.
j. A currency board is the best way to operate a fixed exchange rate.

2. Consider a country operating under fixed exchange rates. The IS curve is given by equation (20.1)

$$Y = Y\left(\frac{\bar{E}P}{P^*}, G, T, i^* - \pi^e, Y^*\right)$$
$$(-,\ +, -,\ \ -,\ +)$$

a. Explain the term $(i^* - \pi^e)$. Why does the foreign nominal interest rate appear in the relation?
b. Explain why when π^e increases, the *IS* curve shifts left.
c. In the following table, how is the real exchange rate evolving from period 1 to period 5? What is domestic inflation? What is foreign inflation? Draw an *IS-LM* diagram with the *IS* curve in period 1 and the *IS* curve in period 5.

Period	P	P*	E	π	π^*	Real exchange rate ε
1	100.0	100.0	0.5			
2	103.0	102.0	0.5			
3	106.1	104.0	0.5			
4	109.3	106.1	0.5			
5	112.6	108.2	0.5			

d. In the following table, how is the real exchange rate evolving from period 1 to period 5? What is domestic inflation? What is foreign inflation? Draw an *IS-LM* diagram with the *IS* curve in period 1 and the *IS* curve in period 5.

Period	P	P*	E	π	π^*	Real exchange rate ε
1	100.0	100.0	0.5			
2	102.0	103.0	0.5			
3	104.0	106.1	0.5			
4	106.1	109.3	0.5			
5	108.2	112.6	0.5			

e. In the table that follows, how is the real exchange rate evolving from period 1 to period 4? What is domestic inflation? What is foreign inflation? What happened between Period 4 and Period 5? Draw an *IS-LM* diagram with the IS curve in period 1 and the IS curve in period 5.

Period	P	P*	E	π	π^*	Real exchange rate ε
1	100.0	100.0	0.5			
2	103.0	102.0	0.5			
3	106.1	104.0	0.5			
4	109.3	106.1	0.5			
5	112.6	108.2	0.5			

3. Policy choices when the real exchange rate is "too high" and the nominal exchange rate is fixed

An overvalued real exchange rate is a rate such that domestic goods are too expensive relative to foreign goods, net exports are too small, and by implication the demand for domestic goods is too low. This leads to difficult policy choices for the government and central bank. The equations (denotes a value for the foreign country) that describe the economy are:*

The IS curve:

$$Y = Y\left(\frac{\bar{E}P}{P^*}, G, T, i^* - \pi^e, Y^*\right)$$
$$(-,\ +, -,\ \ -,\ +)$$

The Phillips curves for the domestic and the foreign economy:

Domestic Phillips curve $\pi - \bar{\pi} = (\alpha/L)(Y - Y_n)$
Foreign Phillips curve $\pi^* - \bar{\pi}^* = (\alpha^*/L^*)(Y^* - Y_n^*)$

In the text and in this question, we make two critical assumptions. These are explored in parts a and b. Then we move to the analysis *of the policy options when a country is experiencing an overvalued exchange rate.*

a. We assume that the foreign economy is always in medium-run equilibrium. What are the implications of that assumption for foreign output and foreign inflation?

b. We assume that the domestic and foreign economies share the same anchored value for the level of expected inflation denoted so that $\bar{\pi} = \bar{\pi}^*$. What is the implication of that assumption once both the domestic and foreign economies are both in medium-run equilibrium?

c. Draw the *IS-LM-UIP* diagram along the lines of Figure 19-2 for the case where the domestic country has an overvalued nominal exchange rate. What is the key feature of that diagram? Under fixed exchange rates without a devaluation, how does the economy return to its medium-run equilibrium?

d. Draw the *IS-LM-UIP* diagram for the case where the domestic country has an overvalued nominal exchange rate. Show how the economy can return to medium-run equilibrium when a devaluation is a policy choice.

e. Recall that the assumption has been made that interest rate parity holds, so $i = i^*$ at all times. Compare the returns on the domestic bond and the returns on the foreign bond in the period of the devaluation. Will bond holders continue to believe there is a completely fixed nominal exchange rate? If bond holders believe that another devaluation is possible, what are the consequences for domestic interest rates?

4. Modeling an exchange rate crisis

An exchange rate crisis occurs when the peg (the fixed exchange rate) loses its credibility. Bond holders no longer believe that next period's exchange rate will be this period's exchange rate.

Period	i_t	i^*_t	E_t	E^e_{t+1}
1		3	0.5	0.5
2		3	0.5	0.45
3		3	0.5	0.45
4		3	0.5	0.5
5	15%	3	0.5	0.4
6		3	0.4	0.4

a. Solve the uncovered interest rate parity condition for the value of the domestic interest rate in period 1. (Use the approximate version of the uncovered interest parity relation in your calculations for all parts of this question below–equation 17.4).

b. In period 2, the crisis begins. Solve the uncovered interest rate parity condition for the value of the domestic interest rate in period 2.

c. The crisis continues in period 3. However, in period 4, the central bank and government resolve the crisis. How does this occur?

d. Unfortunately, in period 5, the crisis returns bigger and deeper than ever. Has the central bank raised interest rates enough to maintain uncovered interest rate parity? What are the consequences for the level of foreign exchange reserves?

e. How is the crisis resolved in period 6? Does this have implications for the future credibility of the central bank and the government?

5. Modeling the movements in the exchange rate

Equation (20.5) provides insight into the movements of nominal exchange rates between a domestic and a foreign country. Remember that the time periods in the equation can refer to any time unit. The equation is:

$$E_t = \frac{(1 + i_t)(1 + i^e_{t+1})\cdots(1 + i^e_{t+n})}{(1 + i^*_t)(1 + i^{*e}_{t+1})\cdots(1 + i^{*e}_{t+n})} E^e_{t+n+1}$$

a. Suppose we are thinking of one-day time periods. There are overnight (1-day) interest rates. How do we interpret a large movement in the exchange rate over the course of the day if we do not observe any change in the one-day interest rate?

b. We learned in Chapter 15 that a one-month (30- or 31-day interest rate) is the average of today's one-day rate and the expected one-day rates over the next 30 days. This will be true in both countries. The following headline is observed on February 1: "ECB predicted to cut interest rates February 14, dollar rises." Does the headline make sense?

c. We learned in Chapter 15 that a two-year bond yield is the average of today's one-year interest rate and the expected one-year rate one year from now. This will be true in both countries. The following headline is observed on February 1: "Fed announces that interest rates will remain low for the foreseeable future, dollar falls." Does the headline make sense?

d. The current account is this period's lending to (if positive) or borrowing from (if negative) the rest of the world. Assume the current account is more negative than expected and this is surprising news. Explain why the exchange rate would depreciate on this surprising news.

DIG DEEPER

6. Realignments of exchange rate

Look at Figure 1 in the box "The 1992 EMS Crisis." European nominal exchange rates had been fixed between the major currencies from roughly 1979 to 1992.

a. Explain how to read the vertical axis of Figure 1. What country experienced the largest depreciation? What country clearly experienced the smallest depreciation?

b. If two-year nominal interest rates in France and Italy had been similar in January 1992, which country would have generated the highest return on a two-year bond?

c. If the changes in the nominal exchange rates returned countries to medium-run equilibrium, which countries had the largest overvaluations in 1992?

7. Real and nominal exchange rates for Canada and Mexico

Two of the largest trading partners of the United States are Canada and Mexico. The FRED database at the Federal Reserve Bank of St. Louis maintains four series that are useful to us: A Real Broad Effective Exchange rate for Mexico (RBMXBIS); A Real Broad Effective Exchange rate for Canada (RBCABIS); the nominal exchange rate of Mexican pesos per U.S. dollar (EXMSUS); and the number of Canadian dollars per US dollar (EXCAUS). Download all the series monthly and organize to a spreadsheet where the start period is January 1994.

a. The exchange rate in FRED is defined as the number of Mexican pesos and the number of Canadian dollars per US dollar. Redefine them as the number of US cents per peso and the number of US cents per Canadian dollar. Why did you do that?
b. Make a time series graph of the redefined Mexican–US nominal exchange rate and the broad real exchange rate index, RBMXBIS, from 1994 to the end of your data. Do you see a period where the nominal exchange rate is pegged? When the peg was released, did the peso appreciate or depreciate? Is there a period where the peso is appreciating in nominal terms and depreciating in real terms? Around what year did the real exchange rate index take its highest value? What is the behavior of the peso in real terms from its peak to 2018? How did that behavior impact the Mexican economy?
c. Make a time series graph of the redefined Canadian-US nominal exchange rate and the broad real exchange rate index, RBCABIS. Estimate the percentage fluctuation in the Canadian-US real exchange rate index from 1994 to the end of your data. Is there a period where the Canadian dollar was pegged? Explain why the real exchange rate index tracks the nominal exchange rate closely in the Canadian–US case. Would there have been any benefits to pegging the Canadian dollar to the US dollar over this period?

EXPLORE FURTHER

8. Exchange rates and expectations

In this chapter, we emphasized that expectations have an important effect on the exchange rate. In this problem, we use data to get a sense of how large a role expectations play. Using the results in Appendix 2 at the end of this chapter, you can show that the uncovered interest parity condition, equation (20.4), can be rewritten as

$$\frac{(E_t - E_{t-1})}{E_{t-1}} \approx (i_t - i_t^*) - (i_{t-1} - i_{t-1}^*) + \frac{(E_t^e - E_{t-1}^e)}{E_{t-1}^e}$$

In words, the percentage change in the exchange rate (the appreciation of the domestic currency) is approximately equal to the change in the interest rate differential (between domestic and foreign interest rates) plus the percentage change in exchange rate expectations (the appreciation of the expected domestic currency value). We shall call the interest rate differential the spread.

a. Go to the Web site of the Bank of Canada (www.bank-banque-canada.ca) and obtain data on the monthly one-year Treasury bill rate in Canada for the past 10 years. Download the data into a spreadsheet. Now go to the Web site of the Federal Reserve Bank of St. Louis (research.stlouisfed.org/fred2) and download data on the monthly US one-year Treasury bill rate for the same time period. (You may need to look under "Constant Maturity" Treasury securities rather than "Treasury Bills.") For each month, subtract the Canadian interest rate from the US interest rate to calculate the spread. Then, for each month, calculate the change in the spread from the preceding month. (Make sure to convert the interest rate data into the proper decimal form.)
b. At the Web site of the St. Louis Fed, obtain data on the monthly exchange rate between the US dollar and the Canadian dollar for the same period as your data from part a. Again, download the data into a spreadsheet. Calculate the percentage appreciation of the US dollar for each month. Using the standard deviation function in your software, calculate the standard deviation of the monthly appreciation of the US dollar. The standard deviation is a measure of the variability of a data series.
c. For each month, subtract the change in the spread (part a) from the percentage appreciation of the dollar (part b). Call this difference the *change in expectations*. Calculate the standard deviation of the change in expectations. How does it compare to the standard deviation of the monthly appreciation of the dollar?

This exercise is too simple. Still, the gist of this analysis survives in more sophisticated work. In the short run, movements in short-term interest rates do not account for much of the change in the exchange rate. Most of the changes in the exchange rate must be attributed to changing expectations.

9. Real and nominal exchange rates between China and the United States

a. Download the series EXCHUS from the FRED database. This expresses the exchange rate between the yuan (one of the names for the Chinese currency) as the number of yuan per US dollar. Convert that exchange rate to the number of US cents per yuan so that we are treating China as the domestic country. Are there periods after 1994 where China maintains a fixed exchange rate?
b. Download the series RBCNBIS. This is a real exchange rate index for China that can be interpreted as the value of $\varepsilon = EP/P^*$ for China. Is the value of the real exchange rate fixed during the period where the nominal exchange rate is fixed? Over that period, how did the value of the real exchange rate affect aggregate demand in China?
c. From July 2007 to June 2010, the value of the yuan was very close to 14.6 US cents. This was a second period of fixed exchange rates. Can you find any evidence from the movements of the nominal and real exchange rate after June 2010 that suggest that China chose a lower value of the yuan to encourage Chinese exports and reduce Chinese imports?

FURTHER READINGS

- For an early skeptical view of the euro, read Martin Feldstein, "The European Central Bank and the Euro: The First Year," 2000, www.nber.org/papers/w7517, and "The Euro and the Stability Pact," 2005, www.nber.org/papers/w11249.

Two good books on the euro crisis (the second one is very critical of the euro in general):

- Jean Pisani-Ferry, *The Euro Crisis and Its Aftermath*, Oxford University Press, 2014.
- Ashoka Mody, *Euro-Tragedy: A Drama in Nine Acts*, Oxford University Press, 2018.

APPENDIX 1: Deriving the IS Relation under Fixed Exchange Rates

Start from the condition for goods market equilibrium we derived in Chapter 19, equation (19.1):

$$Y = C(Y - T) + I(Y, r) + G + NX(Y, Y^*, \varepsilon)$$

This condition states that, for the goods market to be in equilibrium, output must be equal to the demand for domestic goods—that is, the sum of consumption, investment, government spending, and net exports. Next, recall the following relations:

- The real interest rate, r, is equal to the nominal interest rate, i, minus expected inflation, π^e (see Chapter 14):

$$r \equiv i - \pi^e$$

- The real exchange rate, ε, is defined as (see Chapter 17):

$$\varepsilon = \frac{EP}{P^*}$$

- Under fixed exchange rates, the nominal exchange rate, E, is, by definition, fixed. Denote by $\bar{E}$ the value at which the nominal exchange rate is fixed, so:

$$E = \bar{E}$$

- Under fixed exchange rates and perfect capital mobility, the domestic interest rate, i, must be equal to the foreign interest rate, i^* (see Chapter 17):

$$i = i^*$$

Using these four relations, rewrite equation (20.1) as:

$$Y = C(Y - T) + I(Y, i^* - \pi^e) + G + NX\left(Y, Y^*, \frac{\bar{E}P}{P^*}\right)$$

This can be rewritten, using a more compact notation, as:

$$Y = Y\left(\frac{\bar{E}P}{P^*}, G, T, i^* - \pi^e, Y^*\right)$$
$$(-, +, -, \quad -, \quad +)$$

which is equation (20.1) in the text.

APPENDIX 2: The Real Exchange Rate and Domestic and Foreign Real Interest Rates

We derived in Section 20-3 a relation among the current nominal exchange rate, current and expected future domestic and foreign nominal interest rates, and the expected future nominal exchange rate (equation (20.5)). This appendix derives a similar relation, but in terms of real interest rates and the real exchange rate. It then briefly discusses how this alternative relation can be used to think about movements in the real exchange rate.

Deriving the Real Interest Parity Condition

Start from the nominal interest parity condition, equation (19.3):

$$(1 + i_t) = (1 + i_t^*)\left(\frac{E_t}{E_{t+1}^e}\right)$$

Recall the definition of the real interest rate from Chapter 6, equation (6.3):

$$(1 + r_t) = \frac{(1 + i_t)}{(1 + \pi_{t+1}^e)}$$

where $\pi_{t+1}^e \equiv (P_{t+1}^e - P_t)/P_t$ is the expected rate of inflation. Similarly, the foreign real interest rate is given by:

$$(1 + r_t^*) = \frac{(1 + i_t^*)}{(1 + \pi_{t+1}^{*e})}$$

where $\pi_{t+1}^{*e} \equiv (P_{t+1}^{*e} - P_t^*)/P_t^*$ is the expected foreign rate of inflation.

Use these two relations to eliminate nominal interest rates in the interest parity condition, so:

$$(1 + r_t) = (1 + r_t^*)\left[\frac{E_t}{E_{t+1}^e}\frac{(1 + \pi_{t+1}^{*e})}{(1 + \pi_{t+1}^e)}\right] \qquad (20.A1)$$

Note from the definition of inflation that $(1 + \pi_{t+1}^e) = P_{t+1}^e/P_t$ and, similarly, $(1 + \pi_{t+1}^{*e}) = P_{t+1}^{*e}/P_t^*$.

Using these two relations in the term in brackets gives:

$$\frac{E_t}{E_{t+1}^e}\frac{(1 + \pi_{t+1}^{*e})}{(1 + \pi_{t+1}^e)} = \frac{E_t}{E_{t+1}^e}\frac{P_{t+1}^{*e}P_t}{P_t^* P_{t+1}^e}$$

Reorganizing terms:

$$\frac{E_t P_{t+1}^{*e} P_t}{E_{t+1}^e P_t^* P_{t+1}^e} = \frac{E_t P_t/P_t^*}{E_{t+1}^e P_{t+1}^e/P_{t+1}^{*e}}$$

Using the definition of the real exchange rate:

$$\frac{E_t P/P_t^*}{E_{t+1}^e P_{t+1}^e/P_{t+1}^{*e}} = \frac{\varepsilon_t}{\varepsilon_{t+1}^e}$$

Replacing in equation (20.A1) gives:

$$(1 + r_t) = (1 + r_t^*)\frac{\varepsilon_t}{\varepsilon_{t+1}^e}$$

or equivalently,

$$\varepsilon_t = \frac{1 + r_t}{1 + r_t^*}\varepsilon_{t+1}^e \qquad (20.A2)$$

The real exchange rate today depends on the domestic and foreign real interest rates this year and the expected future real exchange rate next year. This equation corresponds to equation (20.4) in the text, but now in terms of the real rather than nominal exchange and interest rates.

Solving the Real Interest Parity Condition Forward

The next step is to solve equation (20.A2) forward, in the same way we did for equation (20.4). The equation above implies that the real exchange rate in year $t + 1$ is given by:

$$\varepsilon_{t+1} = \frac{1 + r_{t+1}^e}{1 + r_{t+1}^{*e}}\varepsilon_{t+2}^e$$

Taking expectations as of year *t*:

$$\varepsilon_{t+1}^e = \frac{1 + r_{t+1}^e}{1 + r_{t+1}^{*e}}\varepsilon_{t+2}^e$$

Replacing in the previous relation:

$$\varepsilon_t = \frac{(1 + r_t)\ (1 + r_{t+1}^e)}{(1 + r_t^*)\ (1 + r_{t+1}^{*e})}\varepsilon_{t+2}^e$$

Solving for ε_{t+2}^e and so on gives:

$$\varepsilon_t = \frac{(1 + r_t)(1 + r_{t+1}^e)\cdots(1 + r_{t+n}^e)}{(1 + r_t^*)\ (1 + r_{t+1}^{*e})(1 + r_{t+n}^{*e})}\varepsilon_{t+n+1}^e$$

This relation gives the current real exchange rate as a function of current and expected future domestic real interest rates, of current and expected future foreign real interest rates, and the expected real exchange rate in year $t + n$.

The advantage of this relation over the relation we derived in the text between the nominal exchange rate and nominal interest rates (equation (20.5)) is that it is typically easier to predict the future real exchange rate than to predict the future nominal exchange rate. If, for example, the economy suffers from a large trade deficit, we can be fairly confident that there will have to be a real depreciation—that ε_{t+n}^e will have to be lower. Whether there will be a nominal depreciation—what happens to E_{t+n}^e—is harder to tell. It depends on what happens to inflation, both at home and abroad, over the next *n* years.

EXTENSIONS

Back to Policy

Nearly every chapter of this text has looked at the role of policy. The next three chapters put it all together.

Chapter **21**

Chapter 21 asks two questions: Given the uncertainty about the effects of macroeconomic policies, wouldn't it be better not to use policy at all? And even if policy can in principle be useful, can we trust policymakers to carry out the right policy? The bottom lines: Uncertainty limits the role of policy. Policymakers do not always do the right thing. But with the right institutions, policy does help and should be used.

Chapter **22**

Chapter 22 looks at fiscal policy. It reviews what we have learned, chapter by chapter, and then looks more closely at the implications of the government budget constraint for the relation between debt, spending, and taxes. It then considers the implications of high levels of public debt, a central issue in advanced countries today.

Chapter **23**

Chapter 23 looks at monetary policy. It reviews what we have learned, chapter by chapter, and then focuses on current challenges. First, it describes the framework, known as inflation targeting, that most central banks had adopted before the crisis. It then turns to several issues raised by the crisis, from the optimal rate of inflation to the role of financial regulation and the use of new instruments, known as macroprudential tools.

This page is intentionally left blank

21

Should Policymakers Be Restrained?

At many points in this text, we saw that the right mix of fiscal and monetary policies could help a country out of a recession, improve its trade position without increasing activity and igniting inflation, slow an overheating economy, and stimulate investment and capital accumulation.

These observations, however, appear to be at odds with frequent demands that policymakers be tightly restrained.

In the United States, there are regular calls for the introduction of a balanced budget amendment to the Constitution to limit the growth of debt. Such a call was the first item in the Contract with America drawn up by Republicans for the midterm US elections in 1994 and reproduced in Figure 21-1. In Europe, the countries that adopted the euro signed a **Stability and Growth Pact (SGP)**, requiring them to keep their budget deficit under 3% of GDP or face large fines. The charter of the central bank of New Zealand, written in 1989, defined monetary policy's role as the maintenance of price stability to the exclusion of any other macroeconomic goal.

Interestingly, the promise of a balanced budget in the United States did not go anywhere. In the euro zone, countries have regularly exceeded the SGP targets. And, in 2018, the central bank of New Zealand changed its mandate to include a goal of high employment. There is a lesson to be drawn here.

This chapter looks at the case for such restraints on macroeconomic policy.

Sections 21-1 and 21-2 look at one line of argument: that policymakers may have good intentions, but they end up doing more harm than good.

Section 21-3 looks at another—more cynical—line, that policymakers do what is best for themselves, which is not necessarily what is best for the country.

If you remember one basic message from this chapter, it should be: It is useful to think of policy as the outcome of games between policymakers and the economy, between policymakers and voters, and among policymakers. As a result, there is a case for putting restraints on policymakers. ▶▶▶

House Republican

Contract with America

A Program for Accountability

We've listened to your concerns and we hear you loud and clear. If you give us the majority, on the first day of Congress, a Republican House will:

Force Congress to live under the same laws as every other American
Cut one out of three Congressional committee staffers
Cut the Congressional budget

Then, in the first 100 days there will be votes on the following 10 bills:

1. Balanced budget amendment and the line item veto: It's time to force the government to live within its means and restore accountability to the budget in Washington.

2. Stop violent criminals: Let's get tough with an effective, able, and timely death penalty for violent offenders. Let's also reduce crime by building more prisons, making sentences longer and putting more police on the streets.

3. Welfare reform: The government should encourage people to work, not have children out of wedlock.

4. Protect our kids: We must strengthen families by giving parents greater control over education, enforcing child support payments, and getting tough on child pornography.

5. Tax cuts for families: Let's make it easier to achieve the American Dream: save money, buy a home, and send their kids to college.

6. Strong national defense: We need to ensure a strong national defense by restoring the essentials of our national security funding.

7. Raise the senior citizens' earning limit: We can put an end to government age discrimination that discourages seniors from working if they want.

8. Roll back government regulations: Let's slash regulations that strangle small business and let's make it easier for people to invest in order to create jobs and increase wages.

9. Common-sense legal reform: We can finally stop excessive legal claims, frivolous lawsuits, and overzealous lawyers.

10. Congressional term limits: Let's replace career politicians with citizen legislators. After all, politics shouldn't be a lifetime job.
(Please see reverse side to know if the candidate from your district has signed the Contract as of October 5, 1994.)

IF WE BREAK THIS CONTRACT, THROW US OUT, WE MEAN IT.

Figure 21-1

The Contract with America

21-1 UNCERTAINTY AND POLICY

A blunt way of stating the first argument in favor of policy restraints is that those who know little should do little. The argument has two parts: Macroeconomists, and by implication the policymakers who rely on their advice, know little, and they should therefore do little. Let's look at each part separately.

How Much Do Macroeconomists Actually Know?

Macroeconomists are like doctors treating cancer. They know a lot, but there is a lot they don't know.

Take an economy with high unemployment, where the central bank is considering lowering interest rates to increase economic activity. Assume that it has room to decrease the interest rate; in other words, leave aside the even more difficult issue of what to do if the economy is in the liquidity trap (Chapter 4). Think of the sequence of links between a reduction in the interest rate that the central bank controls and the increase in output—all the questions the central bank faces when deciding whether, and by how much, to reduce the interest rate:

- Is the current high rate of unemployment above the natural rate of unemployment, or has the natural rate of unemployment itself increased (Chapter 7)?

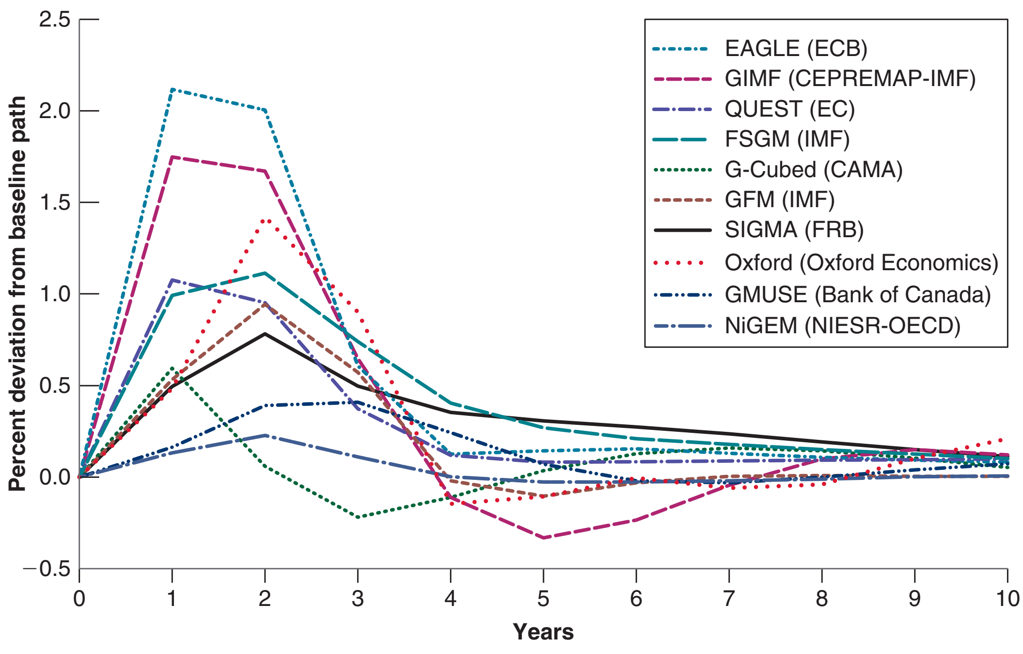


Figure 21-2

The Response of Output to a Monetary Expansion: Predictions from 10 Models

Although all 10 models predict that output will increase for some time in response to a monetary expansion, the range of answers regarding the size and the length of the output response is large.

- If the unemployment rate is close to the natural rate, is there a significant risk that an interest rate reduction will lead to a decrease in unemployment below the natural rate of unemployment and cause an increase in inflation (Chapter 9)?
- What will be the effect of the decrease in the policy rate on the long-term interest rate (Chapter 14)? By how much will stock prices increase (Chapter 14)? By how much will the currency depreciate (Chapters 19 and 20)?
- How long will it take for lower long-term interest rates and higher stock prices to affect investment and consumption spending (Chapter 15)? How much will the exchange rate depreciation improve the trade balance (Chapter 18)? What is the danger that the effects come too late, when the economy has already recovered?

When assessing these questions, central banks—or macroeconomic policymakers in general—do not operate in a vacuum. They rely, in particular, on macroeconometric models. The equations in these models show how these individual links have looked in the past. But different models yield different answers, because they have different structures, different lists of equations, and different lists of variables.

Figure 21-2 gives an example of this diversity. The example comes from a study coordinated by the IMF, asking the builders of 10 main macroeconometric models to answer a similar question: *Trace out the effects of a decrease in the US policy rate by 100 basis points (1%) for two years.*

Three of these models have been developed and are used by central banks; four are used by international organizations, such as the IMF or the OECD; and three are used by academic institutions or commercial firms. They have a roughly similar structure, which you can think of as a more detailed version of the IS-LM-PC framework we developed in Chapter 9. Yet, as you can see, they give rather different answers to the question. Although the average response is for an increase in US output of 0.8% after one year, the answers vary from 0.1% to 2.1%. And after two years, the average response is for an increase of 1.0%, with a range from 0.2% to 2%. In short, if we measure uncertainty by the range of answers from this set of models, there is indeed substantial uncertainty about the effects of policy.

And this is a relatively simple experiment. Think of asking the same question in a country subject to a major financial crisis: How will the change in the interest rate affect the financial system, the perception of risks by investors, and so on?

Should Uncertainty Lead Policymakers to Do Less?

Should uncertainty about the effects of policy lead policymakers to do less? In general, the answer is: yes. Consider the following example, which builds on the simulations we have just looked at.

Suppose the US economy is in recession. The unemployment rate is 6% and the Fed is considering using monetary policy to expand output. To concentrate on uncertainty about the effects of policy, let's assume the Fed knows, with certainty, everything else. Based on its forecasts, it knows that, absent changes in monetary policy, unemployment will still be 6% next year. It knows that the natural rate of unemployment is 4%, and therefore it knows that the unemployment rate is 2% above the natural rate. And it knows, from Okun's law, that 1% more output growth for a year leads to a 0.4% reduction in the unemployment rate.

In the real world, of course, the Fed does not know any of these things with certainty. It can only make educated guesses. It does not know the exact value of the natural rate of unemployment, or the exact coefficient in Okun's law. Introducing these sources of uncertainty would reinforce our basic conclusion.

Under these assumptions, the Fed knows that if it could use monetary policy to achieve 5% more output growth over the coming year, the unemployment rate a year from now would be lower by $0.4 \times 5\% = 2\%$, so it would be down to the natural rate of unemployment, 4%. By how much should the Fed decrease the policy rate?

Taking the average of the responses from the different models in Figure 21-2, a decrease in the policy rate of 1% leads to an increase in output of 0.8% in the first year. Suppose the Fed takes this average relation as holding with *certainty*. What it should then do is straightforward. To return the unemployment rate to the natural rate in one year requires 5% more output growth, which requires the Fed to decrease the policy rate by $5\%/0.8\% = 6.25\%$. If the economy's response is equal to the *average* response from the 10 models, this decrease in the policy rate will return the economy to the natural rate of unemployment at the end of the year.

Suppose the Fed indeed decreases the policy rate by 6.25%. But let's now take into account uncertainty, as measured by the *range* of responses of the different models in Figure 21-2. Recall that the range of responses of output to a 1% decrease in the policy rate varies from 0.1% to 2.1%. This range implies that the decrease in the policy rate leads, across models, to an output response anywhere between 0.625% $(0.1 \times 6.25\%)$ and 13.1% $(2.1 \times 6.25\%)$. These output numbers imply, in turn, a decrease in unemployment anywhere between 0.25% $(0.4 \times 0.625\%)$ and 5.24% $(0.4 \times 13.1\%)$. Put another way, the unemployment rate a year hence could be anywhere between 0.76% $(6\% - 5.24\%)$ and 5.75% $(6\% - 0.25\%)$!

The conclusion is clear: given the range of uncertainty about the effects of monetary policy on output, decreasing the policy rate by 6.25% would be irresponsible. If the effects of the interest rate on output are as strong as suggested by one of the 10 models, unemployment by the end of the year could be 3.24% $(4\% - 0.76\%)$ below the natural rate of unemployment, leading to enormous inflationary pressures. Given this uncertainty, the Fed should decrease the policy rate by much less than 6.25%. For example, decreasing the policy rate by 3% leads to a range for unemployment of 5.9% to 3.5% a year, clearly a safer range of outcomes.

This example relies on the notion of *multiplicative uncertainty*: Because the effects of policy are uncertain, more active policies lead to more uncertainty. See William Brainard, "Uncertainty and the Effectiveness of Policy," *American Economic Review*, 1967, 57(2): pp. 411–425.

Uncertainty and Restraints on Policymakers

Let's summarize: There is substantial uncertainty about the effects of macroeconomic policies. This uncertainty should lead policymakers to be cautious and to limit the use of active policies. Policies should be broadly aimed at avoiding large prolonged recessions, slowing down booms, and avoiding inflationary pressure. The higher unemployment or the higher inflation, the more active the policies should be. One example comes from the recession of 2008–2009 when an unprecedented shift in monetary and fiscal policies probably avoided a repeat of what happened in the

1930s during the Great Depression. But in normal times, macroeconomic policies should stop short of **fine tuning**, of trying to achieve constant unemployment or constant output growth.

These conclusions would have been controversial 20 years ago. Back then, there was a heated debate between two groups of economists. One group, headed by Milton Friedman of the University of Chicago, argued that because of long and variable lags in the effects of policy on activity, activist policy was likely to do more harm than good. The other group, headed by Franco Modigliani of MIT, had just built the first generation of large macroeconometric models and believed that economists' knowledge was becoming good enough to allow for increasingly fine tuning the economy. Today, most economists recognize that there is substantial uncertainty about the effects of policy. They also accept the implication that, except in special circumstances, such as 2008–2009, this uncertainty should lead to less active policies.

Friedman and Modigliani are the same two economists who independently developed the modern theory of consumption we saw in Chapter 15.

Note, however, that what we have developed so far is an argument for *self-restraint by* policymakers, not for *restraints on* policymakers. If policymakers are benevolent—they care about national well-being—and if they understand the implications of uncertainty—and there is no particular reason to think they don't—they will, on their own, follow less active policies. There is no reason to impose further restraints, such as the requirement that money growth be constant or that the budget be balanced. Let's now turn to arguments for restraints *on* policymakers.

21-2 EXPECTATIONS AND POLICY

One of the reasons the effects of macroeconomic policy are uncertain is the interaction of policy and expectations. How a policy works, and sometimes whether it works at all, depends not only on how it affects current variables but also on how it affects expectations about the future (this was the main theme of Chapter 16). The importance of expectations for policy goes, however, beyond uncertainty about the effects of policy. This brings us to a discussion of *games*.

Until 30 years ago, macroeconomic policy was seen in the same way as the control of a complicated machine. Methods of **optimal control**, developed initially to control and guide rockets, were increasingly being used to design macroeconomic policy. Economists no longer think this way. It has become clear that the economy is fundamentally different from a machine, even a very complicated one. Unlike a machine, the economy is composed of people and firms who try to anticipate what policymakers will do, and who react not only to current policy but also to expectations of future policy. Hence, macroeconomic policy must be thought of as a **game** between the policymakers and "the economy"—more concretely, the people and firms in the economy. So, when thinking about policy, what we need is not **optimal control theory** but rather **game theory**.

Even machines are becoming smarter. HAL, the robot in the 1968 movie *2001: A Space Odyssey*, starts anticipating what humans in the space ship will do. The result is not a happy one. (See the movie.)

Warning: When economists say *game*, they do not mean "entertainment"; they mean **strategic interactions** between **players**. In the context of macroeconomic policy, the players are the policymaker on one side and people and firms on the other. The strategic interactions are clear: What people and firms do depends on what they expect policymakers will do; what policymakers do depends on what is happening in the economy.

Game theory has given economists many insights, often explaining how some apparently strange behavior makes sense when one understands the nature of the game. One of these insights is particularly important for our discussion of restraints here. Sometimes you can do better in a game by giving up some of your options. To see why, let's start with an example from outside economics: governments' policies toward hostage takers.

Game theory has become an important tool in all branches of economics. Both the 1994 and the 2005 Nobel Prizes in Economics were awarded to game theorists: in 1994 to John Nash at Princeton, John Harsanyi at Berkeley, and Reinhard Selten in Germany (John Nash's life is portrayed in the movie *A Beautiful Mind*); in 2005 to Robert Aumann in Israel and Tom Schelling at Harvard.

Hostage Takings and Negotiations

Most governments have a stated policy that they will not negotiate with hostage takers. The reason for this stated policy is clear: to deter hostage taking by making it unattractive.

Suppose that, despite the stated policy, someone is taken hostage. Now that a hostage has been taken anyway, why not negotiate? Whatever compensation the hostage takers demand is likely to be less costly than the alternative (i.e., the likelihood that the hostage will be killed). So the best policy would appear to be: Announce that you will not negotiate, but if someone is taken hostage, negotiate.

On reflection, it is clear this would in fact be a very bad policy. Hostage takers' decisions do not depend on the stated policy but on what they expect will actually happen. If they know that negotiations will actually take place, they will rightly consider the stated policy as irrelevant. And hostage takings will happen.

So what is the best policy? Despite the fact that, once a hostage is taken, negotiations typically lead to a better outcome, the best policy is for governments to commit *not* to negotiate. By ruling out the option to negotiate, they are more likely to prevent hostage takings to begin with.

Let's now turn to a macroeconomic example based on the relation between inflation and unemployment. As you will see, even if the circumstances are less dramatic, exactly the same logic is involved.

This example was developed by Finn Kydland, then at Carnegie Mellon and now at the University of California–Santa Barbara, and Edward Prescott, then at Minnesota and now at Arizona State University, in "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 1977, 85(3): pp. 473–492. Kydland and Prescott were awarded the Nobel Prize in Economics in 2004.

Inflation and Unemployment Revisited

Recall the relation between inflation and unemployment we derived in Chapter 8 [equation (8.10), with the time indexes omitted for simplicity]:

$$\pi = \pi^e - \alpha(u - u_n) \tag{21.1}$$

Inflation π depends on expected inflation, π^e, and on the difference between the actual unemployment rate, u, and the natural unemployment rate, u_n. The coefficient α captures the effect of unemployment on inflation, given expected inflation. When unemployment is above the natural rate, inflation is lower than expected; when unemployment is below the natural rate, inflation is higher than expected.

A refresher: Given labor market conditions and given their expectations of the price level, firms and workers set nominal wages through bargaining. Given these nominal wages, firms set prices and this determines the price level. So prices depend on expected prices and labor market conditions. Equivalently, price inflation depends on expected price inflation and labor market conditions. This is what is captured in equation (21.1).

Suppose the Fed announces it will follow a monetary policy consistent with a target of zero inflation ($\bar{\pi} = 0$). On the assumption that people believe the announcement, expected inflation (π^e) as embodied in wage contracts is equal to zero, and the Fed faces the following relation between unemployment and inflation:

$$\pi = -\alpha(u - u_n) \tag{21.2}$$

If the Fed follows through with its announced policy, it will choose an unemployment rate equal to the natural rate; from equation (21.2), inflation will be equal to zero, just as the Fed announced and people expected.

For simplicity, we assume the Fed can choose the unemployment rate—and, by implication, the inflation rate—exactly. In doing so, we ignore the uncertainty about the effects of policy. This was the topic of Section 21-1, but it is not central here.

Achieving zero inflation and an unemployment rate equal to the natural rate is not a bad outcome. But it would seem the Fed can do even better:

- Assume, for the purposes of this computation, α is equal to 0.5. Equation (21.2) implies that, by accepting just 1% inflation, the Fed can achieve an unemployment rate of $1\%/0.5 = 2\%$ below the natural rate of unemployment. Suppose the Fed—and everyone else in the economy—finds the trade-off attractive and decides to decrease unemployment by 2% in exchange for an inflation rate of 1%. This incentive to deviate from the announced policy once the other player has made his move—in this case, once wage setters have set the wage—is known in game theory as the **time inconsistency** of optimal policy. In our example, the Fed can improve

Remember that the natural rate of unemployment is neither natural nor best in any sense (see Chapter 7). It may be reasonable for the Fed and everyone else in the economy to prefer an unemployment rate lower than the natural rate of unemployment.

the outcome this period by deviating from its announced policy of zero inflation: By accepting some inflation, it can achieve a substantial reduction in unemployment.

- Unfortunately, this is not the end of the story. Seeing that the Fed has allowed for more inflation than it announced it would, wage setters are likely to smarten up and begin to expect positive inflation of 1%. If the Fed still wants to achieve an unemployment rate 2% below the natural rate, it will now have to accept 2% inflation because expectations have changed. Accepting an inflation of 1% is no longer enough to sustain lower unemployment. However, if the Fed persists and achieves 2% inflation, wage setters are likely to further increase their expectations of future inflation, and so on.
- The eventual outcome is likely to be persistent high inflation. Because wage setters understand the Fed's motives, expected inflation eventually catches up with actual inflation. The end result is an economy with the *same unemployment rate* that would have prevailed if the Fed had followed its announced policy, but with possibly *much higher inflation.* In short, attempts by the Fed to make things better would lead, in the end, to things being worse.

How relevant is this example? Very relevant. Go back to Chapter 8: We can read the history of the Phillips curve and the increase in inflation in the late 1960s and the 1970s as coming precisely from the Fed's attempts to keep unemployment below the natural rate of unemployment, leading to higher and higher expected inflation, and higher and higher actual inflation. In that light, the shift of the original Phillips curve can be seen as the adjustment of wage setters' expectations to the central bank's behavior.

What, then, is the best policy for the Fed to follow in this case? It is to make a credible commitment that it will not try to decrease unemployment below the natural rate. By giving up the option of deviating from its announced policy, the Fed can achieve unemployment equal to the natural rate of unemployment and zero inflation. The analogy with the hostage taking example is clear. By credibly committing not to do something that would appear desirable at the time, policymakers can achieve a better outcome; no hostage takings (in our previous example), and no inflation here.

Establishing Credibility

How can a central bank credibly commit to not deviate from its announced policy?

One way to establish its credibility is for the central bank to give up—or to be stripped by law of—its policymaking power. For example, the mandate of the central bank can be defined by law in terms of a simple rule, such as keeping money growth at 0% (or more generally, at x%) forever. (An alternative, which we discussed in Chapter 20, is to adopt a hard peg, such as a currency board or, in other countries, dollarization. In this case the central bank must keep interest rates equal to foreign rates no matter what.)

Such a law surely takes care of the problem of time inconsistency. But the tight restraint it creates comes close to throwing the baby out with the bath water. We want to prevent the central bank from pursuing too high a rate of money growth in an attempt to maintain unemployment below the natural unemployment rate. But—subject to the restrictions discussed in Section 21-1—we still want the central bank to be able to decrease the policy rate by expanding the money supply when unemployment is far above the natural rate, and to increase the policy rate by contracting the money supply when unemployment is far below the natural rate. Such actions become impossible under a constant-money-growth rule. There are indeed better ways to deal with time inconsistency. In the case of monetary policy, our discussion suggests various ways of dealing with the problem.

A first step is to make the central bank independent. By an **independent central bank**, we mean a central bank where interest rate and money supply decisions are made independent of the influence of currently elected politicians. Politicians, who face frequent reelections, may want lower unemployment now, even if it leads to inflation later. Making the central bank independent and making it difficult for politicians to fire the central banker makes it easier for the central bank to resist political pressure to decrease unemployment below the natural rate of unemployment.

This may not be enough, however. Even if it is not subject to political pressure, the central bank may be tempted to decrease unemployment below the natural rate. Doing so leads to a better outcome in the short run. So a second step is to give incentives to the central bankers to take the long view—that is, to take into account the long-run costs of higher inflation. One way of doing so is to give them long terms in office, so they have a long horizon and have the incentives to build credibility.

A third, and more controversial, step may be to appoint a "conservative" central banker, someone who dislikes inflation very much and is therefore less willing to accept more inflation in exchange for less unemployment when unemployment is at the natural rate. When the economy is at the natural rate, such a central banker will be less tempted to embark on a monetary expansion. Thus, the problem of time inconsistency will be reduced.

These are the steps many countries have taken over the last two decades. Central banks have been given more independence from governments. Central bankers have been given long terms in office. And governments typically have appointed central bankers who are more "conservative" than the governments themselves—central bankers who appear to care more about inflation and less about unemployment than the government does. (See the Focus Box "Was Alan Blinder Wrong in Speaking the Truth?")

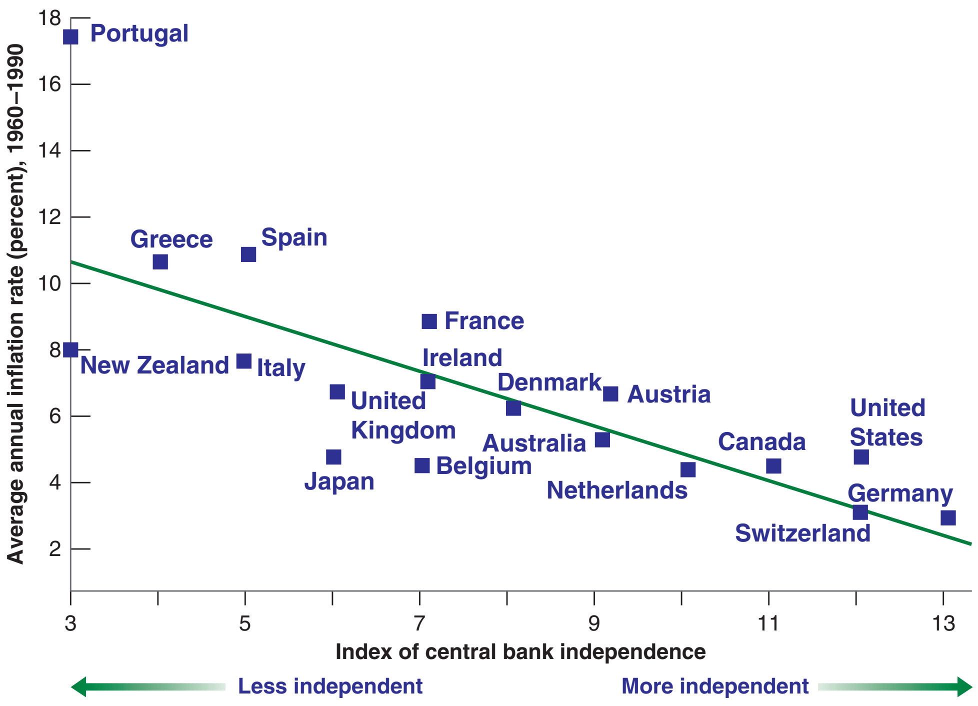
Have these steps been successful? Figure 21-3 suggests that they have, at least in terms of delivering low inflation. The vertical axis gives the average inflation rate in 18 OECD countries over the period 1960–1990, when some central banks were independent, others much less so. (Today, most central banks are, at least formally, independent.) The horizontal axis gives the value of an index of "central bank independence," constructed by looking at a number of legal provisions in the central bank's charter—for example, whether and how the government can remove the head of the bank. There is a striking inverse relation between the two variables, as illustrated by the regression line: More

Figure 21-3

Inflation and Central Bank Independence

Across OECD countries, the higher the degree of central bank independence, the lower the rate of inflation.

Source: Vittorio Grilli, Donato Masciandaro, and Guido Tabellini, "Political and Monetary Institutions and Public Financial Policies in the Industrial Countries," *Economic Policy*, 1991, 6(13): pp. 341–392.



Was Alan Blinder Wrong in Speaking the Truth?

FOCUS

In the summer of 1994, President Bill Clinton appointed Alan Blinder, an economist from Princeton, vice chair (in effect, second in command) of the Federal Reserve Board. A few weeks later, Blinder, speaking at an economic conference, indicated his belief that the Fed had both the responsibility and the ability, when unemployment was high, to use monetary policy to help the economy recover. This statement was badly received. Bond prices fell, and most newspapers ran editorials critical of Blinder.

Why was the reaction of markets and newspapers so negative? It was surely not that Blinder was wrong. There is no doubt that monetary policy can and should help the economy out of a recession. Indeed, the Federal Reserve Bank Act of 1977 requires the Fed to pursue full employment as well as low inflation.

The reaction was negative because, in terms of the argument we developed in the text, Blinder revealed by his words that he was not a conservative central banker, that he cared about unemployment as well as inflation. With the unemployment rate at the time equal to 6.1%, close to what was thought to be the natural rate of unemployment at the time, markets interpreted Blinder's statements as suggesting that he might want to decrease unemployment below the natural rate. Interest rates increased because of higher expected inflation, and bond prices decreased.

The moral of the story: Whatever views central bankers may hold, they should try to look and sound conservative. This is why, for example, many heads of central banks are reluctant to admit, at least in public, the existence of any trade-off between unemployment and inflation, even in the short run.

central bank independence appears to be systematically associated with lower inflation. The mechanism: Central bank independence makes the announced inflation target more credible. This leads inflation expectations to stay close to the announced target, and this in turn makes it easier for the central bank to achieve its target. Chapter 8 showed, however, that establishing credibility does not happen overnight. The Fed indicated a commitment to low inflation as early as the early 1980s, but it took until the mid-1990s for inflation expectations to become anchored at the target inflation rate.

A warning: Figure 21-3 shows correlation, not necessarily causality. It may be that countries that dislike inflation tend to both give more independence to their central bankers and have lower inflation. (This is another example of the difference between correlation and causality, discussed in Appendix 3 at the end of the book.)

Time Consistency and Restraints on Policymakers

Let's summarize what we have learned in this section:

We have examined arguments for putting restraints on policymakers based on the issue of time inconsistency.

When issues of time inconsistency are relevant, tight restraints on policymakers—like a fixed-money-growth rule in the case of monetary policy, or a balanced budget rule in the case of fiscal policy—can provide a rough solution. But the solution has large costs because it prevents the use of macroeconomic policy altogether. Better solutions typically involve designing better institutions (like an independent central bank, or a better budget process) that can reduce the problem of time inconsistency while allowing the use of policy for the stabilization of output. This is not easy to do, however.

21-3 POLITICS AND POLICY

We have assumed so far that policymakers are benevolent, that is, that they try to do what is best for the economy. However, much public discussion challenges that assumption. Politicians or policymakers, the argument goes, do what is best for themselves, and this is not always what is best for the country.

You have heard the arguments. Politicians avoid the hard decisions and they pander to the electorate, partisan politics leads to gridlock, and nothing gets done. Discussing the flaws of democracy goes far beyond the scope of this book. What we can do here is to briefly review how these arguments apply to macroeconomic policy, then look at the empirical evidence and see what light it sheds on the issue.

Games between Policymakers and Voters

Many macroeconomic policy decisions involve trading off short-run losses against long-run gains or, conversely, short-run gains against long-run losses.

Take, for example, tax cuts. By definition, tax cuts lead to lower taxes today. They are also likely to lead to an increase in demand, and therefore to an increase in output for some time. But unless they are matched by equal decreases in government spending, they lead to a larger budget deficit and to the need for an increase in taxes in the future. If voters are shortsighted, the temptation for politicians to cut taxes may prove irresistible. Politics may lead to systematic deficits, at least until the level of government debt has become so high that politicians are scared into action.

Now let's move on from taxes to macroeconomic policy in general. Again, suppose that voters are shortsighted. If the politicians' main goal is to please voters and get reelected, what better policy than to expand aggregate demand before an election, leading to higher growth and lower unemployment? True, growth in excess of the normal growth rate cannot be sustained, and eventually the economy must return to the natural level of output. Higher growth now must be followed by lower growth later. And lower taxes now must be followed by higher taxes later. But with the right timing and shortsighted voters, higher growth and lower taxes can win elections. Thus, we might expect a clear **political business cycle** (i.e., economic fluctuations induced by political elections) associated with higher growth on average before elections than after elections.

We saw in Chapter 9 that, even if monetary policy is used to increase output in the short run, in the medium run, output returns to its natural level and unemployment to its natural rate.

You probably have heard these arguments before, in one form or another. And their logic appears convincing. The question is: How well do they fit the facts?

First, consider deficits and debt. The preceding argument would lead you to expect that budget deficits and high government debt always have been and always will be there. Figure 21-4 takes the long view. It gives the evolution of the ratio of government debt to GDP in the United States since 1900 and suggests a more complex reality.

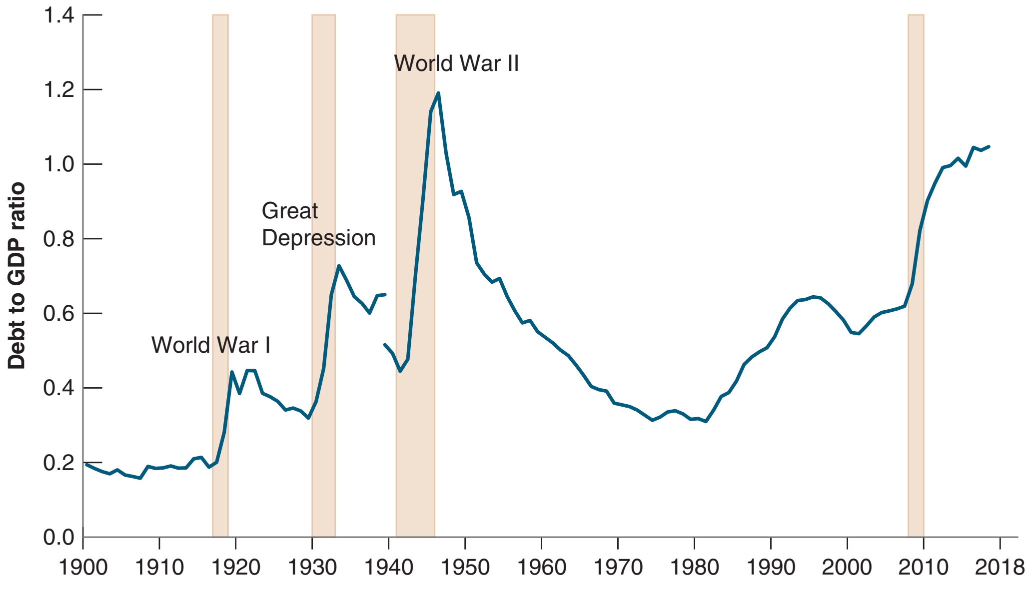
Look first at the evolution of the ratio of debt to GDP from 1900 to 1980. Each of the first three buildups in debt (represented by the shaded bars) was associated with special circumstances: World War I, the Great Depression, and World War II. These were times of unusually high military spending or unusual declines in output. Adverse circumstances—not pandering to voters—were clearly behind the large deficits and the resulting increase

Figure 21-4

The Evolution of the US Debt to GDP Ratio since 1900

The three major buildups of debt since 1900 have been associated with World War I, the Great Depression, and World War II. In contrast, except for the increase in debt due to the financial crisis, the buildup since 1980 has not been caused by either wars or adverse economic shocks.

Source: 1900 to 1939: Historical Statistics of the United States; US Census Bureau. 1939 on: FRED GFDGDPA188S (note the break in the series in 1939)



in debt during each of these three episodes. Note also that, in each of these three cases, the buildup was followed by a steady decrease in debt. The ratio of debt to GDP, which was as high as 120% in 1946, was steadily reduced to a postwar low of 31% in 1981.

The precise relation between the evolution of deficits, debt, and the ratio of debt to GDP is explored in detail in Chapter 22. For the moment, all you need to know is that deficits lead to increases in debt.

The more recent evidence is less reassuring. The ratio of debt to GDP has climbed back up to reach 105% in 2018. The increase has three main causes: large tax cuts under the Reagan administration in the early 1980s, another round of large tax cuts under the Bush administration in the early 2000s, and the large deficits triggered by the financial crisis in 2008 and 2009.

As we saw in Chapter 6, the large deficits during the financial crisis were largely justified on macroeconomic grounds; without them, the recession would have been much worse. But what about the two rounds of tax cuts? Were they motivated by a desire to pander to shortsighted voters? Probably not, or at least not mainly. One motivation, discussed below, was to "starve the beast," to create deficits in order to create the political pressure to cut spending later. Another was based on **supply side theories**, the belief that tax cuts would lead people to work more and firms to be more productive, leading to higher potential output and higher tax revenues in the future. (The hope that tax cuts would lead to higher tax revenues never materialized, and debt ended up higher.)

The main proponent of supply side theories was an economist named Arthur Laffer. It is said that, in 1974, he convinced then President Ford by drawing a graph of tax revenues against the tax rate on a napkin. When tax rates were too high, which he argued was the case in the United States, lower tax rates would increase rather than decrease revenues.

In short, the simple hypothesis that the combination of cynical politicians and shortsighted voters must inevitably lead to steady deficits and high debt does not fit the facts. What about the political business cycle argument, that policymakers try to get high output growth before the elections so they will be reelected? If the political business cycle is important, we would expect to see faster growth before elections rather than after. With this motivation, Table 21-1 gives average output growth rates for each of the four years of each US administration from 1948 to 2018, distinguishing between Republican and Democratic presidential administrations. Look at the last line of the table. Growth has indeed been highest on average in the last year of an administration: 3.6%, compared to 2.8% in the first year. (We shall return below to another interesting feature in the table, namely the difference between Republican and Democratic administrations.) This can be read as supporting evidence of some manipulation of economic policy to win elections.

To summarize: Policymakers are not indifferent to their electoral prospects, and voters are often shortsighted. The evidence on political business cycles suggests that this plays a role in the setting of macroeconomic policy. But, if we are to explain the evolution of debt and deficits, it is clear that other factors are at work: wars, crises, unusual economic theories, and, as we explore in the next subsection, games between policymakers.

Games between Policymakers

Suppose that the party in power wants to reduce spending but faces opposition to spending cuts in Congress. One way of putting pressure both on Congress and on future parties in power is to cut taxes and create deficits. As debt increases over time, the increasing

Table 21-1 Growth during Democratic and Republican Presidential Administrations: 1948–2018

	Year of administration				
	First (%)	**Second (%)**	**Third (%)**	**Fourth (%)**	**Average (%)**
Democratic	2.4	5.1	3.8	3.4	3.7
Republican	3.3	0.9	3.3	3.8	2.8
Average	2.8	3.0	3.5	3.6	3.2

Source: Calculated using Series GDPCA, from 1948 to 2018: Federal Reserve Economic Data (FRED) https://fred.stlouisfed.org/

This strategy goes by the ugly but revealing name of "Starve the Beast." The expression is attributed to a staffer in the Reagan administration.

pressure to reduce deficits may force Congress and future parties in power to reduce spending—something they would not have been willing to do otherwise.

Or suppose that, for any reason, the country is facing large budget deficits. Both parties in Congress want to reduce the deficit, but they disagree about the way to do it. One party wants to preserve spending and reduce deficits primarily through an increase in taxes; the other wants to decrease spending. Both parties may hold out on the hope that the other side will give in first. Only when debt has increased sufficiently, and it becomes urgent to reduce deficits, will one party give up.

Game theorists refer to these situations as **wars of attrition**. The hope that the other side will give in leads to long and often costly delays. Such wars of attrition happen often in the context of fiscal policy, and deficit reduction occurs long after it should.

Another example outside of economics: During the 2004–2005 National Hockey League lockout, where the complete season was cancelled because owners and players could not reach an agreement. The National Basketball Association faced a similar lockout during the summer of 2011.

Wars of attrition arise in other macroeconomic contexts; for example, during episodes of hyperinflation. As we shall see in Chapter 22, hyperinflations come from the use of money creation to finance large budget deficits. Although the need to reduce those deficits is usually recognized early on, support for stabilization programs—which include the elimination of those deficits—typically comes only after inflation has reached such high levels that economic activity is severely affected.

See the discussion in the Focus Box "Monetary Contraction and Fiscal Expansion: The United States in the Early 1980s" in Chapter 19.

More on the current US fiscal situation in Chapter 22.

These games also go a long way in explaining the rise in the ratio of debt to GDP in the United States since the early 1980s. There is little doubt that one of the goals of the Reagan administration, when it decreased taxes from 1981 to 1983, was to slow the growth of government spending. There is also little question that, by the mid-1980s, there was general agreement among policymakers that the deficits should be reduced. But, because of disagreements between Democrats and Republicans about whether this should happen primarily through tax increases or spending cuts, it was not until the late 1990s that deficit reduction was achieved. The same dynamics followed the Bush administration tax cuts of the early 2000s. And these dynamics may well be at play again: The Trump administration tax cuts have led to larger deficits. Whether these are eventually reduced through tax increases or spending cuts may depend on the results of another war of attrition between Republicans and Democrats.

Another example of games between political parties is the movements in economic activity brought about by the alternation of parties in power. Traditionally, Republicans have worried more than Democrats about inflation and worried less than Democrats about unemployment. So we would expect Democratic administrations to show stronger growth—and thus less unemployment and more inflation—than Republican administrations. This prediction appears to fit the facts quite well. Look at Table 21-1 again. Average growth has been 3.7% during Democratic administrations, compared to 2.8% during Republican administrations. The most striking contrast is in the second year: 5.1% during Democratic administrations compared to 0.9% during Republican administrations.

This raises an intriguing question. Why is the effect so much stronger in the administration's second year? It could just be a fluke. Many other factors affect growth. But the theory of unemployment and inflation we developed in Chapter 8 suggests a possible hypothesis. There are lags in the effects of policy, so it takes about a year for a new administration to affect the economy. And sustaining higher growth than normal for too long would lead to increasing inflation, so even a Democratic administration would not want to sustain higher growth throughout its term. Thus, growth rates tend to be much closer to each other during the second half of Democratic and Republican administrations—more so than during the first half.

Politics and Fiscal Restraints

If politics sometimes lead to large and lasting budget deficits, can rules be put in place to limit these adverse effects?

A constitutional amendment to balance the budget each year, such as the amendment proposed by the Republicans in 1994, would eliminate the problem of deficits. But just like a constant-money-growth rule in the case of monetary policy, it also would eliminate the use of fiscal policy as a macroeconomic instrument altogether. This is too high a price to pay.

A better approach is to put in place rules that put limits either on deficits or on debt. But this is more difficult than it sounds. Rules such as limits on the ratio of the deficit to GDP or the ratio of debt to GDP are more flexible than a balanced budget requirement, but they may still not be flexible enough if the economy is affected by particularly bad shocks. This has been made clear by the problems associated with the Stability and Growth Pact; these problems are discussed in the Focus Box "Euro Area Fiscal Rules: A Short History." More flexible or more complex rules—rules that allow for special circumstances or take into account the state of the economy—are harder to design and harder to enforce. For example, allowing the deficit to be higher if the unemployment rate is higher than the natural rate requires a simple and unambiguous way of computing what the natural rate is, a nearly impossible task.

A complementary approach is to put in place mechanisms to reduce deficits, were such deficits to arise. Consider, for example, a mechanism that triggers automatic spending cuts when the deficit gets too large. Suppose the budget deficit is too large and it is desirable to cut spending across the board by 5%. Members of Congress will find it difficult to explain to their constituency why their favorite spending program was cut by 5%. Now suppose the deficit triggers automatic across-the-board spending cuts of 5% without any congressional action. Knowing that other programs will be cut, members of Congress will accept cuts in their favorite programs more easily. They will also be better able to deflect the blame for the cuts. Members of Congress who succeed in limiting the cuts to their favorite program to, say, 4% (by convincing Congress to make deeper cuts in some other programs to maintain the lower overall level of spending) can then return to their constituents and claim they have successfully prevented even larger cuts.

This was the general approach used to reduce deficits in the United States in the 1990s. The Budget Enforcement Act passed in 1990, and extended by new legislation in 1993 and 1997, introduced two main rules:

- It imposed constraints on spending. Spending was divided into two categories: discretionary (roughly, spending on goods and services, including defense) and mandatory (roughly, transfer payments to individuals). Constraints, called **spending caps**, were set on discretionary spending for the following five years and required a small but steady decrease in discretionary spending (in real terms), with explicit provisions for emergencies. For example, spending on Operation Desert Storm during the Gulf War in 1991 was not subject to the caps.
- It stipulated that a new transfer program could be adopted only if it could be shown not to increase deficits in the future (either by raising new revenues or by decreasing spending on an existing program). This rule is known as the pay-as-you-go or **PAYGO rule**.

The focus on spending rather than on the deficit itself had one important implication. If there was a recession, hence a decrease in revenues, the deficit could increase without triggering a decrease in spending. This happened in 1991 and 1992. And the focus on spending had two desirable effects: It allowed for a larger fiscal deficit during a recession—a good thing from the point of view of macroeconomic policy; and it decreased the pressure to break the rules during a recession—a good thing from a political point of view.

By 1998, deficits were gone, and for the first time in 20 years, the federal budget was in surplus. Not all the deficit reduction was due to the Budget Enforcement

FOCUS

Euro Area Fiscal Rules: A Short History

The Maastricht treaty, negotiated by the countries of the European Union in 1991, set a number of convergence criteria that countries had to meet in order to qualify for membership in the euro area (for more on the history of the euro, see the Focus Box "The Euro: A Short History" in Chapter 20). Among them were two restrictions on fiscal policy. First, the ratio of the budget deficit to GDP had to be lower than 3%. Second, the ratio of debt to GDP had to be less than 60%, or at least "approaching this value at a satisfactory pace."

In 1997, would-be members of the euro area agreed to make some of these restrictions permanent. The Stability and Growth Pact (SGP), signed in 1997, required members of the euro area to adhere to the following fiscal rules:

- That countries commit to balance their budget in the medium run. That they agree to present programs to the European authorities, specifying their objectives for the current and following three years to show how they are making progress toward their medium-run goal.
- That countries avoid excessive deficits, except under exceptional circumstances. Following the Maastricht treaty criteria, excessive deficits were defined as deficits in excess of 3% of GDP. Exceptional circumstances were defined as declines of GDP larger than 2%.
- That sanctions be imposed on countries that ran excessive deficits. These sanctions could range from 0.2 to 0.5% of GDP—so, for a country like France, up to roughly 10 billion euros!

Figure 1 plots the evolution of budget deficits since 1995 for the euro area as a whole. Note that from 1995 to 2000 budget balances went from a deficit of 7.5% of euro area GDP to budget balance. The performance of some of the member countries was particularly impressive. Greece reduced its deficit from 13.4% of GDP to a reported 1.4% of GDP. (It was discovered later that the Greek government had cheated in reporting its deficit numbers and that the actual improvement, although impressive, was less than reported; the deficit for 2000 is now estimated to have been 4.1%.) Italy's deficit went from 10.1% of GDP in 1993 to only 0.9% of GDP in 2000.

Was the improvement entirely due to the Maastricht criteria and the SGP rules? No, the strong expansion of the late 1990s played an important role. But so did the fiscal rules. The carrot—the right to become a member of the euro area—was attractive enough to lead several countries to take tough measures to reduce their deficits.

Things changed, however, after 2000. Deficits started increasing. The first country to break the limit was Portugal in 2001, with a deficit of 4.4%. The next two countries were France and Germany, both with deficits in excess of 3% of GDP in 2002. Italy soon followed. In each case, the government of the country decided it was more important to avoid a fiscal contraction that could lead to even slower output growth than to satisfy the rules of the SGP.

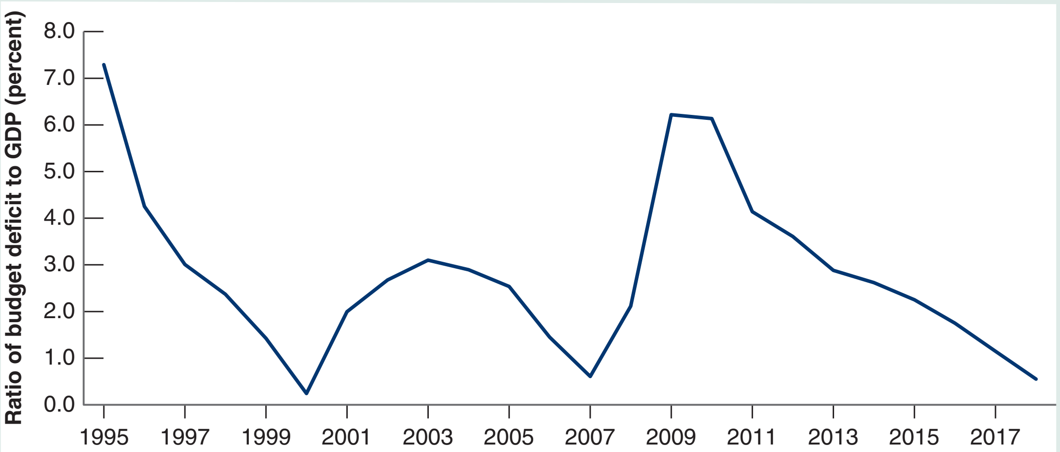
Faced with clear "excessive deficits" (and without the excuse of exceptional circumstances because output growth in each of these countries was low but positive), European authorities found themselves in a quandary. Starting the excessive deficit procedure against Portugal, a small country, might have been politically feasible, although it is doubtful that Portugal would have ever been willing to pay the fine. Starting the same procedure against the two largest members of the euro area, France and Germany, proved politically impossible. After an internal fight between the two main European authorities, the European Commission and the European Council—the European Commission wanted to proceed with the excessive deficit procedure, whereas the European Council, which represents the states, did not—the procedure was suspended.

The crisis made it clear that the initial rules were too inflexible. Romano Prodi, the president of the European Commission, admitted to that much. In an interview in October 2002, he stated, "I know very well that the Stability Pact is stupid, like all decisions that are rigid." And the attitudes of both France and Germany showed that the threat

Figure 1

Euro Area Budget Deficit as a Percentage of GDP since 1995

Source: European Central Bank. References: If you want to get a sense of the complexity of the current rules, read the Wikipedia entry at https://en.wikipedia.org/wiki/European_Fiscal_Compact.



to impose large fines on countries with excessive deficits was simply not credible.

For two years, the European Commission explored ways to improve the rules to make them more flexible and, by implication, more credible. In 2005, a new, revised SGP was adopted. It kept the 3% deficit and 60% debt numbers as thresholds but allowed for more flexibility in deviating from the rules. Growth no longer had to be less than -2% for the rules to be suspended. Exceptions were also made if the deficit came from structural reforms or from public investment. Fines were gone, and the plan was to rely on early public warnings as well as on peer pressure from other euro area countries.

For a while, the ratio of the deficit to GDP declined, again largely due to strong growth and higher revenues. The ratio reached a low of 0.5% in 2007. But the crisis, and the associated sharp decrease in revenues, led again to a sharp increase in budget deficits. In 2010, the ratio was close to 6%, twice the SGP threshold; 23 of 27 EU countries stood in violation of the 3% deficit limit, and it was clear that the rules had to be reconsidered. Eventually, in 2012, a new intergovernmental treaty was signed among the member countries of the European Union: the Treaty on Stability, Coordination and Governance in the Economic and Monetary Union, also known as the Fiscal Compact. It has four main provisions:

- Member countries should introduce a balanced budget rule into national legislation, through either a constitutional amendment or a framework law.
- Government budgets should be balanced or in surplus. The treaty defines a balanced budget as a budget deficit not exceeding 3.0% of GDP, and a cyclically adjusted deficit not exceeding a country-specific objective, which at most can be set to 0.5% of GDP for states with a debt-to-GDP ratio exceeding 60%, or at most 1.0% of GDP for states with debt levels within the 60% limit.
- Countries whose government debt-to-GDP ratio exceeds 60% must reduce it at an average rate of at least one twentieth (5%) per year of the exceeded percentage points. (So, for example, if the actual ratio of debt to GDP is 100%, they must decrease by at least $0.05\ (100 - 60) = 2\%$ of GDP.)
- If a country's budget shows a significant deviation from the second rule, an automatic correction mechanism is triggered with a procedure called the Excessive Deficit Procedure. The exact implementation of this mechanism is defined individually by each country, but it has to comply with the basic principles outlined by the European Commission. The convoluted procedure is graphically well explained here: https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/european-economy-explained/graphs-economic-topics_en (2014–11-10).

In 2015 a new criterion was added to the four, which specifies that in deciding whether a country should be subject to the Excessive Deficit Procedure, its progress in implementing structural reforms (e.g., reforms of pensions, labor, goods, and services markets) will also be considered.

By 2018 the average budget deficit of euro area countries had fallen to 0.7% and only 1 of 19 euro member countries (Spain) was still under the Excessive Deficit Procedure because it was in violation of another of the Fiscal Compact rules. There is wide agreement that the rules have become too complex and too confusing, and that they must be simplified. Work is ongoing, but designing a simpler set of rules is proving difficult.

Act rules. A decrease in defense spending due to the end of the Cold War and a large increase in tax revenues due to the strong expansion in the second half of the 1990s were important factors. But there is wide agreement that the rules played an important role in making sure that decreases in defense spending and increases in tax revenues were used for deficit reduction rather than for increases in other spending programs.

Once budget surpluses appeared, however, Congress became increasingly willing to break its own rules. Spending caps were systematically broken, and the PAYGO rule was allowed to expire in 2002. It was reintroduced by President Obama in 2010, but it has not prevented an increase in deficits due largely to tax cuts. The lesson from this, as well as from the failure of the SGP described in the Focus Box "Euro Area Fiscal Rules: A Short History," is that, although rules can help, they cannot fully substitute for a lack of resolve from policymakers.

SUMMARY

- The effects of macroeconomic policies are always uncertain. This uncertainty should lead policymakers to be more cautious and to use less active policies. Policies must be broadly aimed at avoiding prolonged recessions, slowing down booms, and avoiding inflationary pressure. The higher the level of unemployment or inflation, the stronger the policies should be. But they should stop short of fine tuning, of trying to maintain constant unemployment or constant output growth.
- Using macroeconomic policy to control the economy is fundamentally different from controlling a machine. Unlike a machine, the economy is composed of people and firms who try to anticipate what policymakers will do and who react not only to current policy but also to expectations of future policy. In this sense, macroeconomic policy can be thought of as a game between policymakers and people in the economy.
- When playing a game, it is sometimes better for a player to give up some options. For example, when a hostage taking occurs, it is better to negotiate with the hostage takers. But a government that credibly commits to not negotiating with hostage takers—a government that gives up the option of negotiation—is actually more likely to deter hostage takings.
- The same argument applies to various aspects of macroeconomic policy. By credibly committing not to use monetary policy to decrease unemployment below the natural rate of unemployment, a central bank can alleviate fears that money growth will be high, and in the process decrease both expected and actual inflation. When issues of time inconsistency are relevant, tight restraints on policymakers—such as a fixed-money-growth rule in the case of monetary policy—can provide a rough solution. But the solution can have large costs if it prevents the use of macroeconomic policy altogether. Better methods typically involve designing better institutions (such as an independent central bank) that can reduce the problem of time inconsistency without eliminating monetary policy as a macroeconomic policy tool.
- Another argument for putting restraints on policymakers is that they may play games either with the public or among themselves, and these games may lead to undesirable outcomes. Politicians may try to fool a shortsighted electorate by choosing policies with short-run benefits but large long-term costs—for example, large budget deficits—to be reelected. Political parties may delay painful decisions, hoping that the other party will make the adjustment and take the blame. In cases like this, tight restraints on policy, such as a constitutional amendment to balance the budget, again provide a rough solution. Better ways typically involve better institutions and better ways of designing the process through which policy and decisions are made. However, the design and consistent implementation of such fiscal frameworks has proven difficult in practice, as demonstrated in both the United States and the European Union.

KEY TERMS

Stability and Growth Pact (SGP), 443
fine tuning, 447
optimal control, 447
game, 447
optimal control theory, 447
game theory, 447
strategic interactions, 447
players, 447
time inconsistency, 448
independent central bank, 449
political business cycle, 452
supply side theories, 453
wars of attrition, 454
spending caps, 455
PAYGO rule, 455

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. There is so much uncertainty about the effects of monetary policy that we would be better off not using it.
b. Depending on the model used, a one percentage point reduction in the policy interest rate is estimated to increase output growth in the year of the interest rate cut by as little as 0.1 percentage point.
c. Depending on the model used, a one percentage point reduction in the policy interest rate is estimated to increase output growth in the year of the interest rate cut by as much as 2.1 percentage points.
d. Elect a Democrat as president if you want low unemployment.
e. There is clear evidence of political business cycles in the pattern of growth rates in the United States.
f. Fiscal spending rules in the United States have been ineffective in reducing budget deficits.
g. Balanced budget rules in Europe have been effective in constraining budget deficits.

h. Governments would be wise to announce a no-negotiation policy with hostage takers.

i. If hostages are taken, it is clearly better for governments to negotiate with hostage takers, even if the government has announced a no-negotiation policy.

j. There is some evidence that countries with more independent central banks have generally lower inflation.

k. In a "starve-the-beast" fiscal policy, spending cuts come before tax cuts.

2. Implementing a political business cycle

You are the economic adviser to a newly elected president. In four years he or she will face another election. Voters want a low unemployment rate and a low inflation rate. However, you believe that voting decisions are influenced heavily by the values of unemployment and inflation in the last year before the election, and that the economy's performance in the first three years of a president's administration has little effect on voting behavior.

Assume that inflation last year was 10%, and that the unemployment rate was equal to the natural rate. The Phillips curve is assumed to be given by

$$\pi_t = \pi_{t-1} - \alpha(u_t - u_n)$$

Assume that you can use fiscal and monetary policy to achieve any unemployment rate you want for each of the next four years. Your task is to help the president achieve low unemployment and low inflation in the last year of his or her administration.

a. Suppose you want to achieve a low unemployment rate (i.e., an unemployment rate below the natural rate) in the year before the next election (four years from today). What will happen to inflation in the fourth year?

b. Given the effect on inflation you identified in part a, what would you advise the president to do in the early years of the administration to achieve low inflation in the fourth year?

c. Now suppose the Phillips curve is given by

$$\pi_t = \pi_t^e - \alpha(u_t - u_n)$$

In addition, assume that people form inflation expectations, π_t^e, based on consideration of the future (as opposed to looking only at inflation last year) and are aware that the president has an incentive to carry out the policies you identified in parts a and b. Are the policies you described in those parts likely to be successful? Why or why not?

3. Suppose the government amends the constitution to prevent government officials from negotiating with terrorists.

What are the advantages of such a policy? What are the disadvantages?

4. In 1989, New Zealand rewrote the charter of its central bank to make low inflation its only goal.

Why would New Zealand want to do this?

5. In 2018 New Zealand rewrote the charter of its central bank to include high employment as well as low inflation as its goals.

Why would New Zealand want to do this?

DIG DEEPER

6. Political expectations, inflation, and unemployment

Consider a country with two political parties, Democrats and Republicans. Democrats care more about unemployment than Republicans, and Republicans care more about inflation than Democrats. When Democrats are in power, they choose an inflation rate of π_D, and when Republicans are in power, they choose an inflation rate of π_R. We assume that

$$\pi_D > \pi_R$$

The Phillips curve is given by

$$\pi_t = \pi_t^e - \alpha(u_t - u_n)$$

An election is about to be held. Assume that expectations about inflation for the coming year (represented by π_t^e) are formed before the election. (Essentially, this assumption means that wages for the coming year are set before the election.) Moreover, Democrats and Republicans have an equal chance of winning the election.

a. Solve for expected inflation, in terms of π_D and π_R.

b. Suppose the Democrats win the election and implement their target inflation rate, π_D. Given your solution for expected inflation in part a, how will the unemployment rate compare to the natural rate of unemployment?

c. Suppose the Republicans win the election and implement their target inflation rate, π_R. Given your solution for expected inflation in part a, how will the unemployment rate compare to the natural rate of unemployment?

d. Do these results fit the evidence in Table 21-1? Why or why not?

e. Now suppose that everyone expects the Democrats to win the election, and the Democrats indeed win. If the Democrats implement their target inflation rate, how will the unemployment rate compare to the natural rate?

f. If the central bank sets an inflation target and monetary policy is not affected by who wins the election because the central bank is independent, do the preferences of the Republicans and Democrats matter? If the Federal Reserve were truly independent, how would we explain the results in Table 21-1?

7. Deficit reduction as a prisoner's dilemma game

Suppose there is a budget deficit. It can be reduced by cutting military spending, by cutting welfare programs, or by cutting both. The Democrats have to decide whether to support cuts in welfare programs. The Republicans have to decide whether to support cuts in military spending.

The possible outcomes are represented in the following table:

		Welfare Cuts	
		Yes	**No**
Defense Cuts	**Yes**	$(R = 1, D = -2)$	$(R = -2, D = 3)$
	No	$(R = 3, D = -2)$	$(R = -1, D = -1)$

The table presents payoffs to each party under the various outcomes. Think of a payoff as a measure of happiness for a given party under a given outcome. If Democrats vote for welfare cuts, and Republicans vote against cuts in military spending, the Republicans receive a payoff of 3, and the Democrats receive a payoff of −2.

a. If the Republicans decide to cut military spending, what is the best response of the Democrats? Given this response, what is the payoff for the Republicans?

b. If the Republicans decide not to cut military spending, what is the best response of the Democrats? Given this response, what is the payoff for the Republicans?
c. What will the Republicans do? What will the Democrats do? Will the budget deficit be reduced? Why or why not? (A game with a payoff structure like the one in this problem, and that produces the outcome you have just described, is known as a *prisoner's dilemma*.) Is there a way to improve the outcome?

EXPLORE FURTHER

8. Games, pre-commitment, and time inconsistency in the news

Current events offer abundant examples of disputes in which the parties are involved in a game, try to commit themselves to lines of action in advance, and face issues of time inconsistency. Examples arise in the domestic political process, international affairs, and labor-management relations.

a. Choose a current dispute (or one resolved recently) to investigate. Do an internet search to learn the issues involved in the dispute, the actions taken by the parties to date, and the current state of play.
b. In what ways have the parties tried to pre-commit to certain actions in the future? Do they face issues of time inconsistency? Have the parties failed to carry out any of their threatened actions?
c. Does the dispute resemble a prisoner's dilemma game (a game with a payoff structure like the one described in Problem 6)? In other words, does it seem likely (or did it actually happen) that the individual incentives of the parties will lead them to an unfavorable outcome—one that could be improved for both parties through cooperation? Is there a deal to be made? What attempts have the parties made to negotiate?
d. How do you think the dispute will be resolved (or how has it been resolved)?

9. The legislation governing the Federal Reserve Board

The 1977 Federal Reserve Act, as amended in 1978, 1988, and 2000, governs the behavior of the Federal Reserve.

a. In your opinion, does this excerpt from the Act make the policy goals of the Fed clear?

> ***Section 2B. Monetary policy objectives***
>
> *The Board of Governors of the Federal Reserve System and the Federal Open Market Committee shall maintain long-run growth of the monetary and credit aggregates commensurate with the economy's long-run potential to increase production, so as to promote effectively the goals of maximum employment, stable prices, and moderate long-term interest rates.*

b. In your opinion, are these two excerpts from the act consistent with the position of the United States in Figure 21-3?

> *Section 2B. Appearances Before and Reports to the Congress*
>
> ***Appearances Before The Congress***
>
> *In General. The Chairman of the Board shall appear before the Congress at semi-annual hearings, as specified in paragraph (2), regarding—A. the efforts, activities, objectives and plans of the Board and the Federal Open Market Committee with respect to the conduct of monetary policy; and B. economic developments and prospects for the future described in the report required in subsection (b).*
>
> ***Section 10. Board of Governors of the Federal Reserve System***
>
> *1. Appointment and qualification of members*
>
> *The Board of Governors of the Federal Reserve System (hereinafter referred to as the "Board") shall be composed of seven members, to be appointed by the President, by and with the advice and consent of the Senate, after the date of enactment of the Banking Act of 1935, for terms of fourteen years.*

FURTHER READINGS

- For more model comparisons, you can look at Günter Coenen et al., "Effects of Fiscal Stimulus in Structural Models," *American Economic Journal: Macroeconomics*, 2012, 4(1): pp. 22–68.
- If you want to learn more about political economy issues, a useful reference is *Political Economy in Macroeconomics* by Allan Drazen (2002, Princeton University Press).
- For an argument that inflation decreased as a result of the increased independence of central banks in the 1990s, read "Central Bank Independence and Inflation" in the 2009 Annual Report of the Federal Reserve Bank of St. Louis at www.stlouisfed.org/annual-report/2009/central-bank-independence-and-inflation.
- A leading proponent of the view that governments misbehave and should be tightly restrained was James Buchanan, of George Mason University. He received the Nobel Prize in 1986 for his work on public choice. Read, for example, his book with Richard Wagner, *Democracy in Deficit: The Political Legacy of Lord Keynes* (1977, Liberty Fund).
- For an interpretation of the increase in inflation in the 1970s as the result of time inconsistency, see "Did Time Consistency Contribute to the Great Inflation?" by Henry Chappell and Rob McGregor, *Economics & Politics*, 2004, Vol. 16, No. 3, pp. 233–251.

Fiscal Policy: A Summing Up

22

The financial crisis and the Great Recession led to large budget deficits, and these have resulted in a large increase in debt-to-GDP ratios. The debt-to-GDP ratio, which stood at 73% in advanced countries in 2007, was 104% in 2018. In some countries, it was much higher: 130% in Italy, 188% in Greece, 240% in Japan. A major macroeconomic question confronting governments is whether these ratios should be reduced and, if so, at what speed.

The purpose of this chapter is to review what we have learned about fiscal policy so far, to explore in more depth the dynamics of deficits and debt, to discuss the costs of high debt, and to give a tentative answer to the question in the opening paragraph.

Section 22-1 takes stock of what we have learned about fiscal policy so far.

Section 22-2 looks more closely at the government budget constraint and examines its implications for the relation between budget deficits, the interest rate, the growth rate, and government debt.

Section 22-3 takes up three issues for which the government budget constraint plays a central role, from the proposition that deficits do not really matter, to how to run fiscal policy in the cycle, to whether to finance wars through taxes or through debt.

Section 22-4 discusses the dangers associated with high government debt, from higher taxes to higher interest rates, default, and high inflation.

Section 22-5 looks at the challenges facing US fiscal policy today.

If you remember one basic message from this chapter, it should be: Fiscal policy can be a strong macroeconomic policy tool. When using it, one must think about both its short-, medium-, and long-run effects. ▶▶▶

22-1 WHAT WE HAVE LEARNED

Let's review what we have learned so far about fiscal policy:

- In Chapter 3, we looked at how government spending and taxes affected demand and, in turn, output in the short run.

  We saw that, in the short run, a fiscal expansion—increases in government spending or decreases in taxes—increases output.
- In Chapter 5, we looked at the short-run effects of fiscal policy on output and on the interest rate.

  We saw that a fiscal contraction leads to lower disposable income, which causes people to decrease their consumption. This decrease in demand leads, through a multiplier, to a decrease in output and income. At a given policy rate, the fiscal contraction leads therefore to a decrease in output. A decrease in the policy rate by the central bank can, however, potentially offset the adverse effects of the fiscal contraction.
- In Chapter 6, we saw how fiscal policy was used during the financial crisis to limit the fall in output.

  We saw that when the economy is in a liquidity trap, a reduction in the interest rate can no longer be used to increase output, and thus fiscal policy has an important role to play. Large increases in spending and cuts in taxes, however, were not enough to avoid the recession.
- In Chapter 9, we looked at the effects of fiscal policy in the short run and in the medium run.

  We saw that, in the medium run (that is, taking the capital stock as given), a fiscal consolidation has no effect on output but is reflected in a different composition of spending. In the short run, however, output decreases. In other words, if output was at potential to start with, the fiscal consolidation, as desirable as it may be on other grounds, initially leads to a recession.
- In Chapter 11, we looked at how saving, both private and public, affects the level of capital accumulation and the level of output in the long run.

  We saw that, once capital accumulation is taken into account, a larger budget deficit, and by implication a lower national saving rate, decreases capital accumulation, leading to a lower level of output in the long run.
- In Chapter 16, we returned to the short-run effects of fiscal policy, considering not only fiscal policy's direct effects through taxes and government spending, but also its effects on expectations.

  We saw that the effects of fiscal policy depend on expectations of future fiscal and monetary policy. We saw that a deficit reduction may, in some circumstances, lead to an increase in output even in the short run, thanks to people's expectations of higher future disposable income.
- In Chapter 18, we looked at the effects of fiscal policy when the economy is open in the goods market.

  We saw that fiscal policy affects both output and the trade balance, and we examined the relation between the budget deficit and the trade deficit.
- In Chapter 19, we looked at the role of fiscal policy in an economy open in both goods markets and financial markets.

  We saw that the effects of fiscal policy depend on the exchange rate regime. Fiscal policy is likely to have a stronger effect on output under fixed exchange rates than under flexible exchange rates.
- In Chapter 21, we looked at the problems facing policymakers in general, from uncertainty about the effects of policy to issues of time consistency and credibility.

We looked at the pros and cons of putting restraints on the conduct of fiscal policy, from spending caps to a constitutional amendment to balance the budget.

In deriving these conclusions, we did not pay close attention to the government budget constraint—that is, the relation among debt, deficits, spending, and taxes over time. This relation is important, however, in understanding both how we got to where we are today and the choices faced by policymakers. It is the focus of the next section.

22-2 THE GOVERNMENT BUDGET CONSTRAINT: DEFICITS, DEBT, SPENDING, AND TAXES

Suppose that, starting from a balanced budget, the government decreases taxes, creating a budget deficit. What will happen to the debt over time? Will the government need to increase taxes later? If so, by how much?

The Arithmetic of Deficits and Debt

To answer these questions, we must begin with a definition of the budget deficit. We can write the budget deficit in year t as:

$$\text{deficit}_t = rB_{t-1} + G_t - T_t \tag{22.1}$$

All variables are in real terms:

- B_{t-1} is government debt at the end of year $t - 1$, or, equivalently, at the beginning of year t; r is the real interest rate, which we shall for now assume to be constant. Thus, rB_{t-1} equals the real interest payments on the government debt in year t.
- G_t is government spending on goods and services during year t.
- T_t is taxes minus transfers during year t.

In words: The budget deficit equals spending, including interest payments on the debt, minus taxes net of transfers.

Note two characteristics of equation (22.1):

- We measure interest payments as real interest payments—that is, the product of the *real* interest rate times existing debt—rather than as actual interest payments—that is, the product of the nominal interest rate and the existing debt. As the Focus Box "Inflation Accounting and the Measurement of Deficits" shows, this is the correct way of measuring interest payments. Official measures of the deficit, however, use actual (nominal) interest payments and are therefore incorrect. When inflation is high, official measures can be seriously misleading. The correct measure of the deficit is sometimes called the **inflation-adjusted deficit**.
- For consistency with our definition of G as spending on goods and services, G does not include transfer payments. Transfers are instead subtracted from T, so that T stands for *taxes minus transfers*. Official measures of government spending typically add transfers to spending on goods and services and define revenues as taxes, not taxes net of transfers.
- These are only accounting conventions. Whether transfers are added to spending or subtracted from taxes makes a difference to the measurement of G and T, but clearly does not affect $G - T$, and therefore does not affect the measure of the deficit.

The **government budget constraint** then simply states that the *change in government debt during year t* is equal to the deficit during year t:

$$B_t - B_{t-1} = \text{deficit}_t$$

◀ Do not confuse the words *deficit* and *debt*. (Many journalists and politicians do.) Debt is a stock—what the government owes as a result of past deficits. The deficit is a flow—how much the government borrows during a given year.

◀ Transfer payments are government transfers to individuals, such as unemployment benefits or Medicare.

◀ Let G represent government spending on goods and services; Tr, transfers; and Tax, total taxes. For simplicity, assume interest payments rB equal zero. Then

$$\text{Deficit} = G + Tr - Tax$$

This can be rewritten in two (equivalent) ways:

$$\text{Deficit} = G - (Tax - Tr)$$

The deficit equals spending on goods and services minus net taxes—that is, taxes minus transfers. This is the way we write it in the text. Or it can be written as:

$$\text{Deficit} = (G + Tr) - Tax$$

which is the way it is decomposed in official measures (see, for example, Table A1-4 in Appendix 1 at the end of the book).

FOCUS

Inflation Accounting and the Measurement of Deficits

Official measures of the budget deficit are constructed (dropping the time indexes, which are not needed here) as nominal interest payments, iB, plus spending on goods and services, G, minus taxes net of transfers, T:

$$\text{official measure of the deficit} = iB + G - T$$

This is an accurate measure of the cash flow position of the government. If it is positive, the government is spending more than it receives and must therefore issue new debt. If it is negative, the government buys back previously issued debt.

But this is not an accurate measure of the *change in real debt*—that is, the change in how much the government owes, expressed in terms of goods rather than dollars.

To see why, consider the following example: Suppose the official measure of the deficit is equal to zero, so the government neither issues nor buys back debt. Suppose inflation is positive and equal to 10%. Then, at the end of the year, the real value of the debt has decreased by 10%. If we define—as we should—the deficit as the change in the real value of government debt, the government has decreased its real debt by 10% over the year. In other words, it has in fact run a budget surplus equal to 10% times the initial level of debt.

More generally: If B is debt and π is inflation, the official measure of the deficit overstates the correct measure by an amount equal to πB. Put another way, the correct measure of the deficit is obtained by subtracting πB from the official measure:

$$\begin{aligned}\text{correct measure of the deficit} &= iB + G - T - \pi B \\ &= (i - \pi)B + G - T \\ &= rB + G - T\end{aligned}$$

where $r = i - \pi$ is the (realized) real interest rate. The correct measure of the deficit is equal to real interest payments plus government spending minus taxes net of transfers.

The difference between the official and the correct measures of the deficit equals πB. The higher the rate of inflation, π, or the higher the level of debt, B, the more inaccurate the official measure. In countries where both inflation and debt are high, the official measure may record a very large budget deficit, when in fact real government debt is decreasing. This is why you should always do the inflation adjustment before deriving conclusions about the position of fiscal policy.

Figure 1 plots the official measure and the inflation-adjusted measure of the (federal) budget deficit for the United States as a percent of GDP since 1969. The official measure shows a deficit in every year from 1970 to 1997. The inflation-adjusted measure shows instead alternating deficits and surpluses until the late 1970s. Both measures, however, show that the deficit became much larger after 1980, that things improved in the 1990s, and that they have deteriorated in the 2000s. Today, with inflation running at about 2% a year and the ratio of debt-to-GDP around 80%, the difference between the two measures is roughly equal to 2% times 80%, or 1.6% of GDP.

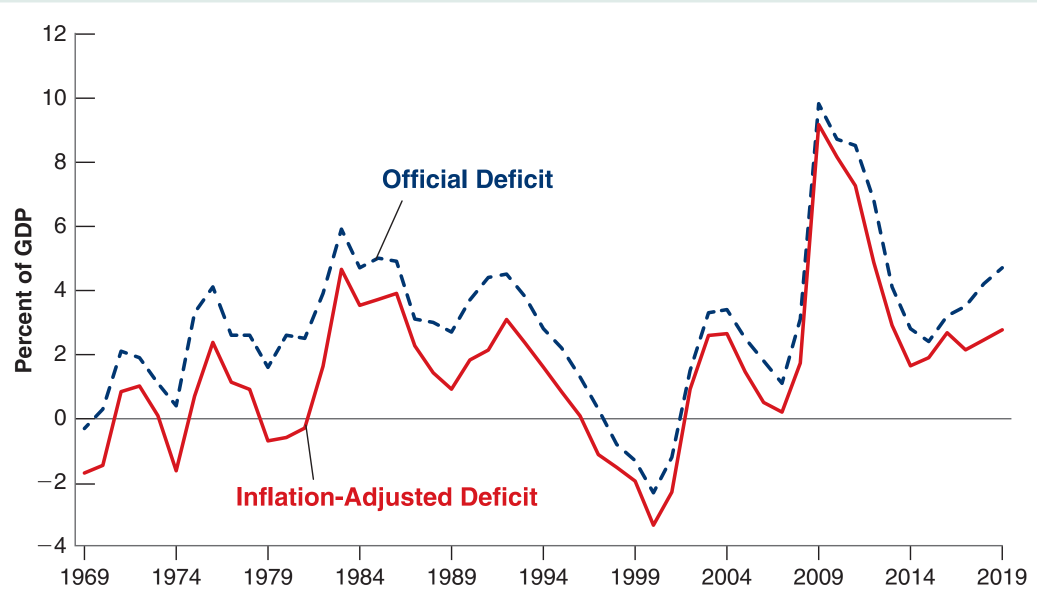


Figure 1

Official and Inflation-Adjusted Federal Budget Deficits for the United States since 1969

Source: FRED: CPIAUCSL, and Tables B-18 and B-19 *Economic Report of the President.*

If the government runs a deficit, government debt increases as the government borrows to fund the part of spending in excess of revenues. If the government runs a surplus, government debt decreases as the government uses the budget surplus to repay part of its outstanding debt.

Using the definition of the deficit (equation (22.1)), we can rewrite the government budget constraint as

$$B_t - B_{t-1} = rB_{t-1} + G_t - T_t \tag{22.2}$$

The government budget constraint links the change in government debt to the initial level of debt (which affects interest payments) and to current government spending and taxes. It is often convenient to decompose the deficit into the sum of two terms:

- Interest payments on the debt, rB_{t-1}.
- The difference between spending and taxes, $G_t - T_t$. This term is called the **primary deficit** (equivalently, $T_t - G_t$ is called the **primary surplus**).

Using this decomposition, we can rewrite equation (22.2) as

$$\overbrace{B_t - B_{t-1}}^{\text{Change in the Debt}} = \overbrace{rB_{t-1}}^{\text{Interest Payments}} + \overbrace{(G_t - T_t)}^{\text{Primary Deficit}}$$

Or, moving B_{t-1} to the right side of the equation and reorganizing,

$$B_t = (1 + r)B_{t-1} + \overbrace{(G_t - T_t)}^{\text{Primary Deficit}} \tag{22.3}$$

This relation states that the debt at the end of year t equals $(1 + r)$ times the debt at the end of year $t - 1$ plus the primary deficit during year t, $(G_t - T_t)$. Let's look at some of its implications. (In this section, I shall assume that the real interest rate is positive, so that, if the primary deficit is equal to zero, debt increases over time. This is not necessarily the case, and I shall return to the assumption later on.)

Current versus Future Taxes

Consider first a one-year decrease in taxes for the path of debt and future taxes. Start from a situation where, until year 1, the government has balanced its budget, so that initial debt is equal to zero. During year 1, the government decreases taxes by 1 (think 1 billion dollars, for example) for one year. Thus, debt at the end of year 1, B_1, is equal to 1. The question we take up: What happens thereafter?

Full Repayment in Year 2

Suppose the government decides to fully repay the debt during year 2. From equation (22.3), the budget constraint for year 2 is given by

$$B_2 = (1 + r)B_1 + (G_2 - T_2)$$

If the debt is fully repaid during year 2, then the debt at the end of year 2 is equal to zero, $B_2 = 0$. Replacing B_1 by 1 and B_2 by 0 and transposing terms gives

$$T_2 - G_2 = (1 + r)1 = (1 + r)$$

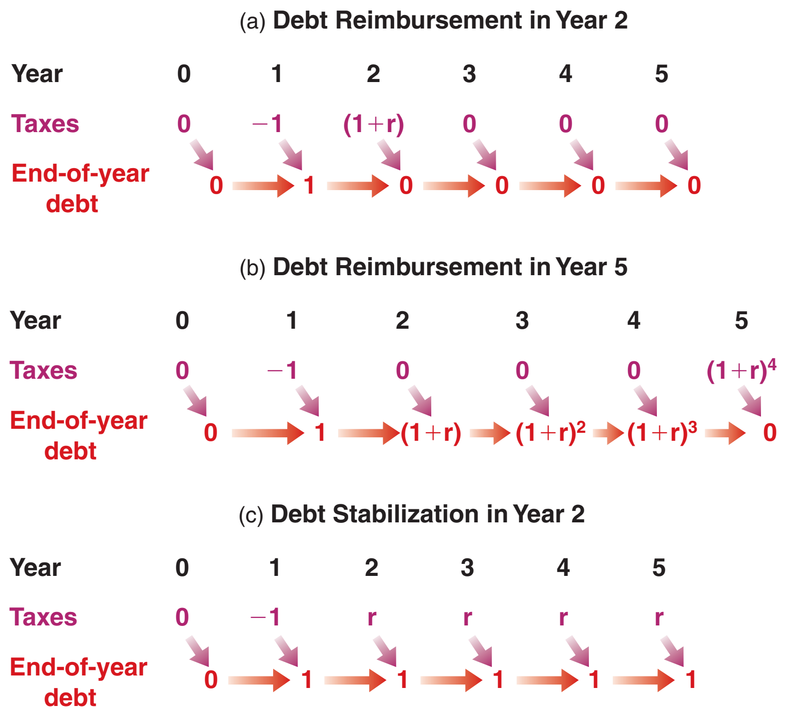
To repay the debt fully during year 2, the government must run a primary surplus equal to $(1 + r)$. It can do so in one of two ways: a decrease in spending or an increase in taxes. I shall assume here and in the rest of this section that the adjustment comes through taxes, so that the path of spending is unaffected. It follows that the decrease in taxes by 1 during year 1 must be offset by an increase in taxes by $(1 + r)$ during year 2. ◄

Full repayment in year 2: T_1 decreases by 1 $\Rightarrow$ T_2 increases by $(1 + r)$

Figure 22-1

Tax Cuts, Debt Repayment, and Debt Stabilization

(a) If the debt is fully repaid during year 2, the decrease in taxes of 1 in year 1 requires an increase in taxes equal to $(1 + r)$ in year 2. (b) If the debt is fully repaid during year 5, the decrease in taxes of 1 in year 1 requires an increase in taxes equal to $(1 + r)^4$ during year 5. (c) If the debt is stabilized from year 2 on, then taxes must be permanently higher by r from year 2 on.



The path of taxes and debt corresponding to this case is given in Figure 22-1(a): If the debt is fully repaid during year 2, the decrease in taxes of 1 in year 1 requires an increase in taxes equal to $(1 + r)$ in year 2.

Full Repayment in Year *t*

Now suppose the government decides to wait until year t to repay the debt. From year 2 to year $t - 1$ the primary deficit is equal to zero; taxes are equal to spending, not including interest payments on the debt.

During year 2, the primary deficit is zero. So, from equation (22.3), debt at the end of year 2 is:

$$B_2 = (1 + r)B_1 + 0 = (1 + r)1 = (1 + r)$$

where the second equality uses the fact that $B_1 = 1$.

With the primary deficit still equal to zero during year 3, debt at the end of year 3 is

$$B_3 = (1 + r)B_2 + 0 = (1 + r)(1 + r)1 = (1 + r)^2$$

Solving for debt at the end of year 4 and so on, it is clear that *as long as the government keeps a primary deficit equal to zero*, debt grows at a rate equal to the interest rate, and thus debt at the end of year $t - 1$ is given by

$$B_{t-1} = (1 + r)^{t-2} \tag{22.4}$$

Even though taxes are cut only in year 1, debt keeps increasing over time, at a rate equal to the interest rate. The reason is simple; although the primary deficit is equal to zero, debt is now positive, and so are interest payments on it. Each year, the government must issue more debt to pay the interest on existing debt.

In year t, the year in which the government decides to repay the debt, the budget constraint is

$$B_t = (1 + r)B_{t-1} + (G_t - T_t)$$

If debt is fully repaid during year t, then B_t (debt at the end of year t) is zero. Replacing B_t by zero and B_{t-1} by its expression from equation (22.4) gives

$$0 = (1 + r)(1 + r)^{t-2} + (G_t - T_t)$$

Add exponents: $(1 + r)(1 + r)^{t-2} = (1 + r)^{t-1}$. See Appendix 2 at the end of this book.

Reorganizing and bringing $(G_t - T_t)$ to the left side of the equation implies

$$T_t - G_t = (1 + r)^{t-1}$$

To repay the debt, the government must run a primary surplus equal to $(1 + r)^{t-1}$ during year t. If the adjustment is done through taxes, the initial decrease in taxes of 1 during year 1 leads to an increase in taxes of $(1 + r)^{t-1}$ during year t. The path of taxes and debt corresponding to the case where debt is repaid in year 5 is given in Figure 22-1(b).

Full repayment in year 5: T_1 decreases by 1 ⇒ T_5 increases by $(1 + r)^4$

This example yields our first set of conclusions:

- If government spending is unchanged, a decrease in taxes must eventually be offset by an increase in taxes in the future.
- The longer the government waits to increase taxes, or the higher the real interest rate is, the higher the eventual increase in taxes must be.

Debt Stabilization in Year 2

We have assumed so far that the government fully repays the debt. Let's now look at what happens to taxes if the government only stabilizes the debt. (Stabilizing the debt means changing taxes or spending so that debt remains constant from then on.)

Suppose the government decides to stabilize the debt from year 2 on: the debt at the end of year 2 and thereafter remains at the same level as at the end of year 1.

From equation (22.3), the budget constraint for year 2 is

$$B_2 = (1 + r)B_1 + (G_2 - T_2)$$

Under our assumption that debt is stabilized in year 2, $B_2 = B_1 = 1$. Setting $B_2 = B_1 = 1$ in the preceding equation yields

$$1 = (1 + r) + (G_2 - T_2)$$

Reorganizing and bringing $(G_2 - T_2)$ to the left side of the equation,

$$T_2 - G_2 = (1 + r) - 1 = r$$

To avoid a further increase in debt during year 2, the government must run a primary surplus equal to real interest payments on the existing debt (recall our assumption that the real interest rate is positive). It must do so in each of the following years as well. Each year, the primary surplus must be sufficient to cover interest payments, leaving the debt level unchanged. The path of taxes and debt is shown in Figure 22-1(c). Debt remains equal to 1 from year 1 on. Taxes are permanently higher from year 2 on, by an amount equal to r; equivalently, from year 2 on, the government runs a primary surplus equal to r.

Stabilizing the debt from year 2 on: T_1 decreases by 1 ⇒ $T_2, T_3, \ldots$ increase by r

The logic of this argument extends directly to the case where the government waits until year t to stabilize the debt. Whenever the government stabilizes, it must, each year from then on, run a primary surplus sufficient to pay the interest on the debt.

This example yields our second set of conclusions:

- The legacy of past deficits is higher government debt today.
- To stabilize the debt, the government must eliminate the deficit.
- To eliminate the deficit, and if the interest rate is positive, the government must run a primary surplus equal to the interest payments on the existing debt. This requires higher taxes forever.

The Evolution of the Debt-to-GDP Ratio

We have focused so far on the evolution of the *level* of debt. But in an economy in which output grows over time, it makes more sense to focus instead on the *ratio of debt to output.*

To see how this change in focus modifies our conclusions, we need to go from equation (22.3) to an equation that gives the evolution of the **debt-to-GDP ratio**—the **debt ratio** for short.

Deriving the evolution of the debt ratio takes a few steps. Do not worry; the final equation is easy to understand.

First divide both sides of equation (22.3) by real output, Y_t, to get

$$\frac{B_t}{Y_t} = (1 + r)\frac{B_{t-1}}{Y_t} + \frac{G_t - T_t}{Y_t}$$

Next rewrite B_{t-1}/Y_t as $(B_{t-1}/Y_{t-1})\,(Y_{t-1}/Y_t)$ (in other words, multiply the numerator and the denominator by Y_{t-1}):

Start from $Y_t = (1 + g)Y_{t-1}$. Divide both sides by Y_t to get $1 = (1 + g)Y_{t-1}/Y_t$. Reorganize to get $Y_{t-1}/Y_t = 1/(1 + g)$. This approximation is derived as proposition 6 in Appendix 2 at the end of this book.

$$\frac{B_t}{Y_t} = (1 + r)\left(\frac{Y_{t-1}}{Y_t}\right)\frac{B_{t-1}}{Y_{t-1}} + \frac{G_t - T_t}{Y_t}$$

Note that all the terms in the equation are now in terms of ratios to output, Y. To simplify this equation, assume that output growth is constant and denote the growth rate of output by g, so Y_{t-1}/Y_t can be written as $1/(1 + g)$. And use the approximation $(1 + r)/(1 + g) \approx 1 + r - g$.

Using these two assumptions, rewrite the preceding equation as

$$\frac{B_t}{Y_t} = (1 + r - g)\frac{B_{t-1}}{Y_{t-1}} + \frac{G_t - T_t}{Y_t}$$

Finally, reorganize to get

$$\frac{B_t}{Y_t} - \frac{B_{t-1}}{Y_{t-1}} = (r - g)\frac{B_{t-1}}{Y_{t-1}} + \frac{G_t - T_t}{Y_t} \qquad (22.5)$$

This took many steps, but the final relation has a simple interpretation.
The change in the debt ratio over time (the left side of the equation) is equal to the sum of two terms:

- The first term is the difference between the real interest rate and the growth rate, times the initial debt ratio.
- The second term is the ratio of the primary deficit to GDP.

If two variables (here debt and GDP) grow at rates r and g respectively, then their ratio (here the ratio of debt to GDP) will grow at rate $r - g$. See proposition 8 in Appendix 2 at the end of this book.

Compare equation (22.5), which gives the evolution of the ratio of debt to GDP, to equation (22.2), which gives the evolution of the level of debt itself. Assume that the primary deficit (and therefore the ratio of the primary deficit to GDP is equal to zero. Then, whether *debt* increases or decreases depends on whether the interest rate is positive or not. However, whether the *debt-to-GDP* ratio increases or decreases depends on whether the interest rate is larger or smaller than the growth rate. (More on this in Section 22-5.)

Equation (22.5) implies that the increase in the ratio of debt to GDP will be larger:

- the higher the real interest rate,
- the lower the growth rate of output,
- the higher the initial debt ratio,
- the higher the ratio of the primary deficit to GDP.

Building on this relation, the Focus Box "How Countries Decreased Their Debt Ratios after World War II" shows how governments that inherited high debt ratios at the end of the war steadily decreased them through a combination of low real interest rates, high growth rates, and primary surpluses. The next section shows how our analysis can also be used to shed light on a number of other fiscal policy issues.

FOCUS

How Countries Decreased Their Debt Ratios after World War II

After World War II, many countries had high debt ratios, often in excess of 100% of GDP. Yet, two or three decades later, the debt ratios were much lower, often below 50%. (We saw in Chapter 21 that this was the case for the United States.) How did they do it? The answer is given in Table 1.

Table 1 looks at four countries: Australia, Canada, New Zealand, and the United Kingdom. Column 1 gives the period during which debt ratios decreased. The first year is either 1945 or 1946. The last year is the year in which the debt ratio reached its lowest point; the period of adjustment varies from 13 years in Canada to 30 years in the United Kingdom. Column 2 gives debt ratios at the start and at the end of the period. The most striking numbers are those for the United Kingdom: an initial debt ratio of 270% of GDP in 1946 and an impressive decline to 47% in 1974.

To interpret the numbers in the table, go back to equation (22.5). It tells us that there are two, not mutually exclusive, ways in which a country can reduce its debt ratio. The first is through high primary surpluses. Suppose, for example, that $(r - g)$ was equal to 0. Then the decrease in the debt ratio over some period would just be the sum of the ratios of primary surpluses to GDP over the period. The second is through a low $(r - g)$, through either low real interest rates or high growth, or both.

With this in mind, columns 3 to 5 give the average ratio of the primary balance to GDP, average growth rate of GDP, and average real interest rate over the relevant period.

Look first at primary balances in column 3. All four countries ran primary surpluses on average over the period. But note that these primary surpluses account only for a small part of the decline in the debt ratio. Look, for example, at the United Kingdom. The sum of the ratios of the primary surpluses to GDP over the period is equal to 2.1% multiplied by $30 = 63\%$ of GDP, accounting for less than a third of the decline in the debt ratio, 223% $(270\% - 47\%)$ of GDP.

Now look at the growth rates and real interest rates in columns 4 and 5. Note how high growth rates and how low real interest rates were during the period. Take Australia, for example. The average value of $(r - g)$ during the period was -6.9% $(-2.3\% - 4.6\%)$. This implies that, even if the primary balance had been equal to zero, the debt ratio would have declined each year by 6.9%. In other words, the decline in debt was mainly the result not of primary surpluses but of sustained high growth and sustained negative real interest rates.

This leads to a final question: Why were real interest rates so low? The answer is given in column 6. During the period, average inflation was relatively high. This inflation, combined with consistently low nominal interest rates, is what accounts for the negative real interest rates. Put another way, a large part of the decrease in debt ratios was achieved by paying bond holders a negative real return on their bonds for many years.

Table 1 Changes in Debt Ratios Following World War II

	1	2	3	4	5	6
Country	Start/End Year	Start/End Debt Ratio	Primary Balance	Growth Rate	Real Interest Rate	Inflation Rate
Australia	1946–1963	92–29	1.1	4.6	− 2.3	5.7
Canada	1945–1957	115–59	3.6	4.3	− 1.4	4.0
New Zealand	1946–1974	148–41	2.3	3.9	− 2.9	4.9
United Kingdom	1946–1975	270–47	2.1	2.6	− 1.5	5.5

Columns 2 and 3: Percent of GDP. Columns 4 to 6: Percent.

Source: S. M. A. Abbas et al., "Historical Patterns and Dynamics of Public Debt: Evidence from a New Database," *IMF Economic Review*, 2011 (November): pp. 717–742.

22-3 RICARDIAN EQUIVALENCE, CYCLICAL ADJUSTED DEFICITS, AND WAR FINANCE

Having looked at the mechanics of the government budget constraint, we take up three issues in which this constraint plays a central role.

Ricardian Equivalence

How does taking into account the government budget constraint affect the way we should think of the effects of deficits on output?

Although Ricardo stated the logic of the argument, he also argued that there were many reasons why it would not hold in practice. In contrast, Barro argued that the argument was not only logically correct but also a good description of reality.

One extreme view is that once the government budget constraint is taken into account, neither deficits nor debt have an effect on economic activity! This argument is known as the **Ricardian equivalence** proposition. David Ricardo, a 19th-century English economist, was the first to articulate its logic. His argument was further developed and given prominence in the 1970s by Robert Barro, then at Chicago, now at Harvard University. For this reason, the argument is also known as the **Ricardo-Barro proposition**.

The best way to understand the logic of the proposition is to use the example of tax changes from Section 22-2:

- Suppose that the government decreases taxes by 1 (again, think 1 billion dollars) this year and announces that, to repay the debt, it will increase taxes by $(1 + r)$ next year. What will be the effect of the initial tax cut on consumption?
- One possible answer is: No effect at all. Why? Because consumers realize that the tax cut is not much of a gift. Lower taxes this year are exactly offset, in present value, by higher taxes next year. Put another way, their human wealth —the present value of after-tax labor income—is unaffected. Current taxes go down by 1, but the present value of next year's taxes goes up by $(1 + r)/(1 + r) = 1$, and the net effect of the two changes is exactly equal to zero.
- Another way of coming to the same answer—this time looking at saving rather than consumption—is as follows: To say that consumers do not change their consumption in response to the tax cut is the same as saying that *private saving increases one-for-one with the deficit.* The Ricardian equivalence proposition says that if a government finances a given path of spending through deficits, private saving will increase one-for-one with the decrease in public saving, leaving total saving unchanged. The total amount left for investment will not be affected. Over time, the mechanics of the government budget constraint imply that government debt will increase. But this increase will not come at the expense of capital accumulation.

See Chapter 15 for a definition of human wealth and a discussion of its role in consumption.

Go back to the IS-LM model. If Ricardian equivalence holds, what is the multiplier associated with a decrease in current taxes?

Under the Ricardian equivalence proposition, a long sequence of deficits and the associated increase in government debt are no cause for worry. As the government is dissaving, the argument goes, people are saving more in anticipation of the higher taxes to come. The decrease in public saving is offset by an equal increase in private saving. Total saving is therefore unaffected, and so is investment. The economy has the same capital stock today that it would have had if there had been no increase in debt. Therefore, high debt is no cause for concern.

How seriously should we take the Ricardian equivalence proposition? Most economists would answer: "Seriously, but surely not seriously enough to think that deficits and debt are irrelevant." A major theme of this book has been that expectations matter, that consumption decisions depend not only on current income but also on future income. If it were widely believed that a tax cut this year will be followed by an offsetting increase in taxes *next year*, the effect on consumption would indeed probably be small. Many consumers would save most or all of the tax cut in anticipation of higher taxes next year. (Replace *year* by *month* or *week* and the argument becomes even more convincing.)

Recall that this assumes that government spending is unchanged. If people expect government spending to be decreased in the future, what will they do?

Of course, tax cuts rarely come with the announcement of corresponding tax increases a year later. Consumers must guess when and how taxes will eventually be increased. This fact does not by itself invalidate the Ricardian equivalence argument. No matter when taxes will be increased, the government budget constraint still implies that the present value of future tax increases must always be equal to the decrease in taxes today.

Take the second example we looked at in Section 22-2, drawn in Figure 22-1(b) in which the government waits t years to increase taxes and so increases taxes by $(1 + r)^{t-1}$. The present value in year 0 of this expected tax increase is $(1 + r)^{t-1}/(1 + r)^{t-1} = 1$—exactly equal to the original tax cut. The change in human wealth from the tax cut is still zero.

The increase in taxes in t years is $(1 + r)^{t-1}$. The discount factor for a dollar t years from now is $1/(1 + r)^{t-1}$. So the value of the increase in taxes t years from now as of today is $(1 + r)^{t-1}/(1 + r)^{t-1} = 1$.

But insofar as future tax increases appear more distant and their timing more uncertain, consumers are in fact more likely to ignore them. This may be because they expect to die before taxes go up or, more likely, because they just do not think that far into the future. In either case, Ricardian equivalence is likely to fail.

So it is safe to conclude that budget deficits have an important effect on activity, although perhaps a smaller effect than you thought before going through the Ricardian equivalence argument. In the short run, larger deficits are likely to lead to higher demand and higher output. In the long run, higher government debt lowers capital accumulation and, as a result, lowers output.

Deficits, Output Stabilization, and the Cyclically Adjusted Deficit

The fact that budget deficits do, indeed, have long-run adverse effects on capital accumulation, and in turn adverse effects on output, does not imply that fiscal policy should not be used to reduce output fluctuations. Rather, it implies that deficits during recessions should be offset by surpluses during booms, so as not to lead to a steady increase in debt.

Note the analogy with monetary policy: The fact that higher money growth leads in the long run to more inflation does not imply that monetary policy should not be used for output stabilization.

To help assess whether fiscal policy is on track, economists have constructed deficit measures that tell them what the deficit would be, under existing tax and spending rules, if output were equal to potential output. Such measures come under many names—the **full-employment deficit**, the **midcycle deficit**, the **standardized employment deficit**, and the **structural deficit** (the term used by the OECD). I shall use **cyclically adjusted deficit**, the term I find the most intuitive.

Such a measure gives a simple benchmark against which to judge the direction of fiscal policy. If the actual deficit is large but the cyclically adjusted deficit is zero, then current fiscal policy is consistent with no systematic increase in debt over time. The debt will increase as long as output is below the potential level of output; but as output returns to potential, the deficit will disappear and the debt will stabilize.

It does not follow that the goal of fiscal policy should be to maintain a cyclically adjusted deficit equal to zero at all times. In a recession, the government may want to run a deficit large enough that even the cyclically adjusted deficit is positive. In this case, the fact that the cyclically adjusted deficit is positive provides a useful warning: that the return of output to potential will not be enough to stabilize the debt. The government will have to take specific measures, from tax increases to cuts in spending, to decrease the deficit at some point in the future.

The theory underlying the concept of the cyclically adjusted deficit is simple. The practice of it has proven tricky. To see why, we need to look at how measures of the cyclically adjusted deficit are constructed. Construction requires two steps: First, establish how much lower the deficit would be if output were, say, 1% higher. Second, assess how far output is from potential.

- The first step is relatively straightforward. For the United States, a reliable rule of thumb is that a 1% decrease in output leads automatically to an increase in the deficit of about 0.5% of GDP. This increase occurs because most taxes are proportional to output, whereas most government spending does not depend on the level of output. That means a decrease in output, which leads to a decrease in revenues and not much change in spending, naturally leads to a larger deficit.

  If output is, say, 5% below potential, the deficit as a ratio to GDP will be about 2.5% larger than if output were at potential. (This effect of activity on the deficit has been called an **automatic stabilizer**. A recession naturally generates a deficit, and therefore a fiscal expansion, which partly counteracts the recession.)
- The second step is more difficult. Recall from Chapter 7 that potential output is the output level that would be produced if the economy were operating at the natural rate of unemployment. Too low an estimate of the natural rate of unemployment will lead to too high an estimate of potential output and therefore too optimistic a measure of the cyclically adjusted deficit.

This difficulty explains in part what happened to debt in Europe in the 1980s. Based on the assumption of an unchanged natural unemployment rate, the cyclically adjusted deficits of the 1980s did not look that bad. If European unemployment had returned to its level of the 1970s, the associated increase in output would have been enough to reestablish budget balance in most countries. But it turned out that much of the increase in unemployment reflected an increase in the natural unemployment rate, and unemployment remained high during the 1980s. As a result, the decade was characterized by high deficits and large increases in debt ratios in most European countries.

Wars and Deficits

Look at the two peaks associated with World War I and World War II in Figure 21–4. ▶

Wars typically bring about large budget deficits. As we saw in Chapter 21, the two largest increases in US government debt in the 20th century took place during World War I and World War II. We examine the case of World War II further in the Focus Box "Deficits, Consumption, and Investment in the United States during World War II."

Is it right for governments to rely so much on deficits to finance wars? After all, war economies are usually operating at low unemployment, so output stabilization reasons for running deficits are irrelevant. The answer, nevertheless, is yes. In fact, there are two good reasons to run deficits during wars:

- The first is distributional: Deficit finance is a way to pass some of the burden of the war to those alive after the war. It seems only fair for future generations to share in the sacrifices the war requires.
- The second is more narrowly economic: Deficit spending helps reduce tax distortions. Let's look at each reason in turn.

Passing on The Burden of the War

Wars lead to large increases in government spending. Consider the implications of financing this increased spending through either increased taxes or debt finance. To distinguish this case from our previous discussion of output stabilization, let's also assume that output is and remains at its potential level.

Deficits, Consumption, and Investment in the United States during World War II

FOCUS

In 1939, the share of US government spending on goods and services in GDP was 15%. By 1944, it had increased to 45%! The increase was due to increased spending on national defense, which went from 1% of GDP in 1939 to 36% in 1944.

Faced with such a massive increase in spending, the US government reacted with large tax increases. For the first time in US history, the individual income tax became a major source of revenues: individual income tax revenues, which were 1% of GDP in 1939, increased to 8.5% in 1944. But the tax increases were still far less than the increase in government expenditures. The increase in federal revenues, from 7.2% of GDP in 1939 to 22.7% in 1944, was only a little more than half the increase in expenditures.

The result was a sequence of large budget deficits. By 1944, the federal deficit reached 22% of GDP. The ratio of debt to GDP, already high at 53% in 1939 because of the deficits the government had run during the Great Depression, reached 110%!

Was the increase in government spending achieved at the expense of consumption or private investment? (As we saw in Chapter 18, it could in principle have come from higher imports and a current account deficit. But the United States had nobody to borrow from during the war. Rather, it was lending to some of its allies. Loans and gifts from the US government to foreign countries were equal to 6% of US GDP in 1944.)

- It was met in large part by a decrease in consumption. The share of consumption in GDP fell by 23 percentage points, from 74% to 51%. Part of the decrease in consumption may have been due to anticipations of higher taxes after the war; part of it was due to the unavailability of many consumer durables. Patriotism also probably motivated people to save more and buy the war bonds issued by the government to finance the war.
- It was also met by a 6% decrease in the share of (private) investment in GDP—from 10% to 4%. Part of the burden of the war was therefore passed on, in the form of lower capital accumulation, to those living after the war.

- Suppose that the government relies on deficit finance. With government spending sharply up, there will be a large increase in the demand for goods. Given our assumption that output stays the same, the interest rate will have to increase enough to maintain equilibrium. Investment, which depends on the interest rate, will decrease sharply. (More realistically, in a war economy, the government may use direct measures to decrease non–war-related investment, without resorting to a high interest rate.)
- Suppose instead that the government finances the spending increase through an increase in taxes—say, income taxes. Consumption will decline sharply. Exactly how much depends on consumers' expectations: The longer they expect the war to last, the longer they will expect higher taxes to last, and the more they will decrease their consumption. In any case, the increase in government spending will be partly offset by a decrease in consumption. Interest rates will increase by less than they would have increased under deficit spending, and investment will therefore decrease by less.

Assume that the economy is closed, so that $Y = C + I + G$. Suppose that G goes up, and Y remains the same. Then $C + I$ must decrease. If taxes are not increased, most of the decrease will come from a decrease in I. If taxes are increased, most of the decrease will come from a decrease in C.

In short, for a given output, the increase in government spending requires either a decrease in consumption or a decrease in investment. Whether the government relies on tax increases or deficits determines whether consumption or investment does more of the adjustment when government spending increases.

How does this affect who bears the burden of the war? The more the government relies on deficits, the smaller the decrease in consumption during the war and the larger the decrease in investment. Lower investment means a lower capital stock after the war and therefore lower output after the war. By reducing capital accumulation, deficits become a way of passing some of the burden of the war onto future generations.

Reducing Tax Distortions

There is another argument for running deficits, not only during wars but also, more generally, in times when government spending is exceptionally high. Think, for example, of reconstruction after an earthquake or the costs involved in the reunification of Germany in the early 1990s.

The argument is as follows: If the government were to increase taxes to finance the temporary increase in spending, tax rates would have to be very high. Very high tax rates can lead to very high economic distortions. Faced with very high tax rates, people work less or engage in illegal, untaxed activities. Rather than moving the tax rate up and down to always balance the budget, it is better (from the point of view of reducing distortions) to maintain a relatively constant tax rate—to *smooth taxes*. **Tax smoothing** implies running large deficits when government spending is exceptionally high and small surpluses the rest of the time.

22-4 THE DANGERS OF HIGH DEBT

We have seen that high debt requires higher taxes in the future. A lesson from history is that high debt can also lead to vicious cycles, making the conduct of fiscal policy extremely difficult. Let's look at this more closely.

High Debt, Default Risk, and Vicious Cycles

Return to equation (22.5):

$$\frac{B_t}{Y_t} - \frac{B_{t-1}}{Y_{t-1}} = (r - g)\frac{B_{t-1}}{Y_{t-1}} + \frac{(G_t - T_t)}{Y_t}$$

Take a country with a high debt ratio, say, 100%. Suppose the real interest rate is 3% and the growth rate is 2%. The first term on the right is $(3\% - 2\%)$ times $100\% = 1\%$ of GDP. Suppose further that the government is running a primary surplus of 1% of output, so just enough to keep the debt ratio constant (the right side of the equation equals $(3\% - 2\%) \times 100\% + (-1\%) = 0\%$).

Now suppose financial investors start to worry that the government may not be able to fully repay the debt. They ask for a higher interest rate to compensate for what they perceive as a higher risk of default on the debt. This makes it more difficult for the government to stabilize the debt. Suppose, for example, that the interest rate increases from 3% to 8%. Then, just to stabilize the debt, the government needs to run a primary surplus of 6% of output (the right side of the equation is then equal to $(8\% - 2\%) \times 100\% + (-6\%) = 0$).

Suppose that, in response to the increase in the interest rate, the government takes measures to increase the primary surplus to 6% of output. The spending cuts or tax increases that are needed are likely to prove politically costly, potentially generating more political uncertainty, a higher risk of default, and thus a further increase in the interest rate. Also, the sharp fiscal contraction is likely to lead to a recession, decreasing the growth rate. Both the increase in the real interest rate and the decrease in growth further increase $(r - g)$, requiring an even larger budget surplus to stabilize the debt. At some point, the government may become unable to increase the primary surplus sufficiently, and the debt ratio starts increasing, leading investors to become even more worried and to require an even higher interest rate. Increases in the interest rate and increases in the debt ratio feed on each other. In short, the higher the ratio of debt to GDP, the larger the potential for catastrophic debt dynamics. Even if the fear that the government may not fully repay the debt was initially unfounded, it can easily become self-fulfilling. The higher interest that the government must pay on its debt can lead the government to lose control of its budget and lead to an increase in debt to a level such that the government is unable to repay the debt, thus validating the initial fears.

This should remind you of bank runs and the discussion in Chapter 6. If people believe a bank is not solvent and decide to withdraw their funds, the bank may have to sell its assets at fire sale prices and become insolvent, validating the initial fears. Here, investors do not ask for their funds, but require a higher interest rate. The result is the same.

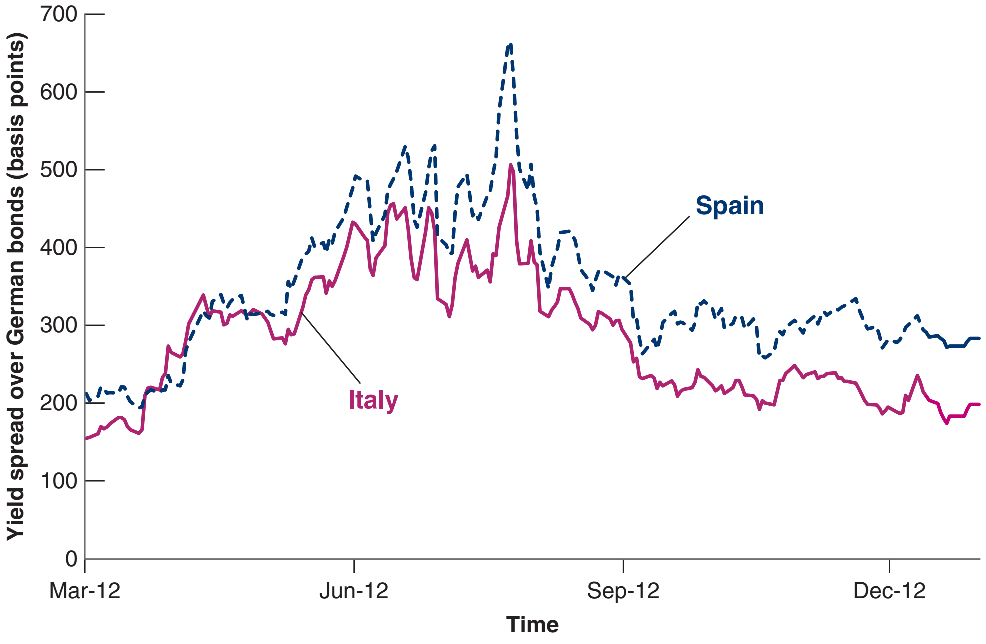


Figure 22-2

The Increase in European Bond Spreads

The spreads on Italian and Spanish two-year government bonds over German two-year bonds increased sharply between March and July 2012. At the end of July, when the European Central Bank stated that it would do whatever was necessary to prevent a breakup of the euro, the spreads decreased.

Source: Haver Analytics.

This is far from an abstract issue. Let's look again at what happened in the euro area during the crisis. Figure 22-2 shows the evolution of interest rates on Italian and Spanish government bonds from March to December 2012. For each country, it plots the difference, also called the **spread**, between the two-year interest rate on the country's government bonds and the two-year interest rate on German government bonds. The reason for comparing interest rates to German interest rates is that German bonds are considered nearly riskless. The spreads are measured, on the vertical axis, in **basis points** (a basis point is a hundredth of a percent).

Go back to Section 20-2 for a discussion of how, under fixed exchange rates, the expectation of a devaluation leads to high interest rates. The same is true for a country in a common currency area.

Both spreads started rising in March 2012. Toward the end of July, the spread on Italian bonds reached 500 basis points (equivalently, 5%), and the spread on Spanish bonds, 660 basis points (6.6%). These spreads reflected two worries: first, that the Italian and Spanish governments might default on their debt, and second, that they might devalue. In principle, in a monetary union, such as the euro area, no one should expect a devaluation unless markets start thinking that the monetary union might break up and that countries might reintroduce national currencies at a devalued exchange rate. This is exactly what investors worried about in the spring and summer of 2012. We can understand why by going back to our discussion of self-fulfilling debt crises.

Consider Italy, for instance. In March, the interest on Italian two-year bonds was below 3%; this was the sum of the interest on German two-year bonds, slightly below 1%, plus a 2% risk spread due to investors' concerns about the Italian government's creditworthiness. The country had at the time (and still has) a debt-to-GDP ratio above 130%. With interest below 3% such a high debt burden was sustainable; Italy was generating primary budget surpluses sufficient to keep the debt stable, although at that high level. Italy was fragile (because the debt was so high) but in a "good equilibrium." At this point investors started asking themselves what would happen if, for some reason, interest rates on Italian bonds were to double, to 6%. They concluded that if that happened, it was unlikely that Italy would be able to raise its primary surplus high enough to keep the debt stable. It was more likely that the country would enter a debt spiral and end up defaulting. At that point Italy might decide to abandon the monetary union and rely on a devaluation to improve its competitiveness and support growth because defaults are usually accompanied by sharp recessions. The fear that this might happen shifted Italy from a "good" to a "bad" equilibrium. As investors recognized that a default and an exit from the euro were a possibility, interest rates jumped to 6%, and this increase in interest rates validated the initial fears. Eventually, it was the European Central Bank (ECB) that shifted Italy back to the good equilibrium. On July 26, 2012, the president of the bank, Mario Draghi, clearly said that a breakup of the euro was out of question and that the ECB would do whatever was necessary to avoid it. Investors believed the promise and Italy shifted back to the good equilibrium.

By this statement, Mario Draghi meant that the ECB would be ready to buy Spanish or Italian bonds to maintain a low yield and go back to the "good equilibrium." In this event, the commitment was enough to decrease rates and the ECB did not have to intervene at all.

Thus, Italy and Spain succeeded, with the help of the ECB, in avoiding bad debt dynamics and default. What if a government does not succeed in stabilizing the debt and enters a debt spiral? Then, historically, one of two things happens: The government either explicitly defaults on its debt or relies increasingly on money finance. Let's look at each outcome.

The danger is not completely gone. At the time of writing, the spread on Italian bonds remains high, around 2.5%, due partly to still-high debt and deficits, which slightly exceed EU rules, and worries that the government, despite its statements to the contrary, may consider a euro exit.

Debt Default

At some point, when a government finds itself unable to repay the outstanding debt, it may decide to default. Default is often partial, however, and creditors take what is known as a **haircut**. A haircut of 30%, for example, means that creditors receive only 70% of what they were owed. Default also comes under many names, many of them euphemisms—probably to make the prospects more appealing (or less unappealing) to creditors. It is called **debt restructuring**, or **debt rescheduling** (when interest payments are deferred rather than cancelled), or, quite ironically, **private sector involvement** (the private sector, that is, the creditors, are asked to *get involved*, i.e., to accept a haircut). It may be unilaterally imposed by the government, or it may be the result of a negotiation with creditors, who, knowing that they will not be fully repaid in any case, may prefer to work out a deal with the government. This is what happened to Greece in 2012 when private creditors accepted a haircut of roughly 50%.

When debt is very high, default would seem to be an appealing solution. Having a lower level of debt after default reduces the size of the required fiscal consolidation and thus makes it more credible. It lowers required taxes, potentially allowing for higher growth. But default comes with high costs. If debt is held, for example, by domestic pension funds, as is often the case, retirees may suffer very much from the default. If it is held by domestic banks, then some banks may go bankrupt, with major adverse effects on the economy. If debt is held instead mostly by foreigners, then the country's international reputation may be lost and it may be difficult for the government to borrow from abroad for a long time. So, in general, and rightly so, governments are very reluctant to default on their debt.

Money Finance

For a refresher on how the central bank creates money, go back to Chapter 4, Section 4-3.

The other outcome is money finance. So far we have assumed that the only way a government could finance itself was by selling bonds. There is however another possibility. The government can finance itself by, in effect, printing money. The way it does this is not actually by printing money itself but by issuing bonds and then requiring the central bank to buy its bonds in exchange for money. This process is called **money finance** or **debt monetization**. Because, in this case, the rate of money creation is determined by the government deficit rather than by decisions of the central bank, this is also known as **fiscal dominance** of monetary policy.

How large a deficit can a government finance through such money creation? Let H be the amount of central bank money in the economy. (I shall refer to *central bank money* simply as *money* for short in what follows.) Let ΔH be money creation; that is, the change in the nominal money stock from one month to the next. The revenue, in real terms (that is, in terms of goods), that the government generates by creating an amount of money equal to ΔH is therefore $\Delta H/P$—money creation during the period divided by the price level. This revenue from money creation is called **seignorage**.

The word seignorage is revealing. The right to issue money was a precious source of revenue for the *seigneurs*, or lords, of the past. They could buy the goods they wanted by issuing their own money and using it to pay for the goods.

$$\text{seignorage} = \frac{\Delta H}{P}$$

Seignorage is equal to money creation divided by the price level. To see what rate of (central bank) nominal money growth is required to generate a given amount of seignorage, rewrite $\Delta H/P$ as

$$\frac{\Delta H}{P} = \frac{\Delta H}{H}\frac{H}{P}$$

In words: We can think of seignorage $(\Delta H/P)$ as the product of the rate of nominal money growth $(\Delta H/H)$ and the real money stock (H/P). Replacing this expression in the previous equation gives

$$\text{seignorage} = \frac{\Delta H}{H}\frac{H}{P}$$

This gives us a relation between seignorage, the rate of nominal money growth, and real money balances. To think about relevant magnitudes, it is convenient to take one more step and divide both sides of the equation by, say, monthly GDP, Y, to get

$$\frac{\text{seignorage}}{Y} = \frac{\Delta H}{H}\left(\frac{H/P}{Y}\right) \qquad (22.6)$$

Suppose the government is running a budget deficit equal to 10% of GDP and decides to finance it through seignorage, so $(\text{deficit}/Y) = (\text{seignorage}/Y) = 10\%$. The average ratio of central bank money to monthly GDP in advanced countries is roughly equal to 1, so choose $(H/P)/Y = 1$. This implies that nominal money growth must satisfy

$$10\% = \frac{\Delta H}{H} \text{ times } 1 \Rightarrow \frac{\Delta H}{H} = 10\%$$

Thus, to finance a deficit of 10% of GDP through seignorage, given a ratio of central bank money to monthly GDP of 1, the monthly growth rate of nominal money must be equal to 10%.

This is surely a high rate of money growth, but one might conclude that, in exceptional circumstances, it may be an acceptable price to pay to finance the deficit. Unfortunately, this conclusion would be wrong. As money growth increases, inflation typically follows. And high inflation leads people to want to reduce their demand for money, and in turn the demand for central bank money. In other words, as the rate of money growth increases, the real money balances that people want to hold decreases. If, for example, they were willing to hold money balances equal to one month of income when inflation was low, they may decide to reduce it to one week of income or less when inflation reaches 10%. In terms of equation (22.6), as $(\Delta H/H)$ increases, $(H/P)/Y$ decreases. And so, to achieve the same level of revenues, the government needs to increase the rate of money growth further. But higher money growth leads to further inflation, a further decrease in $(H/P)/Y$, and the need for further money growth.

This is an example of a general proposition. As you increase the tax rate (here the rate of inflation), the tax base (here real money balances) decreases.

Soon, high inflation turns into **hyperinflation**, the term that economists use for very high inflation—typically inflation in excess of 30% per month. The Focus Box "Money Financing and Hyperinflations" describes some of the most famous episodes.

FOCUS

Money Financing and Hyperinflations

We saw in this chapter that the attempt to finance a large fiscal deficit through money creation can lead to high inflation, or even hyperinflation. This scenario has played out many times in the past. (At the time of writing, it is playing out in Venezuela, where annual inflation hit 80,000% in 2018!) You probably have heard of the hyperinflation that took place in post–World War I Germany. In 1913, the value of all currency circulating in Germany was 6 billion marks. Ten years later, in October 1923, 6 billion marks was barely enough to buy a one-kilo loaf of rye bread in Berlin. A month later, the price of the same loaf of bread had increased to 428 billion marks. But the German hyperinflation is not the only example.

Table 1 summarizes the seven major hyperinflations that followed World War I and World War II. They share a number of features. They were all short (lasting a year or so) but intense, with money growth and inflation running at 50% *per month* or more. The increases in the price levels were staggering. As you can see, the largest price increase occurred not in Germany but in Hungary after World War II: What cost 1 Hungarian pengő in August 1945 cost 3,800 trillions of trillions of pengős less than a year later! And Hungary has the distinction of having not one but two hyperinflations, one after World War I and the other after World War II.

Inflation rates of this magnitude have not been seen since the 1940s. But many countries have experienced high inflation as a result of money finance. Monthly inflation ran above 20% in many Latin American countries in the late 1980s. A recent example of high inflation is Zimbabwe, where, in 2008, monthly inflation reached 500% before a stabilization program was adopted in early 2009. At the time of writing, monthly inflation has reached 300% in Venezuela. The photo shows a creative and more valuable use of bolívares, the Venezuelan money.

It will come as no surprise that hyperinflations have enormous economic costs:

- The transaction system works less and less well. One famous example of inefficient exchange occurred in Germany at the end of its hyperinflation in 1923. People had to use wheelbarrows to cart around the huge amounts of currency needed for daily transactions.
- Price signals become less and less useful. Because prices change so often, it is difficult for consumers and producers to assess the relative prices of goods and to make informed decisions. The evidence shows that the higher the rate of inflation, the higher the variation in the relative prices of different goods. Thus, the price system, which is crucial to the functioning of a market economy, also becomes less and less efficient. A joke heard in Israel during the high inflation of the 1980s: "Why is it cheaper to take the taxi rather than the bus? Because in the bus, you have to pay the fare at the beginning of the ride. In the taxi, you pay only at the end."
- Swings in the inflation rate become larger. It becomes harder to predict what inflation will be in the near future, whether it will be, say, 500% or 1,000% over the next year. Borrowing at a given nominal interest rate becomes more and more of a gamble. If we borrow at, say, 1,000% for a year, we may end up paying a real interest rate of 500% or 0%: a large difference! The result is that borrowing and lending typically come to a stop in the final months of hyperinflation, leading to a large decline in investment.

Table 1 Seven Hyperinflations of the 1920s and 1940s

Country	Start	End	P_T/P_0	Average Monthly Inflation Rate (%)	Average Monthly Money Growth (%)
Austria	Oct. 1921	Aug. 1922	70	47	31
Germany	Aug. 1922	Nov. 1923	1.0×10^{10}	322	314
Greece	Nov. 1943	Nov. 1944	4.7×10^{6}	365	220
Hungary 1	Mar. 1923	Feb. 1924	44	46	33
Hungary 2	Aug. 1945	Jul. 1946	3.8×10^{27}	19,800	12,200
Poland	Jan. 1923	Jan. 1924	699	82	72
Russia	Dec. 1921	Jan. 1924	1.2×10^{5}	57	49

P_T/P_0: Price level in the last month of hyperinflation divided by the price level in the first month.

Source: Philip Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman ed., *Studies in the Quantity Theory of Money* (University of Chicago Press, 1956), Table 1.

Hyperinflation ends only when fiscal policy is dramatically improved and the need to finance the deficit through money creation is eliminated. By then, the damage has been done.

As inflation becomes very high, there is typically an increasing consensus that it should be stopped. Eventually, the government reduces the deficit and no longer has recourse to money finance. Inflation stops, but not before the economy has suffered substantial costs.

22-5 THE CHALLENGES FACING US FISCAL POLICY TODAY

How bad is the debt situation in the United States today? In 2018, the debt issued by the federal government—the number that is typically quoted in newspapers, called **gross debt**—stood at $21,300 billion, or 104% of GDP. Part of that debt was held, however, by some government entities—for example, by the social security trust fund. The more relevant number, namely debt held by the public, called **net debt**, stood at $5,750 billion, or 77% of GDP, a substantially lower ratio than the gross debt ratio, although still a historically high number.

The federal deficit stood at $780 billion, or 3.8% of GDP. The ratio of interest payments to GDP was 1.6%, so the primary deficit (that is, the deficit net of interest payments) was equal to $3.8\% - 1.6\% = 2.2\%$; again, not a very high number, but still historically high in the absence of a war or a recession.

To see what this implied for debt dynamics, let's return to equation (22.5). Recall that the change in the debt ratio is equal to the sum of two terms: first, the difference between the interest rate and the growth rate, times the debt-to-GDP ratio; second, the ratio of the primary deficit to GDP:

$$\frac{B_t}{Y_t} - \frac{B_{t-1}}{Y_{t-1}} = (r - g)\frac{B_{t-1}}{Y_{t-1}} + \frac{(G_t - T_t)}{Y_t}$$

In 2018, the real interest rate on government debt was 0.5%—the difference between the nominal interest rate, 2.5%, and inflation, 2%. The growth rate was equal to 2.9%. The debt-to-GDP ratio was 77%, so the first term above was equal to $(0.5\% - 2.9\%) \times 77\% = -1.8\%$. Absent a primary deficit, the debt-to-GDP ratio would have declined by -1.8%. The primary deficit was positive, however, equal to 2.2%. Putting the two together, this implies that the debt ratio increased slightly, by $-1.8\% + 2.2\% = 0.4\%$ of GDP.

An increase in the debt-to-GDP ratio of 0.4% of GDP is small and suggests little reason for concern. The future, however, may see larger increases in debt: At 2.9%, growth was unusually high in 2018 and is forecast to be closer to 2%. And looking forward, two major government spending items look set to substantially increase:

- Social security payments are projected to increase from 4.9% of GDP in 2018 to 6.0% in 2029, reflecting the aging of America—the rapid increase in the proportion of people older than age 65 that will take place as the Baby Boom generation reaches retirement age.
- Medicare (the program that provides health care to retirees) and Medicaid (the program that provides health care to poor people) are projected to increase from 5.4% of GDP in 2018 to 7.2% in 2029. This large increase reflects the increasing cost of health care in the case of Medicaid and the increasing number of retirees in the case of Medicare.

Given these expected increases in spending, and for the reasons we discussed earlier in the chapter, it may be desirable to increase taxes and decrease the debt-to-GDP ratio now so as to have current taxpayers assume more of the burden of the increased future spending, as well as to avoid large, potentially distortionary, increases in taxes later on. This suggests starting to decrease the debt-to-GDP ratio now by reducing primary deficits or even generating primary surpluses, rather than letting debt increase further.

How much and at what rate should primary deficits be reduced? There are three relevant arguments here.

The first is that the real interest rate the government pays on its debt is currently very low, indeed lower than the growth rate. Figure 22-3 shows the evolution of the real interest rate, constructed as the nominal one-year T-bill rate minus actual CPI inflation since 1960. It shows that, after a sharp increase in the 1980s, the real rate has decreased from a peak of nearly 7% in 1984 to −0.3% in 2018. Current forecasts are that the interest rate will remain lower than the growth rate for the foreseeable future.

The average real interest rate on government debt was, as we saw earlier, a bit higher (0.5%) because the maturity of the government debt is longer than one year and long rates were higher than the one-year rate.

This has two implications. The first is that, while government debt is high, debt service, i.e., the interest paid on the debt, is still low. The second is that, if the interest rate remains lower than the growth rate, the government can decrease the ratio of debt to GDP over time while still running primary deficits. To see this, go back to equation (22.5) and the computation of debt dynamics we went through earlier. In 2018, absent a primary deficit, the debt-to-GDP ratio would have decreased by 1.8%. Put another way, the government could have run a primary deficit of up to 1.8% of GDP, and the debt-to-GDP ratio would still have declined. The difference between the growth rate and the interest rate is likely to be smaller in the future, but so long as the interest rate remains less than the growth rate, this implies that the government can run a (limited) primary deficit and still have the debt-to-GDP ratio decrease over time.

The second argument focuses on public investment. One of the ways the government deficit was reduced after the large increase during the Great Recession was by cutting public investment. In 2018, the ratio of government investment to GDP was the lowest it had been since 1950. There is a strong case for increasing it, and the case is strengthened by the fact that the cost of borrowing for the government is very low. Thus, there is an argument for accepting or even increasing the current primary deficits, if these are used to finance public investment instead of current spending or reductions in taxes.

This has been true not only in the United States, but in most countries. Fiscal consolidation often leads to cuts in public investment rather than in current spending, because it is less visible, affects current voters less, and is thus politically less costly.

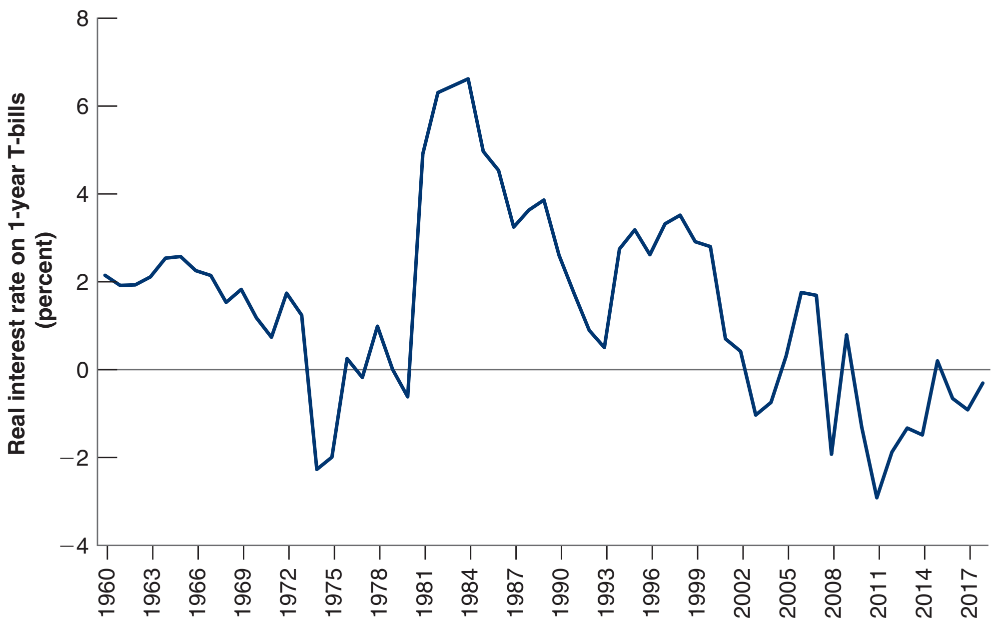
The third argument may be the most important. We saw in Chapter 5 that to avoid a decrease in output in response to a reduction in the budget deficit the central bank must

Figure 22-3

The Evolution of the Real Interest Rate on One-Year T-Bills since 1960

After sharply increasing in the early 1980s, the real interest rate has steadily declined.

Source: FRED: TB1YR, CPIAUCSL



decrease the interest rate. While, at the time of writing, the US economy is no longer at the zero lower bound, the nominal policy rate, at 2.4%, is still low. A large reduction in primary deficits would require a large decrease in the policy rate, a decrease that may be infeasible given the zero lower bound.

A back-of-the-envelope computation is useful here: Suppose that the multiplier associated with a reduction in government spending is equal to 1.5, so a 1% reduction in the primary deficit decreases demand by 1.5%. Suppose that a 1% decrease in the policy rate increases demand by 1%. Then, to avoid a decrease in output, a reduction in the ratio of the primary deficit to GDP by 2% would require a decrease in the policy rate of 3%, thus a larger decrease than is feasible given the zero lower bound. Put another way, given the limits on monetary policy, primary deficits may be needed to maintain output at potential.

To summarize: The US debt-to-GDP ratio is high and slowly increasing over time. Given the likely increases in spending in the future, it would be desirable to reduce primary deficits and reduce the debt ratio. Low interest rates, and by implication the low cost of debt, imply that it can be done slowly. And because low interest rates, together with the zero lower bound, put sharp limits on the use of monetary policy, they also imply that it has to be done slowly so as to avoid a decrease in demand and output.

SUMMARY

- The government budget constraint gives the evolution of government debt as a function of spending and taxes. One way of expressing the constraint is that the change in debt (the deficit) is equal to the primary deficit plus interest payments on the debt. The primary deficit is the difference between government spending on goods and services, G, and taxes net of transfers, T.
- The evolution of the ratio of debt to GDP depends on four factors: the interest rate, the growth rate, the initial debt ratio, and the primary surplus.
- Under the Ricardian equivalence proposition, a larger deficit is offset by an equal increase in private saving. Deficits have no effect on demand or output. The accumulation of debt does not affect capital accumulation. In practice, however, Ricardian equivalence fails and larger deficits lead to higher demand and higher output in the short run. The accumulation of debt leads to lower capital accumulation, and thus to lower output in the long run.
- To stabilize the economy, the government should run deficits during recessions and surpluses during booms. The cyclically adjusted deficit tells us what the deficit would be, under existing tax and spending rules, if output were at its potential level.
- Deficits are justified in times of high spending, such as wars. Relative to an increase in taxes, deficits lead to higher consumption and lower investment during wars. They therefore shift some of the burden of the war from people living during the war to those living after the war. Deficits also help smooth taxes and reduce tax distortions.
- High debt ratios increase the risk of vicious cycles. A higher perceived risk of default can lead to a higher interest rate and an increase in debt. The increase in debt can lead to a higher perceived risk of default and a higher interest rate. Together, both can combine to lead to a debt explosion. Governments may have no choice but to default or to rely on money finance. Money finance may in turn lead to hyperinflation. In either case, the economic costs are likely to be high.
- The US debt-to-GDP ratio is high and slowly increasing. It would be desirable to reduce this ratio over time. The limits on the use of monetary policy imply that the reduction will have to be gradual and limited.

KEY TERMS

inflation-adjusted deficit, 463
government budget constraint, 463
primary deficit, 465
primary surplus, 465
debt-to-GDP ratio, 468
debt ratio, 468
Ricardian equivalence, 470
Ricardo-Barro proposition, 470
full-employment deficit, 471
mid-cycle deficit, 471
standardized employment deficit, 471
structural deficit, 471
cyclically adjusted deficit, 471
automatic stabilizer, 472

tax smoothing, 474
spread, 475
basis points, 475
haircut, 476
debt restructuring, 476
debt rescheduling, 476
private sector involvement, 476
money finance, 476
debt monetization, 476
fiscal dominance, 476
seignorage, 477
hyperinflation, 477
gross debt, 479
net debt, 479

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The deficit is the difference between real government spending and taxes net of transfers.
b. The primary deficit is the difference between government spending on goods and services and taxes net of transfers.
c. The United States has experienced wide fluctuations in the ratio of debt to GDP in the past century.
d. Tax smoothing and deficit finance help spread the burden of war across generations.
e. The government should always take immediate action to eliminate a cyclically adjusted budget deficit.
f. If Ricardian equivalence holds, then an increase in income taxes will affect neither consumption nor saving.
g. The ratio of debt to GDP cannot exceed 100%.
h. A haircut reduces the value of outstanding government debt outstanding.
i. The cyclically adjusted deficit is always smaller than the actual deficit.
j. The inflation-adjusted deficit is always smaller than the actual deficit.
k. When the ratio of debt to GDP is high, the best policy is a fiscal consolidation.
l. A hyperinflation is an inflation rate greater than 30% per month.
m. Hyperinflations may distort prices, but they have no effect on real output.

2. Consider the following statement:

A deficit during a war can be a good thing. First, the deficit is temporary, so after the war is over, the government can go right back to its old level of spending and taxes. Second, given that the evidence supports the Ricardian equivalence proposition, the deficit will stimulate the economy during wartime, helping to keep the unemployment rate low.

Identify the mistakes in this statement. Is anything in this statement correct?

3. Consider an economy characterized by the following facts:

i. *The official budget deficit is 4% of GDP.*
ii. *The debt-to-GDP ratio is 100%.*
iii. *The nominal interest rate is 10%.*
iv. *The inflation rate is 7%.*

a. What is the primary deficit/surplus ratio to GDP?
b. What is the inflation-adjusted deficit/surplus ratio to GDP? *There is a "rule-of-thumb" in the chapter to understand how the actual deficit changes when output falls below its natural level.*
c. Suppose that output is 2% below its natural level. What is the cyclically adjusted, inflation-adjusted deficit/surplus ratio to GDP?
d. Suppose instead that output begins at its natural level and that output growth remains constant at the normal rate of 2%. How will the debt-to-GDP ratio change over time?

4. Assume that money demand takes the form

$$\frac{M}{P} = Y[1 - (r + \pi^e)]$$

where $Y = 1{,}000$ *and* $r = 0.1$.

a. Assume that, in the short run, π^e is constant and equal to 25%. Calculate the amount of seignorage for each annual rate of money growth, $\Delta M/M$, listed.
   a. 25%
   b. 50%
   c. 75%
b. In the medium run, $\pi^e = \pi = \Delta M/M$. Compute the amount of seignorage associated with the three rates of annual money growth in part a. Explain why the answers differ from those in part a.

DIG DEEPER

5. Consider the economy described in Problem 3 and assume that there is a fixed exchange rate, $\bar{E}$. *Suppose that financial investors worry that the level of debt is too high and that the government may devalue to stimulate output (and therefore tax revenues) to help pay down the debt. Financial investors begin to expect a devaluation of 10%. In other words, the expected exchange rate,* E^e_{t+1}, *decreases by 10% from its previous value of* $\bar{E}$.

a. Recall the uncovered interest parity condition:

$$i_t = i^*_t - \frac{E^e_{t+1} - \bar{E}}{\bar{E}}$$

If the foreign interest rate remains constant at 10% a year, what must happen to the domestic interest rate when E^e_{t+1} decreases by 10%?

b. Suppose that domestic inflation remains the same. What happens to the domestic real interest rate? What is likely to happen to the growth rate?
c. What happens to the official budget deficit? What happens to the inflation-adjusted deficit?

d. Suppose the growth rate decreases from 2% to 0%. What happens to the change in the debt ratio? (Assume that the primary deficit/surplus ratio to GDP is unchanged, even though the fall in growth may reduce tax revenues.)

6. Ricardian equivalence and fiscal policy

First consider an economy in which Ricardian equivalence does not hold.

a. Suppose the government starts with a balanced budget. Then, there is an increase in government spending, but there is no change in taxes. Show in an *IS-LM* diagram the effect of this policy on output in the short run when the central bank keeps the real interest rate constant. How will the government finance the increase in government spending?

b. Suppose, as in part a, that the government starts with a balanced budget and then increases government spending. This time, however, assume that taxes increase by the same amount as government spending. Show in an *IS-LM* diagram the effect of this policy on output in the short run. (It may help to recall the discussion of the multiplier in Chapter 3. Does government spending or tax policy have a bigger multiplier?) How does the output effect compare with the effect in part a?

Now suppose Ricardian equivalence holds in this economy. [Parts c and d do not require use of diagrams.]

c. Consider again an increase in government spending with no change in taxes. How does the output effect compare to the output effects in parts a and b?

d. Consider again an increase in government spending combined with an increase in taxes of the same amount. How does this output effect compare to the output effects in parts a and b?

e. Comment on each of the following statements:

   i. "Under Ricardian equivalence, government spending has no effect on output."

   ii. "Under Ricardian equivalence, changes in taxes have no effect on output."

EXPLORE FURTHER

7. Consider an economy characterized by the following facts:

i. The debt-to-GDP ratio is 40%.

ii. The primary deficit is 4% of GDP.

iii. The normal growth rate is 3%.

iv. The real interest rate is 3%.

a. Using your favorite spreadsheet software, compute the debt-to-GDP ratio in 10 years, assuming that the primary deficit stays at 4% of GDP each year; the economy grows at the normal growth rate in each year; and the real interest rate is constant at 3%.

b. Suppose the real interest rate increases to 5%, but everything else remains as in part a. Compute the debt-to-GDP ratio in 10 years.

c. Suppose the normal growth rate falls to 1%, and the economy grows at the normal growth rate each year. Everything else remains as in part a. Calculate the debt-to-GDP ratio in 10 years. Compare your answer to part b.

d. Return to the assumptions of part a. Suppose policymakers decide that a debt-to-GDP ratio of more than 50% is dangerous. Verify that reducing the primary deficit to 1% immediately and that maintaining this deficit for 10 years will produce a debt-to-GDP ratio of 50% in 10 years. Thereafter, what value of the primary deficit will be required to maintain the debt-to-GDP ratio of 50%?

e. Continuing with part d, suppose policy makers wait 5 years before changing fiscal policy. For five years, the primary deficit remains at 4% of GDP. What is the debt-to-GDP ratio in 5 years? Suppose that after five years, policy makers decide to reduce the debt-to-GDP ratio to 50%. In years 6 through 10, what constant value of the primary deficit will produce a debt-to-GDP ratio of 50% at the end of year 10?

f. Suppose that policy makers carry out the policy in either parts d or e. If these policies reduce the growth rate of output for a while, how will this affect the size of the reduction in the primary deficit required to achieve a debt-to-GDP ratio of 50% in 10 years?

g. Which policy—the one in part d or the one in part e—do you think is more dangerous to the stability of the economy?

8. The fiscal situation in the United States and in other countries

From the FRED economic database at the Federal Reserve Bank of St. Louis, you can retrieve two series: General Government Gross Debt of the United States (GGGDTAUSA188N) and a measure of the primary deficit (USAGGXONLBGDP). Both are measured as a percent of GDP. These are measures that incorporate all levels of government. This data are constructed by the International Monetary Fund (IMF). Using data from the IMF or other international organizations helps make a better comparison across countries.

a. What is the ratio of debt to GDP in the United States in the most recent available year? Describe the path of this variable in the last decade?

b. What is the change in the ratio of debt-to-GDP in the last year of the data? Can the debt-to-GDP ratio fall even if the primary deficit is positive?

c. Use the information on the change in the debt-to-GDP ratio and the primary deficit ratio to infer the missing term in equation (22.5) in the last year of the data. Does your calculation make sense to you?

d. Similar data are constructed for all countries. A convenient source that compares the fiscal situation for the overall government sector in the G7 countries is published by Canada's Department of Finance in a document called the "Fiscal Reference Tables." The section titled International Fiscal Comparisons at the end of the document presents the most recent data. Which large economy has the highest and lowest ratio of gross debt to GDP? Which country has the highest and lowest deficit as a percent of GDP? Are these overall deficits or primary deficits?

9. The Congressional Budget Office (CBO) is required to produce a forecast of the federal fiscal situation each year. This question uses the version published in January 2019. There is a document entitled "A Visual Summary of The Budget and Economic Outlook 2019 to 2029." Find the relevant figures that help answer the questions below.

a. Is the federal government expected to be in deficit or surplus over the 10-year horizon?
b. In this presentation do outlays include the interest on the debt? Do they include transfers?
c. Is the ratio of debt to GDP expected to rise or fall over the 10-year horizon?
d. What is the CBO assuming about the term *(r-g)* in the equation describing the change in the debt-to-GDP ratio?

FURTHER READINGS

- The modern statement of the Ricardian equivalence proposition is Robert Barro's "Are Government Bonds Net Wealth?", *Journal of Political Economy*, 1974, 82(6): pp. 1095–1117.
- Each year, the Congressional Budget Office publishes *The Economic and Budget Outlook* for the current and future fiscal years. The document provides a clear and unbiased presentation of the current US budget, current budget issues, and budget trends. Available at www.cbo.gov/.
- For more on the German hyperinflation, read Steven Webb, *Hyperinflation and Stabilization in the Weimar Republic* (Oxford University Press, 1989).
- A good review of what economists know and don't know about hyperinflation is given in Rudiger Dornbusch, Federico Sturzenegger, and Holger Wolf, "Extreme Inflation: Dynamics and Stabilization," *Brookings Papers on Economic Activity*, 1990, Vol. 2, pp. 1–84.
- For the debate on "fiscal austerity" in Europe, see www.voxeu.org/debates/has-austerity-gone-too-far.

23

Monetary Policy: A Summing Up

For the two decades before the Great Recession, most central banks had converged toward a framework for monetary policy, called **inflation targeting**. It was based on two principles: The first was that the primary goal of monetary policy was to keep inflation stable and low. The second was that the best way to achieve this goal was to follow, explicitly or implicitly, an **interest rate rule**, allowing the policy rate to respond to movements in inflation and in activity.

Until the crisis, this framework appeared to work well. Inflation decreased and remained low and stable in most countries. Output fluctuations decreased in amplitude. The period became known as the **Great Moderation**. Many researchers concluded that better monetary policy was one of the main factors behind the improvement, consolidating the support for this monetary policy framework.

Then the crisis came. It forced macroeconomists and central bankers to reassess along at least two dimensions.

The first is the set of issues raised by the liquidity trap. If and when an economy reaches the zero lower bound, the policy rate can no longer be used to increase activity. This raises two questions: First, can monetary policy be conducted in such a way as to avoid getting to the zero lower bound in the first place? Second, once the economy is at the zero lower bound, are there other tools that the central bank can use to help increase activity?

The second concerns the mandate of the central bank and the tools of monetary policy. From the early 2000s to the start of the crisis, most advanced economies appeared to do well, with sustained output growth and stable inflation. Yet, as we saw in Chapter 6, behind the scenes not everything was fine. Important changes were taking place in the financial system, such as the large increase in leverage and the increased reliance on wholesale funding by banks. In many countries, also, there were sharp increases in housing prices. These factors turned out to be at the source of the crisis. This raises two more questions: Looking forward, should the central bank worry not only about inflation and activity but also about asset prices, stock market booms, housing booms, and risk in the financial sector? And if so, what tools does it have at its disposal?

This chapter reviews what we have learned about monetary policy so far, then describes the logic of inflation targeting and the use of an interest rate rule, and finally discusses where macroeconomists stand on the issues raised by the crisis.

Section 23-1 takes stock of what we have learned in this book so far.

Section 23-2 describes the inflation-targeting framework.

Section 23-3 reviews the costs and benefits of inflation and draws implications for the choice of a target inflation rate.

Section 23-4 describes the monetary policy measures taken by central banks when they hit the zero lower bound and since then.

Section 23-5 discusses the potential role of central banks in ensuring financial stability.

If you remember one basic message from this chapter, it should be: Before the crisis, monetary policy had converged to a framework called inflation targeting. The crisis has forced a reassessment of both the mandate and of the tools, a reassessment that is still going on. ▶▶▶

23-1 WHAT WE HAVE LEARNED

- In Chapter 4 we looked at money demand and money supply and the determination of the interest rate.

  We saw how the central bank controls the policy rate through changes in the money supply. We saw also that, when the policy rate reaches zero, a case known as the liquidity trap or the zero lower bound, further increases in the money supply have no effect on the policy rate.
- In Chapter 5 we looked at the short-run effects of monetary policy on output.

  We saw that a decrease in the interest rate leads to an increase in spending and, in turn, an increase in output. We saw how monetary and fiscal policy can be used to affect both the level of output and its composition.
- In Chapter 6, we introduced two important distinctions, first between the nominal and the real interest rate and, second, between the borrowing rate and the policy rate. The real interest rate is equal to the nominal interest rate minus expected inflation. The borrowing rate is equal to the policy rate plus a risk premium.

  We saw that what matters for private spending decisions is the real borrowing rate. We discussed how the state of the financial system affects the relation between the policy rate and the borrowing rate.
- In Chapter 9 we looked at the effects of monetary policy in the medium run.

  We saw that, in the medium run, monetary policy affects neither output nor the real interest rate. Output returns to potential, and the real interest rate returns to its natural rate, also called the *neutral rate* or the *Wicksellian rate of interest*. Higher money growth does not affect either output or the real interest rate, but leads to higher inflation.

  We saw, however, how the zero lower bound may derail this adjustment. High unemployment may lead to deflation, which, at the zero lower bound, leads to a higher real interest rate, which further decreases demand and further increases unemployment.
- In Chapter 14 we introduced another important distinction, between short-term and long-term interest rates.

  We saw that long term-interest rates depend on expectations of future short-term rates and a term premium. We saw that stock prices depend on expected future short-term rates, future dividends, and an equity premium.
  We saw, however, that stock prices may be subject to bubbles or fads, making the prices differ from the fundamental values of the stocks.
- In Chapter 16 we looked at the effects of expectations on spending and output, and the role of monetary policy in this context.

  We saw that monetary policy affects the short-term nominal interest rate, but that spending depends on current and expected future short-term real interest rates. We saw that the effects of monetary policy on output depend crucially on how expectations respond to monetary policy.
- In Chapter 19 we looked at the effects of monetary policy in an economy open in both goods markets and financial markets.

We saw that, in an open economy, monetary policy affects spending and output not only through the interest rate but also through the exchange rate. An increase in money leads to both a decrease in the interest rate and a depreciation, both of which increase spending and output. We saw that, under fixed exchange rates, the central bank gives up monetary policy as a policy instrument.

- In Chapter 20 we discussed the pros and cons of flexible exchange rates versus fixed exchange rates.

  We saw that, under flexible exchange rates, interest rate movements can lead to large changes in exchange rates; under fixed exchange rates, speculation can lead to an exchange rate crisis and a sharp devaluation. We discussed the pros and cons of adopting a common currency such as the euro, or even giving up monetary policy altogether through the adoption of a currency board or dollarization.
- In Chapter 21 we looked at the problems facing macroeconomic policy in general, and monetary policy in particular.

  We saw that uncertainty about the effects of policy should lead to more cautious policies. We saw that even well-intentioned policymakers may sometimes not do what is best, and that there is a case to be made for putting restraints on policymakers. We also looked at the benefits of having an independent central bank and of appointing a conservative central banker.

In this chapter we extend the analysis to look first at the inflation targeting framework in place before the crisis, and then at the challenges to monetary policy raised by the crisis.

23-2 FROM MONEY TARGETING TO INFLATION TARGETING

One can think of the goals of monetary policy as twofold: First, to maintain low and stable inflation; second, to stabilize output around potential—to avoid or at least limit recessions or booms.

Money Targeting

Until the 1980s, the strategy was to choose a target rate of money growth and to allow for deviations from that target rate as a function of activity. The rationale was simple. A low target rate of money growth implied a low average rate of inflation. In recessions, the central bank could increase money growth, leading to a decrease in interest rates and an increase in output. In booms, the central bank could decrease money growth, leading to an increase in interest rates and a slowdown in output.

That strategy did not work well.

First, the relation between money growth and inflation turned out to be far from tight, even *in the medium run*. This is shown in Figure 23-1, which plots 10-year averages of the US inflation rate against 10-year averages of the growth rate of money from 1970 to the crisis (the way to read the figure: The numbers for inflation and for money growth for 2000, for example, are the average inflation rate and the average growth rate of money from 1991 to 2000). The inflation rate is constructed using the consumer price index (CPI) as the price index. The growth rate of nominal money is constructed using the sum of currency and checkable deposits, known as **M1**, as the measure for the money stock.

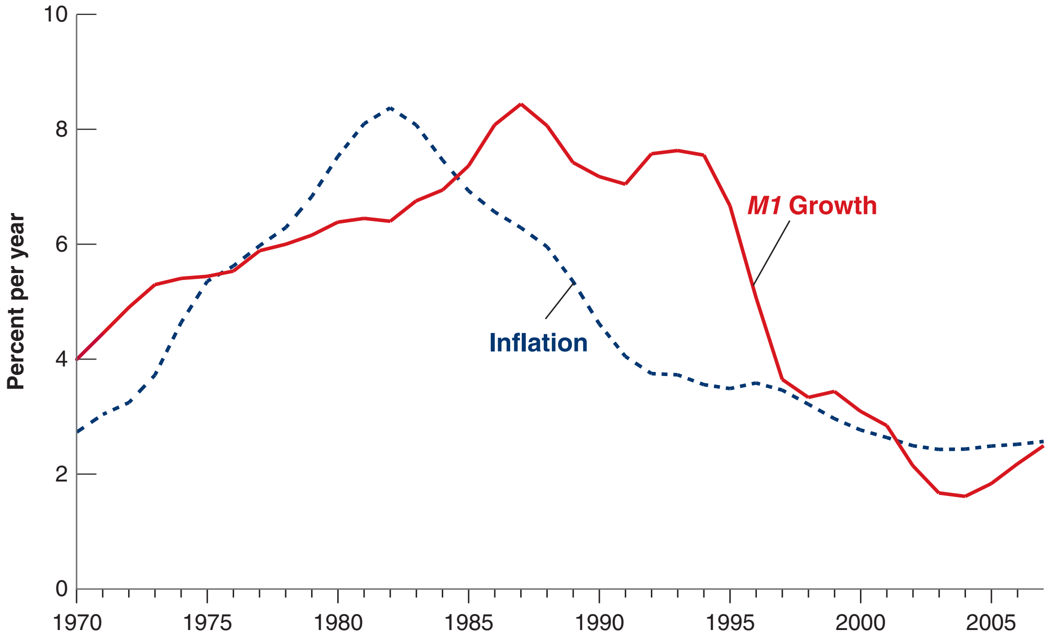
The reason for using 10-year averages should be clear. In the short run, changes in nominal money growth affect mostly interest rates and output rather than inflation. It

Figure 23-1

M1 Growth and Inflation: 10-Year Averages, 1970 to 2007

There is no tight relation between M1 growth and inflation, even in the medium run.

Source: FRED: CPIAUSL, M1SL



is only in the medium run that a relation between nominal money growth and inflation should emerge. Taking 10-year averages of both nominal money growth and inflation is a way of detecting such a medium-run relation. The reason for stopping at the crisis is that, as we saw in Chapter 4, when an economy hits the zero lower bound (which the US economy did at the end of 2008), increases in the money supply no longer have an effect on the policy rate and, by implication, the central bank is no longer able to affect output and inflation; so we exclude the period during which the US economy was stuck at the zero lower bound.

Figure 23-1 shows that, for the United States, the relation between M1 growth and inflation was not tight. True, both went up in the 1970s, and both came down later. But note that inflation started declining in the early 1980s, whereas nominal money growth remained high for another decade and came down only in the 1990s. Average inflation from 1981 to 1990 was down to 4%, while average money growth over the same period was still running at 7.5%.

From Chapter 5, equation (5.3): The real money supply (the left side) must be equal to the real demand for money (the right side):

$$\frac{M}{P} = YL(i)$$

If, as a result of the introduction of credit cards, the real demand for money halves, then

$$\frac{M}{P} = \frac{1}{2}YL(i)$$

In the short run, P does not change, and so the interest rate must adjust. In the medium run, P adjusts. For a given level of output and a given interest rate, M/P must halve. Given M, this implies that P must double.

Second, even the relation between the money supply and the interest rate *in the short run* also turned out to be unreliable. This made money growth an unreliable instrument to affect demand and output.

Both problems—the poor relation between money growth and inflation in the medium run, and the poor relation of the interest rate to the money supply in the short run—had the same origin: *shifts in the demand for money*. An example will help here. Suppose, as a result of the introduction of credit cards, people decide to hold only half the amount of money they held before; in other words, the real demand for money decreases by half. In the short run, at a given price level, this large decrease in the demand for money will lead to a large decrease in the interest rate. In other words, we will see a large decrease in the interest rate with no change in the money supply. (In terms of Figure 4–2 in Chapter 4, money demand will shift to the left, leading to a decrease in the equilibrium interest rate.) In the medium run, at a given interest rate, the price level will adjust and the real money stock will eventually decrease by half. For a given nominal money stock, the price level will eventually double. So, even if the nominal money stock remains constant, there will still be a period of inflation as the price level doubles. During this period, there will be no tight relation between nominal money growth (which is zero) and inflation (which would be positive).

Throughout the 1970s and 1980s, these frequent and large shifts in money demand created serious problems. Central banks found themselves torn between trying to keep a stable target for money growth and staying within announced bands (to maintain credibility), or adjusting to shifts in money demand (to stabilize the interest rate and output in the short run and inflation in the medium run). Starting in the early 1990s, a dramatic rethinking of monetary policy took place based on targeting inflation rather than money growth and the use of an interest rate rule to achieve it. Let's look at it more closely.

Inflation Targeting

If one of the main goals of the central bank is to achieve low and stable inflation, why not target inflation directly rather than money growth? And if the way to affect activity in the short run is to rely on the effect of the interest rate on spending, why not focus directly on the interest rate rather than on money growth? This is the reasoning that led to the elaboration of inflation targeting. Central banks committed to achieving a target inflation rate, and decided to use the interest rate as the instrument to achieve it. Let's look at both parts of the strategy.

Committing to a given inflation target *in the medium run* is hardly controversial. Trying to achieve a given inflation target *in the short run* would appear to be much more controversial. Focusing exclusively on inflation would seem to eliminate any role for monetary policy in reducing output fluctuations. But in fact this is not the case.

To see why, return to the Phillips curve relation between inflation, π_t, expected inflation, π_t^e, and the deviation of the unemployment rate, u_t, from the natural rate of unemployment, u_n (equation (8.10)):

$$\pi_t = \pi_t^e - \alpha(u_t - u_n)$$

Let the inflation target be $\bar{\pi}$. Assume that, thanks to the central bank's reputation, this target is credible, so that people expect inflation to be equal to the target. The relation becomes:

$$\pi_t = \bar{\pi} - \alpha(u_t - u_n)$$

Note that, if the central bank can hit its inflation target exactly, so $\pi_t = \bar{\pi}$, unemployment will be equal to its natural rate. By targeting and achieving a constant rate of inflation in line with inflation expectations, the central bank also keeps unemployment at the natural rate, and by implication keeps output at potential.

◀ $0 = -\alpha(u_t - u_n) \Rightarrow u_t = u_n$.

Put strongly: Even if policymakers did not care about inflation per se (they do) but cared only about output, inflation targeting would still make sense. Keeping inflation stable is a way to keep unemployment at the natural rate, or equivalently to keep output at potential. This result has been dubbed the **divine coincidence**. With a Phillips curve of the form given in equation (8.10), there is no conflict between keeping inflation constant and keeping output at potential. A focus on keeping stable inflation is thus the right approach to monetary policy, in both the short and medium run.

This result is a useful benchmark, but it is too strong. Life is not that nice. The main objection is that, as we saw in Chapter 8, the Phillips curve relation is far from an exact relation. Rather than try to achieve the target inflation rate each period, at the risk of large fluctuations in unemployment, it is better to try to achieve it only over time. So most central banks adopted what is called **flexible inflation targeting**: When inflation is away from target, rather than trying to return it to target right away, they adjust the interest rate to return to the target inflation rate over time.

The Interest Rate Rule

Inflation is not under the direct control of the central bank. The policy rate is. Thus, the question is how to set the policy rate to achieve the target rate of inflation. The answer is a simple one: When inflation is higher than the target, increase the policy rate to decrease demand and the pressure on prices; when it is below the target rate of inflation, decrease the policy rate. With this in mind, in the 1990s, John Taylor, at Stanford University, suggested the following rule for the policy rate, a rule now known as the **Taylor rule**:

- Let π_t be the rate of inflation and $\bar{\pi}$ be the target rate of inflation.
- Let i_t be the policy rate, that is, the nominal interest rate controlled by the central bank, and $\bar{i}$ be the target nominal interest rate—the nominal interest rate associated with the neutral rate of interest, r_n, and the target rate of inflation, $\bar{\pi}$, so $\bar{i} = r_n + \bar{\pi}$.
- Let u_t be the unemployment rate and u_n be the natural unemployment rate.

For a refresher, go back to Chapter 9, Section 9-1.

Think of the central bank as choosing the nominal interest rate, *i*. (Recall, from Chapter 4, that, through open market operations, and ignoring the liquidity trap, the central bank can achieve any short-term nominal interest rate that it wants.) Then, Taylor argued, the central bank should use the following rule:

$$i_t = \bar{i} + a(\pi_t - \bar{\pi}) - b(u_t - u_n)$$

where *a* and *b* are positive coefficients chosen by the central bank.

Let's look at what the rule says:

- If inflation is equal to target inflation $(\pi_t = \bar{\pi})$ and the unemployment rate is equal to the natural rate of unemployment $(u_t = u_n)$, then the central bank should set the nominal interest rate, i_t, equal to its target value, $\bar{i}$. This way, the economy can stay on the same path, with inflation equal to the target inflation rate and unemployment equal to the natural rate of unemployment.
- If inflation is higher than the target $(\pi_t > \bar{\pi})$, the central bank should increase the nominal interest rate, i_t, above $\bar{i}$. This higher interest rate will lead to an increase in unemployment, and this increase in unemployment will lead to a decrease in inflation. The coefficient *a* should therefore reflect how much the central bank cares about inflation. The higher *a*, the more the central bank will increase the interest rate in response to inflation, the more the economy will slow down, the more unemployment will increase, and the faster inflation will return to the target inflation rate.
- In any case, as Taylor pointed out, *a* should be larger than one. Why? Because what matters for spending is the real interest rate, not the nominal interest rate. When inflation increases, the central bank, if it wants to decrease spending and output, must increase the *real* interest rate. In other words, it must increase the nominal interest rate more than one-for-one with inflation.
- If unemployment is higher than the natural rate of unemployment $(u_t > u_n)$, the central bank should decrease the nominal interest rate. The lower nominal interest rate will lead to an increase in output, leading to a decrease in unemployment. The coefficient *b* should reflect how much the central bank cares about unemployment. The higher *b*, the more the central bank will be willing to deviate from target inflation to keep unemployment close to the natural rate of unemployment.

Some economists argue that the increase in US inflation in the 1970s was due to the fact that the Fed increased the nominal interest rate less than one-for-one with inflation. The result, they argue, was that an increase in inflation led to a decrease in the real interest rate, which led to higher demand, lower unemployment, more inflation, a further decrease in the real interest rate, and so on.

In stating this rule, Taylor did not argue that it should be followed blindly. Many other events, such as an exchange rate crisis or the need to change the composition of spending on goods, and thus the mix between monetary policy and fiscal policy, justify changing the nominal interest rate for reasons other than those included in the rule. But, he argued, the rule provided a useful way of thinking about monetary policy. Once the central bank has chosen a target rate of inflation, it should try to achieve it by adjusting the nominal interest rate. The rule it should follow should consider not only current inflation but also current unemployment.

The logic of the rule was convincing, and, by the mid-2000s, most central banks in advanced economies had adopted some form of flexible inflation targeting, that is, the choice of an inflation target together with the use of an interest rule.

Then the crisis came and raised many questions, about the choice of the inflation target, what to do when the interest rate suggested by the interest rule reaches the zero lower bound, and whether and how the central bank should worry about financial stability in addition to inflation and output. The next section discusses the choice of the inflation target, and the following sections discuss other questions raised by the crisis.

23-3 THE OPTIMAL INFLATION RATE

Table 23-1 shows that inflation steadily decreased in advanced economies from the early 1980s on. In 1981, average inflation in the OECD was 10.5%; in 2018, it was down to 2.3%. In 1981, only two countries (out of the 24 OECD members at the time) had an inflation rate below 5%; in 2018, the number had increased to 35 out of 36.

◀ The country with inflation above 5% is Turkey (15%).

Before the crisis, most central banks aimed for an inflation rate of about 2%. Was this the right goal? The answer depends on the costs and benefits of inflation.

The Costs of Inflation

We saw in Chapter 22 that very high inflation, say a rate of 30% per month or more, can disrupt economic activity. The debate in advanced economies today, however, is not about the costs of inflation rates of 30% or more per month. Rather, it centers on the advantages of, say, 0% versus 4% inflation per year. Within that range, economists identify four main costs of inflation: (1) shoe-leather costs, (2) tax distortions, (3) money illusion, and (4) inflation variability.

Shoe-Leather Costs

Recall from Chapter 9 that, in the medium run, a higher inflation rate leads to a higher nominal interest rate and so a higher opportunity cost of holding money. As a result, people decrease their money balances by making more trips to the bank—thus the expression **shoe-leather costs**. These trips would be avoided if inflation were lower and people could be doing other things instead, such as working more or enjoying leisure.

During hyperinflations, shoe-leather costs become quite large. But their importance in times of moderate inflation is limited, at best. If an inflation rate of 4% leads people to go to the bank, say, one more time every month, or to do one more transaction between their money market fund and their checking account each month, this hardly qualifies as a major cost of inflation.

Table 23-1 Inflation Rates in the OECD, 1981–2018

Year	1981	1990	2000	2010	2018
OECD average*	10.5%	6.2%	2.8%	1.2%	2.3%
Number of countries with inflation below 5%**	2/24	15/24	24/27	27/30	35/36

*Average of GDP deflator inflation rates, using relative GDPs measured at PPP prices as weights.
**The second number denotes the number of member countries at the time.

Tax Distortions

The second cost of inflation comes from the interaction between the tax system and inflation.

Consider, for example, the taxation of capital gains. Taxes on capital gains are typically based on the change in the price in dollars of the asset between the time it was purchased and the time it is sold. This implies that the higher the rate of inflation, the higher the tax. An example will make this clear:

- Suppose inflation has been running at $\pi\%$ a year for the last 10 years.
- Suppose also that you bought your house for \$100,000 10 years ago, and you are selling it today for \$100,000 times $(1 + \pi\%)^{10}$; so its real value is unchanged.
- If the capital gains tax is 30%, the *effective tax rate* on the sale of your house—defined as the ratio of the tax you pay to the price for which you sell your house—is

The numerator of the fraction equals the sale price minus the purchase price. The denominator is the sale price.

$$(30\%)\ \frac{100{,}000\,(1 + \pi\%)^{10} - 100{,}000}{100{,}000\,(1 + \pi\%)^{10}}$$

- Because you are selling your house for the same real price at which you bought it, your real capital gain is zero, so you should not pay any tax. Indeed, if $\pi = 0$—there has been no inflation—then the effective tax rate is 0. But if, for example, $\pi = 4\%$, then the effective tax rate is 9.7%: Despite the fact that your real capital gain is zero, you end up paying a high tax.

The problems created by the interactions between taxation and inflation extend beyond capital gains taxes. Although we know that the real rate of return on an asset is the real interest rate, not the nominal interest rate, income for the purpose of income taxation includes nominal interest payments, not real interest payments. Or to take yet another example, until the early 1980s in the United States, the income levels corresponding to different income tax rates were not increased automatically with inflation. As a result, people were pushed into higher tax brackets as their nominal income—but not necessarily their real income—increased over time, an effect known as **bracket creep**.

Some economists argue that the costs of bracket creep were much larger. As tax revenues steadily increased, there was little pressure on the government to control spending. The result, they argue, was an increase in the overall size of the government in the 1960s and 1970s far beyond what would have been desirable.

You might argue that this cost is not a cost of inflation per se but rather the result of a badly designed tax system. In the house example we just discussed, the government could eliminate the problem if it *indexed* the purchase price to the price level—that is, it adjusted the purchase price for inflation since the time of purchase—and computed the tax on the difference between the sale price and the adjusted purchase price. Under this computation, there would be no capital gains and therefore no capital gains tax to pay. But because tax codes around the world rarely define the tax base in real terms, the inflation rate matters and leads to distortions.

Money Illusion

The third cost comes from **money illusion**—the notion that people appear to make systematic mistakes in assessing nominal versus real changes in incomes and interest rates. A number of computations that would be simple when prices are stable become more complicated when there is inflation. When people compare their income this year to their income in previous years, they must keep track of the history of inflation. When choosing between different assets or deciding how much to consume or save, they must keep track of the difference between the real interest rate and the nominal interest rate. Casual evidence suggests that many people find these computations difficult and often fail to make the relevant distinctions. Economists and psychologists have gathered more formal evidence, and it suggests that inflation often leads people and firms to make incorrect decisions (see the Focus Box "Money Illusion"). If this is the case, then a seemingly simple solution is to have zero inflation.

Money Illusion

FOCUS

There is a lot of anecdotal evidence that many people fail to properly adjust for inflation in their financial computations. Recently, economists and psychologists have started looking at money illusion more closely. In a recent study, two psychologists, Eldar Shafir at Princeton and Amos Tversky at Stanford, and one economist, Peter Diamond at MIT, designed a survey aimed at finding out how prevalent money illusion is and what causes it. Among the many questions they asked of people in various groups (people at Newark International Airport, people at two New Jersey shopping malls, and a group of Princeton undergraduates) is the following:

Suppose Adam, Barbara, and Carlos each received an inheritance of $200,000 and each used it immediately to purchase a house. Suppose each sold the house one year after buying it. Economic conditions were, however, different in each case:

- During the time Adam owned the house, there was a 25% deflation—the prices of all goods and services decreased by approximately 25%. A year after he bought the house, he sold it for $154,000 (23% less than what he had paid).
- During the time Barbara owned the house, there was no inflation or deflation—the prices of all goods and services did not change significantly during the year. A year after she bought the house, she sold it for $198,000 (1% less than what he had paid).
- During the time Carlos owned the house, there was a 25% inflation—the prices of all goods and services increased by approximately 25%. A year after Carlos bought the house, he sold it for $246,000 (23% more than what he had paid).

Please rank Adam, Barbara, and Carlos in terms of the success of their house transactions. Assign "1" to the person who made the best deal and "3" to the person who made the worst deal.

In nominal terms, Carlos clearly made the best deal, followed by Barbara, followed by Adam. But what is relevant is how they did in real terms, adjusting for inflation. In real terms, the ranking is reversed. Adam, with a 2% real gain, made the best deal, followed by Barbara (with a 1% loss), followed by Carlos (with a 2% loss).

The survey's answers are shown below.

Rank	Adam	Barbara	Carlos
1st	37%	15%	48%
2nd	10%	74%	16%
3rd	53%	11%	36%

Carlos was ranked first by 48% of the respondents, and Adam was ranked third by 53% of the respondents. These answers suggest that money illusion is prevalent. In other words, people (even Princeton undergraduates) have a hard time adjusting for inflation.[1]

Inflation Variability

Yet another cost comes from the fact that higher inflation is typically associated with *more variable inflation*. And more variable inflation means that financial assets such as bonds, which promise fixed nominal payments in the future, become riskier.

Take a bond that pays $1,000 in 10 years. With constant inflation over the next 10 years, not only the nominal value but also the real value of the bond in 10 years is known with certainty—we can compute exactly how much a dollar will be worth in 10 years. But with variable inflation, the real value of $1,000 in 10 years becomes uncertain. The more variability there is, the more uncertainty it creates. Saving for retirement becomes more difficult. For those who have invested in bonds, lower inflation than they expected means a better retirement; but higher inflation may mean poverty. This is one of the reasons retirees, for whom part of their income is fixed in dollar terms, typically worry more about inflation than other groups in the population.

A good, and sad, movie about surviving on a fixed pension in post–World War II Italy: *Umberto D*, by Vittorio de Sica (1952).

You might argue, as in the case of taxes, that these costs are not due to inflation per se but rather to the financial markets' inability to provide assets that protect their holders against inflation. Rather than issuing only nominal bonds (which promise a fixed nominal amount in the future), governments or firms could also issue *indexed bonds*—bonds that promise a nominal amount adjusted for inflation so people do not have to worry about the real value of the bond when they retire. Indeed, as we saw in Chapter 14,

[1]Source: *Eldar Shafir, Peter Diamond, and Amos Tversky, "Money Illusion," in D. Kahneman and A. Tversky, eds.,* Choices, Values, and Frames *(New York: Cambridge University Press & Russell Sage Foundation, 2000).*

several countries, including the United States, have introduced such bonds so people can better protect themselves against movements in inflation.

The Benefits of Inflation

This may surprise you, but inflation is not all bad. There are three benefits of inflation: (1) seignorage, (2) (somewhat paradoxically) the use of the interaction between money illusion and inflation in facilitating real wage adjustments, and (3) the option of negative real interest rates for macroeconomic policy.

Seignorage

Money creation—the ultimate source of inflation—is one of the ways the government can finance its spending. Put another way, money creation is an alternative to borrowing from the public or raising taxes.

As we saw in Chapter 22, the government typically does not "create" money to pay for its spending. Rather, the government issues and sells bonds and spends the proceeds. But if the bonds are bought by the central bank, which then creates money to pay for them, the result is the same. Other things being equal, the revenues from money creation—that is, *seignorage*—allow the government to borrow less from the public or to lower taxes.

Recall equation (22.6): Let H denote the monetary base—the money issued by the central bank. Then

$$\frac{\text{Seignorage}}{Y} = \frac{\Delta H}{PY} = \frac{\Delta H}{H}\frac{H}{PY}$$

where $\Delta H/H$ is the rate of growth of the monetary base, and H/PY is the ratio of the monetary base to nominal GDP.

How large is seignorage in practice? During hyperinflations, seignorage often becomes an important source of government finance. But its importance in OECD economies today, and for the range of inflation rates we are considering, is much more limited. Take the case of the United States. The ratio of the non-interest-paying money issued by the Fed to GDP is around 6%. An increase in the rate of nominal money growth of 4% per year (which eventually leads to a 4% increase in the inflation rate) would lead therefore to an increase in seignorage of 4% × 6%, or 0.24% of GDP. This is a small amount of revenue in exchange for 4% higher inflation.

Note that I say "non–interest-paying money." This is because one of the changes triggered by the crisis is that many central banks now issue both non-interest-paying money and interest-paying money. More on this in Section 23-4. Given that interest-paying money balances pay an interest rate similar to that on bonds, they are like bonds and do not yield seignorage revenue.

Therefore, while the seignorage argument is sometimes relevant (for example, in economies that do not have a good tax collection system), it hardly seems relevant in the discussion of whether OECD countries today should have, say, 0% versus 4% inflation.

Money Illusion Revisited

Paradoxically, the presence of money illusion provides at least one argument *for* having a positive inflation rate.

To see why, consider two situations: In the first, inflation is 4% and your wage goes up by 1% in nominal terms—in dollars. In the second, inflation is 0% and your wage goes down by 3% in nominal terms. Both lead to the same 3% decrease in your real wage, so you should be indifferent. The evidence, however, shows that many people will accept the real wage cut more readily in the first case than in the second case.

Why is this example relevant to our discussion? As we saw in Chapter 13, the constant process of change that characterizes modern economies means some workers must sometimes take a real pay cut. Thus, the argument goes, the presence of inflation allows for these downward real wage adjustments more easily than if inflation is equal to zero. The evidence on the distribution of wage changes in Portugal under high and low inflation in Chapter 8 suggests that this is indeed a relevant argument.

A conflict of metaphors: Because inflation makes these real wage adjustments easier to achieve, some economists say inflation "greases the wheels" of the economy. Others, emphasizing the adverse effects of inflation on relative prices, say that inflation "puts sand" in the workings of the economy.

The Option Of Negative Real Interest Rates

Higher inflation decreases the probability of hitting the zero lower bound. This argument, which may be the most important, follows from our discussion of the zero lower bound in Chapter 4. A numerical example will help here.

- Consider two economies, both with a natural real interest rate equal to 2%.
- In the first economy, the central bank maintains an average inflation rate of 4%, so the nominal interest rate is on average equal to 2% + 4% = 6%.
- In the second economy, the central bank maintains an average inflation rate of 0%, so the nominal interest rate is on average equal to 2% + 0% = 2%.
- Suppose both economies are hit by a similar adverse shock, which leads, at a given interest rate, to a decrease in spending and a decrease in output in the short run.
- In the first economy, the central bank can decrease the nominal interest rate from 6% to 0% before it hits the liquidity trap, thus achieving a decrease of 6%. Under the assumption that expected inflation does not change immediately and remains equal to 4%, the real interest rate decreases from 2% to −4%.This is likely to have a strong positive effect on spending and help the economy recover.
- In the second economy, the central bank can decrease the nominal interest rate only from 2% to 0%, a decrease of 2%. Under the assumption that expected inflation does not change right away and remains equal to 0%, the real interest rate decreases only by 2%, from 2% to 0%. This small decrease in the real interest rate may not increase spending by very much.

In short, an economy with a higher average inflation rate has more room to use monetary policy to fight a recession. An economy with a low average inflation rate may find itself unable to use monetary policy to return output to the natural level of output. As we saw in Chapter 6, this possibility is far from being just theoretical. At the start of the crisis, central banks quickly hit the zero lower bound, unable to decrease interest rates further. With this experience in mind, the question is whether this should lead central banks to choose higher average inflation in the future. Some economists argue that the Great Recession was an exceptional event, that it is unlikely that countries will face a liquidity trap again in the future, and so there is no need to adopt a higher average inflation rate. Others, and this includes me, argue that the problems faced by a country in a liquidity trap are so serious that we should avoid taking the risk that it happens again, and that a higher target rate of inflation is in fact justified. What is undisputed, though, is that permanently low inflation reduces the central bank's ability to affect the real interest rate.

The Optimal Inflation Rate: The State of the Debate

At this stage, most central banks in advanced economies have an inflation target of about 2%. But they are being challenged on two fronts: Some economists want to achieve price stability—that is, 0% inflation. Others want, instead, a higher target rate of inflation, say 4%.

Those who want to aim for 0% make the point that 0% is a different target rate from all others; it corresponds to price stability. This is desirable in itself. Knowing that the price level will be roughly the same in 10 or 20 years as it is today simplifies a number of complicated decisions and eliminates the scope for money illusion. Also, given the time consistency problem facing central banks (discussed in Chapter 21), the credibility and simplicity of the target inflation rate are important. Some economists and some central bankers believe price stability—that is, a 0% target—can achieve these goals better than a target inflation rate of 2%. So far, however, no central bank has actually adopted a 0% inflation target.

Those who want to aim for a higher rate argue that it is essential not to fall in the liquidity trap in the future, and that, for these purposes, a higher target rate of inflation, say 4%, would be helpful. They argue that the choice of a 2% target was based on the belief that countries would be unlikely to hit the zero lower bound, and that this belief has proven false. Their argument has gained little support among central bankers, who argue that if central banks increase their target from its current value of 2% to 4%, people may start anticipating that the target will soon become 5%, then 6%, and so on, and inflation expectations will no longer be anchored. Thus, they see it as essential to keep current target levels.

This reasoning, which applies to many other issues, is called the "slippery slope" argument.

The debate goes on. For the time being, most central banks continue to aim for low but positive inflation—that is, inflation rates of about 2%.

23-4 UNCONVENTIONAL MONETARY POLICY

When, at the start of the crisis, the interest rate reached the zero lower bound, central banks were unable to decrease it further and thus lost the use of **conventional monetary policy**. In this book, I have assumed until now that monetary policy became impotent. But this was a simplification. Central banks explored other ways to affect activity, a set of measures known as **unconventional monetary policy**.

The idea was simple. While the policy rate was equal to zero, other interest rates remained positive, reflecting various risk premiums. Although I introduced a risk premium in Chapter 6 in the relation of the borrowing rate to the policy rate, I did not discuss in detail what it depended on and how it could be affected by monetary policy. In fact, we can think of the premium on an asset as determined by supply and demand for the asset. If the demand for an asset decreases, whether because buyers become more risk averse or because some investors just decide not to hold the asset, the premium will increase. If, instead, the demand increases, the premium will decrease. This is true whether the increased demand comes from private investors or from the central bank.

This is the logic that led central banks to buy assets (other than short-term bonds) with the intention of decreasing the premium on them and thus decreasing the corresponding borrowing rates with the aim of stimulating economic activity. They did this by financing their purchases through money creation, leading to a large increase in the money supply. Although the increase in the money supply had no effect on the policy rate, the purchase of these other assets decreased their premium, leading to lower borrowing rates and higher spending. These purchase programs are known as **quantitative easing**, or **credit easing**, policies.

In the United States, the Fed started its first quantitative easing program in November 2008, even before it had reached the zero lower bound. In what has become known as Quantitative Easing 1 (**QE1** for short), the Fed started buying certain types of mortgage-based securities. We saw the reason for it in Chapter 6: One of the triggers of the crisis was the difficulty of assessing the value of the underlying mortgages on which those securities were based; as a result, many investors had decided to stop holding any kind of mortgage-based security, and the premium even on securities that seemed relatively safe had jumped to very high levels. By buying these securities, the Fed decreased their premium and limited the effect on the financial system and on spending. The second quantitative easing program, known as **QE2**, started in November 2010, when the Fed starting buying longer-term Treasury bonds, with the intent of decreasing the term premium on these long-term bonds. The third quantitative easing program, **QE3**, started in September 2012, with the further purchase of mortgage-based securities, to decrease the cost of mortgages and further help the housing market to recover.

Much research has gone into assessing the effectiveness of quantitative easing in reducing risk premia. There is wide agreement that QE1 made a large difference.

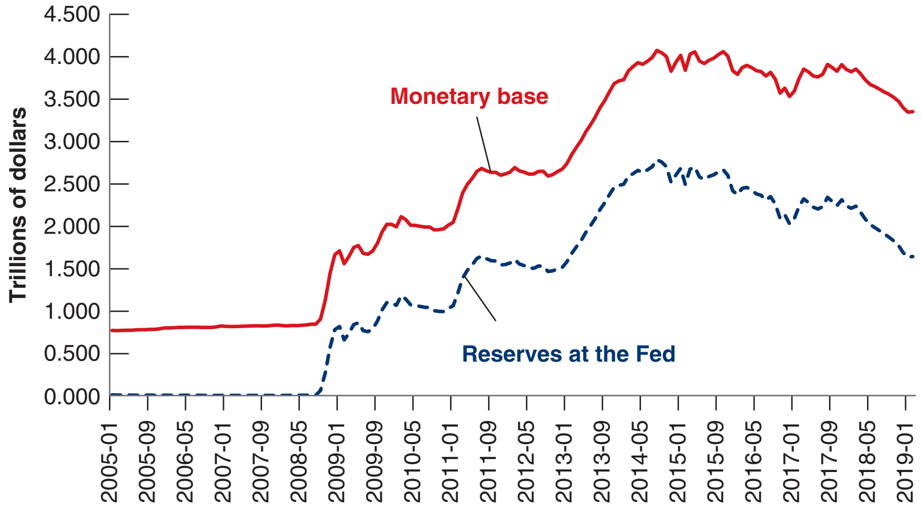


Figure 23-2

The Evolution of the US Monetary Base from 2005 to 2018

As a result of quantitative easing, the monetary base more than quadrupled between 2005 and 2015.

Source: FRED: RESBALS BOGMBASE

By intervening in a market that had become dysfunctional, the Fed limited the increase in premiums. The effects of QE2 and QE3, in which the Fed intervened in markets that were no longer dysfunctional, are more controversial. It is widely accepted that they decreased the term premium on long-term government bonds. The question is by how much.

The general assessment of quantitative easing policies, in the United States and elsewhere, is that they had some effect on borrowing rates, and thus monetary policy can still have some effect on activity even at the zero lower bound. But there is also wide agreement that they work in more complicated and less reliable ways than conventional monetary policy. Put another way, the zero lower bound may not make monetary policy impotent, but it surely limits its efficiency.

We have looked at what the Fed bought. How did it finance itself? By issuing money. As the interest rate was (nearly) equal to zero, banks were indifferent between holding bonds or holding money. So they were willing to accumulate central bank money in the form of reserves at the Fed. (Actually the Fed was willing to pay a slightly positive interest rate, 0.25%, which made it attractive for banks to hold reserves.) As a result, the monetary base, i.e., the money issued by the central bank increased from $850 billion, or about 6.6% of GDP in September 2008, to $4,000 billion, or about 22% of GDP in 2015. This evolution is represented by the red line in Figure 23-2. The blue line shows the reserves held by banks; most of the increase was reflected in an increase in bank reserves at the Fed. As a result, the balance sheet of the Fed is much larger than it was before the crisis.

Monetary Policy since the End of the Liquidity Trap

When the Fed decided at the end of 2015 that the economy was stronger and it should increase the federal funds rate again, it faced an obvious problem. If it continued to pay zero interest on reserves (or, more precisely, to pay a very low one), banks would not be willing to hold reserves. They would try to sell all of them in the federal funds market, and, unless the Fed was willing to buy back all of them—and by implication, to sell all the bonds it had accumulated as a result of QE—the federal funds rate would go back to zero. If the Fed did not want this to happen, it had to start paying interest on reserves. And this is what it did, and still does today.

The Fed did not want to sell all the assets it had acquired during the crisis right away. It thought, rightly, that this would lead to large disruptions in the asset markets, and selling should happen slowly over time, if at all.

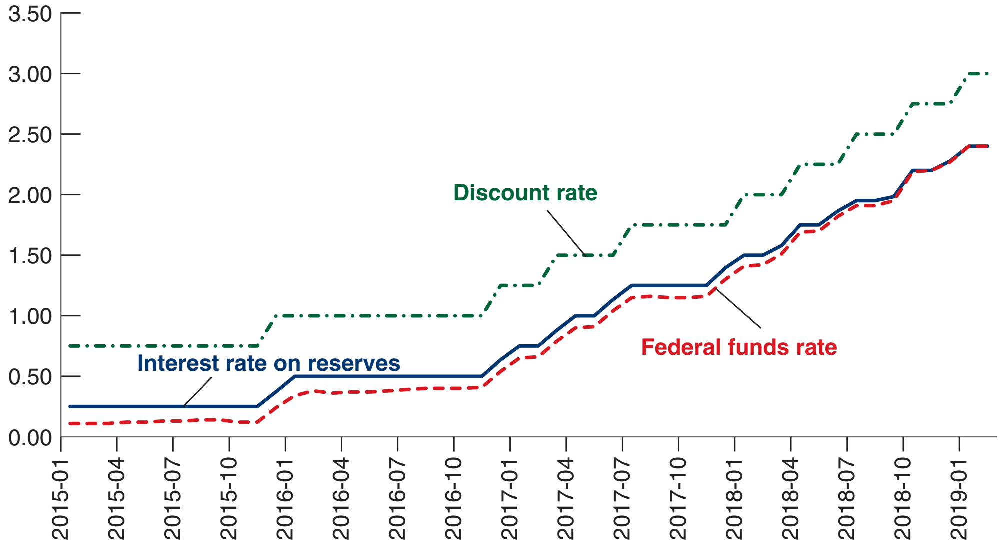
The Fed now operates a **corridor system**. It sets two rates: the rate on reserves, the rate at which banks can in effect lend to the Fed; and the discount rate, the rate at which banks can borrow from the Fed. The federal funds rate is still determined in the federal funds market, but it must be in the corridor created by the two rates. To see why, suppose

Figure 23-3

The Interest Rate on Reserves, the Federal Funds Rate, and the Discount Rate since 2015

The federal funds rate has moved in line with the interest rate on reserves.

Source: FRED: IOER, FEDFUNDS, INTDSRUSM193N



Contrary to the argument in the text, you can see that the federal funds rate has remained slightly below the interest rate on reserves! This is because of a technical issue: Some financial institutions do not have access to Fed reserves and are willing to lend at a lower rate than the rate on reserves. The general message remains: The floor on the federal funds rate is mostly determined by the rate on reserves.

that the federal funds rate was lower than the rate on reserves: No bank would be willing to lend to another bank, as lending to the Fed would give a higher rate. Suppose instead that the federal funds rate was higher than the discount rate: no bank would want to borrow from another bank, as borrowing from the Fed would give a lower rate. Thus, while the federal funds rate is still determined in the federal funds market, it can vary only within the corridor created by the two rates. The evolution of the three rates is shown in Figure 23-3, and it shows how the federal funds rate has moved with the interest rate paid on reserves.

To summarize: Monetary policy is in flux. The crisis forced central bankers in general, and the Fed in particular, to question the focus on inflation targeting and to explore tools other than the policy rate. As a result of the use of unconventional policies during the zero lower bound period, central banks' balance sheets are much larger than they used to be. They hold a much larger quantity of assets than before the crisis and have much larger liabilities. They pay interest on a large part of their liabilities, the reserves held by banks at the central bank. The main policy tool has become the interest rate paid to banks on their reserves at the Fed. The main choice facing central banks at this point is whether to keep their large balance sheets, or to return to a situation closer to the precrisis situation, with a small balance sheet and mostly non–interest-paying liabilities. While they have in most cases either stabilized or started to decrease their balance sheets, the endpoint has not been decided yet.

23-5 MONETARY POLICY AND FINANCIAL STABILITY

When the financial crisis started, central banks were confronted not only with a major decline in demand but also with serious problems in the financial system. As we saw in Chapter 6, the decline in housing prices had been the trigger for the crisis, which was then amplified by failures of the financial system. Opacity of assets led to doubts about the solvency of financial institutions, and these doubts led to runs, in which investors tried to get their funds back, forcing fire sales and generating further doubts about solvency. The first urgent issue facing the central banks was thus what measures to take—beyond the measures already described in the previous sections. The second issue was whether and how, in the future, monetary policy should try to decrease the probability of another such financial crisis. We look at both issues in turn.

Liquidity Provision and Lender of Last Resort

Central banks have long known about bank runs. As we saw in Chapter 6, the structure of the balance sheet of banks exposes them to runs. Many of their assets, such as loans, are illiquid. Many of their liabilities, such as demand deposits, are liquid; as the term indicates, demand deposits can be withdrawn *on demand*. Thus, depositors' worries, founded or unfounded, can lead them to want to withdraw their funds, forcing the bank either to close or to sell the assets at fire sale prices. In most countries, two measures have traditionally been taken to limit such runs:

- Deposit insurance, which gives investors the confidence that they will get their funds back even if the bank is insolvent, so that they do not have an incentive to run.
- And, in case the run actually happens, the provision of liquidity by the central bank to the bank against some collateral, namely some of the assets of the bank. This way, the bank can get the liquidity it needs to pay the depositors without having to sell the assets. This function of the central bank is known as **lender of last resort**, and it has been one of the functions of the Fed since its creation in 1913.

What the crisis showed, however, was that banks were not the only financial institutions that could be subject to runs. Any institution whose assets are less liquid than its liabilities is exposed to similar risks of a run. If investors want their funds back, it may be difficult for the financial institution to get the liquidity it needs. Given the urgency during the crisis, the Fed extended liquidity provision to some financial institutions other than banks. It had little choice other than to do so, but, looking forward, the question is what the rules should be—which institutions can expect to receive liquidity from the central bank and which cannot. The question is far from settled. Do central banks really want to provide such liquidity to institutions they do not regulate?

Macroprudential Tools

Starting in the mid-2000s, the Fed became worried about the increase in housing prices. But the Fed and other central banks facing similar housing price increases were reluctant to intervene. This was for a number of reasons: First, they found it difficult to assess whether the price increases reflected increases in fundamentals (e.g., low interest rates) or a bubble (i.e., increases in prices above what were justified by fundamentals). Second, they worried that an increase in the interest rate, although it might indeed stop the increase in housing prices, would also slow the whole economy and trigger a recession. Third, they thought that, even if the increase in housing prices was indeed a bubble, and the bubble were to burst and lead to a decrease in housing prices later, they could counter the adverse effects on demand through an appropriate decrease in the interest rate.

The crisis has forced them to reconsider. As we saw, housing price declines combined with the buildup of risk in the financial system led to a major financial and macroeconomic crisis, which they could neither avoid nor counter.

As a result, a broad consensus is emerging along two lines:

- It is risky to wait. Even if there is doubt about whether an increase in asset prices reflects fundamentals or a bubble, it may be better to do something than nothing. Better to stand for a while in the way of a fundamental increase and turn out to be wrong than to let a bubble build up and burst, with major adverse macroeconomic effects. The same applies to buildups of financial risk; for example, excessive bank leverage. Better to prevent high leverage, even at the risk of decreasing bank credit, than to allow it to build up, increasing the risk of a financial crisis.

This has led to the dictum: Better lean [against increases in asset prices] than clean [up after asset prices have crashed].

- To deal with bubbles, credit booms, or dangerous behavior in the financial system, the interest rate is not, however, the right policy instrument. It is too blunt a tool, affecting the whole economy rather than resolving the problem at hand. The right instruments are **macroprudential tools**, rules that are aimed directly at borrowers, or lenders, or banks and other financial institutions, as the case may require.

What form might some of the macroprudential tools take? Some tools may be aimed at borrowers:

- Suppose the central bank is worried about what it perceives to be an excessive increase in housing prices. It can tighten conditions under which borrowers can obtain mortgages. A measure used in many countries is a ceiling on the size of the loan borrowers can take relative to the value of the house they buy, a measure known as the maximum **loan-to-value (LTV) ratio**, or maximum LTV for short. Reducing the maximum LTV is likely to decrease demand and thus slow down the price increase. (The Focus Box "LTV Ratios and Housing Price Increases from 2000 to 2007" examines the relation between maximum LTVs and housing price increases in the period leading up to the crisis.)
- Suppose the central bank is worried that people are borrowing too much in foreign currency. An example will help to make the point. In the early 2010s, more than two-thirds of mortgages in Hungary were denominated in Swiss francs! The reason was simple. Swiss interest rates were very low, making it apparently attractive for Hungarians to borrow at the Swiss rather than the Hungarian interest rate. The risk that borrowers did not consider, however, was that the Hungarian currency, the forint, would depreciate vis-à-vis the Swiss franc. Such a depreciation took place, increasing, on average, the real value of the mortgages Hungarians had to pay by more than 50%. Many households could no longer make their mortgage payments. This suggests that it would have been wise to put restrictions from the start on the amount of borrowing in foreign currency by households.

This led the Hungarian government to allow for a conversion of mortgages in Swiss francs to mortgages in forints at a better exchange rate. Hungarian households were better off, but the banks that loaned to them were worse off.

Some tools may be aimed at lenders, such as banks or foreign investors:

- Suppose the central bank is worried about an increase in bank leverage. We saw in Chapter 6 why this should be a concern. High leverage was one of the main reasons why housing price declines led to the financial crisis. The central bank can impose minimum capital ratios to limit leverage. These may take various forms (for example, a minimum value for the ratio of capital to all assets, or a minimum value for the ratio of capital to risk-weighted assets, with riskier assets having a higher weight). In fact, in a series of agreements known as **Basel II** and **Basel III**, many countries have agreed to impose the same minimums on their banks. A more difficult and unresolved issue is whether and how such capital ratios should be adjusted over time as a function of economic and financial conditions (whether, for example, they should be increased if there appears to be excessive credit growth).

Go back to Chapter 6 for a refresher on the relation between leverage and capital ratios.

- Suppose the central bank is worried about high capital inflows, as, for example, in the Hungarian case we just discussed. The central bank worries that, although investors are willing to lend at low interest rates to the country, they may change their mind, and this might lead to a sudden stop. The central bank may then want to limit the capital inflows by imposing **capital controls** on inflows. These may take the form of taxes on different types of inflows, with lower taxes on capital flows that are less prone to sudden stops, such as **foreign direct investment** (the purchase of physical assets by foreigners), or a direct limit on the ability of domestic residents to take out foreign loans.

See the Focus Box "Sudden Stops, Safe Havens, and the Limits to the Interest Parity Condition" in Chapter 19.

LTV Ratios and Housing Price Increases from 2000 to 2007

Is it the case that countries that had more stringent restrictions on borrowing had lower housing price increases from 2000 to 2007? An answer is given in Figure 1. The figure, taken from an IMF study, shows the evidence for 21 countries for which the data could be obtained.

The horizontal axis plots the maximum loan-to-value (LTV) ratio on new mortgages across countries. This maximum is not necessarily a legal maximum but may be a guideline, or a limit over which additional requirements, such as mortgage insurance, may be asked of the borrower. A ratio of 100% means that a borrower may be able to get a loan equal to the value of the house. Actual values vary from 60% in Korea; to 100% in many countries, including the United States; to 125% in the Netherlands. The vertical axis plots the increase in the nominal price of housing from 2000 to 2007 (measuring the real price increase would lead to a similar picture). The figure also plots the regression line, the line that best fits the set of observations.

The figure suggests two conclusions:

The first is that there indeed appears to be a positive relation between the LTV ratio and the housing price increase. Korea and Hong Kong, which both imposed low LTV ratios, had smaller housing price increases. Spain and the United Kingdom, with much higher ratios, had much larger price increases.

The second is that the relation is far from tight. This should not come as a surprise, as surely many other factors played a role in the increase in housing prices. But even controlling for other factors, it is difficult to identify with much confidence the precise effect of the LTV ratio. Looking forward, we shall have to learn a lot more about how an LTV-based regulatory tool might work before it can be used as a reliable macroprudential tool.[2]

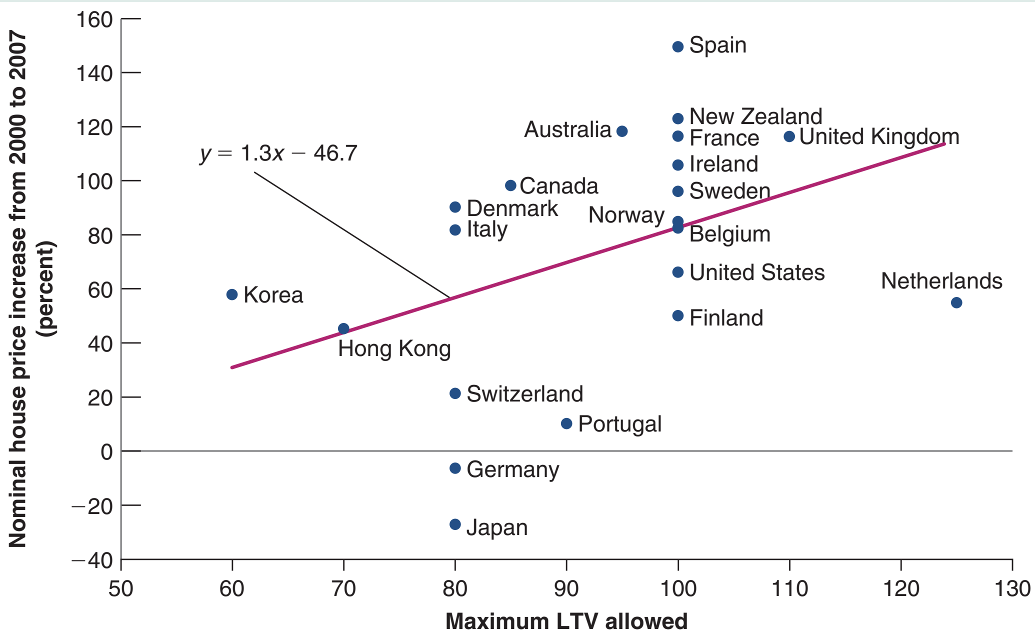


Figure 1

Maximum LTV Ratios and Housing Price Increases, 2000–2007

Although there is wide agreement that the use of such *macroprudential* tools is desirable, many questions remain:

- In many cases, we do not know how well these tools work (e.g., how much a decrease in the maximum LTV ratio affects the demand for housing, or whether foreign investors can find ways of avoiding capital controls).

[2]Source: *Christopher Crowe, Giovanni Dell'Ariccia, Deniz Igan, and Pau Rabanal, "Policies for Macrofinancial Stability: Options to Deal with Real Estate Booms," Staff Discussion Note, International Monetary Fund, February 2011.*

- There are likely to be complex interactions between the traditional monetary policy tools and these macroprudential tools. For example, there is some evidence that very low interest rates lead to excessive risk taking, whether by investors or by financial institutions. If this is the case, a central bank that decides, for macroeconomic reasons, to lower the interest rate may have to use various macroprudential tools to offset the potential increase in risk taking. Again, we know little about how best to do it.
- The question arises of whether macroprudential tools, together with traditional monetary policy tools, should be under the control of the central bank or of a separate authority. The argument for having the central bank in charge of both monetary and macroprudential tools is that these tools interact, and thus only one centralized authority can use them in the right way. The argument against it is that such a consolidation of tools may give too much power to an independent central bank.

  At this stage, some countries have taken one route and others have taken another. In the United Kingdom, the central bank has been given power over both monetary and macroprudential tools. In the United States, the responsibility has been given to a council under the formal authority of the US Treasury, but with the Fed playing a major role within the council.

To summarize: The crisis has shown that macroeconomic stability requires the use not only of traditional monetary instruments but also of macroprudential tools. How best to use them is one of the challenges facing macroeconomic policymakers today.

SUMMARY

- Until the 1980s, the design of monetary policy focused on nominal money growth. But because of the poor relation between inflation and nominal money growth, this approach was abandoned by most central banks.
- Central banks now focus on an inflation rate target rather than a nominal money growth rate target. And they think about monetary policy in terms of choosing the nominal interest rate rather than choosing the rate of nominal money growth.
- The Taylor rule gives a useful way of thinking about the choice of the nominal interest rate. The rule states that the central bank should move its interest rate in response to two main factors: the deviation of the inflation rate from the target rate of inflation, and the deviation of the unemployment rate from the natural rate of unemployment. A central bank that follows this rule will stabilize activity and achieve its target inflation rate in the medium run.
- The optimal rate of inflation depends on the costs and benefits of inflation. Higher inflation leads to more distortions, especially when it interacts with the tax system. On the other hand, because higher inflation implies higher average nominal interest rates, it decreases the probability of hitting the zero lower bound, a bound that was costly in the recent crisis.
- When advanced economies hit the zero lower bound during the Great Financial Crisis, central banks explored unconventional monetary policy tools, such as quantitative easing. These policies worked through the effects of central bank purchases on the risk premiums associated with different assets. These purchases led to large increases in the balance sheets of central banks. The issue is now whether the central banks should reduce those balance sheets, and whether unconventional measures should be used in normal times.
- The crisis has shown that stable inflation is not a sufficient condition for macroeconomic stability. This is leading central banks to explore the use of macroprudential tools. These tools can, in principle, help limit bubbles, control credit growth, and decrease risk in the financial system. How best to use them, however, is still poorly understood and is one the challenges facing monetary policy today.

KEY TERMS

inflation targeting, 485
interest rate rule, 485
Great Moderation, 485
M1, 487
divine coincidence, 489
flexible inflation targeting, 489
Taylor rule, 490
shoe-leather costs, 491
bracket creep, 492
money illusion, 492
conventional monetary policy, 496
unconventional monetary policy, 496
quantitative easing, 496
credit easing, 496
QE1, 496
QE2, 496
QE3, 496
corridor system, 497
lender of last resort, 499
macroprudential tools, 500
loan-to-value (LTV) ratio, 500
Basel II, 500
Basel III, 500
capital controls, 500
foreign direct investment, 500

QUESTIONS AND PROBLEMS

QUICK CHECK

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.

a. The most important argument in favor of a positive rate of inflation in OECD countries is seignorage.
b. Fighting inflation should be the Fed's only purpose.
c. Inflation and money growth moved together from 1970 to 2009.
d. Because most people have little trouble distinguishing between nominal and real values, inflation does not distort decision making.
e. Most central banks around the world have an inflation target of 4%.
f. The higher the inflation rate, the higher the effective tax rate on capital gains.
g. The Taylor rule describes how central banks adjust the policy interest rate across recessions and booms.
h. The zero lower bound on the nominal policy rate was expected to be a regular feature of monetary policy when inflation targeting began.
i. The Federal Reserve pays interest on reserves held by member banks.
j. Quantitative easing refers to central bank purchases of assets with the intention of directly affecting the yield on these assets.
k. In the crisis, central banks provided liquidity to financial institutions they did not regulate.
l. One consequence of the crisis was higher capital requirements and a more extensive regulatory regime for banks.

2. Breaking the link between money growth and inflation in the medium run

The money demand relationship in Chapter 4 is used implicitly in Figure 23-1. That relation is

$$\frac{M}{P} = YL(i)$$

The central bank in conjunction with the political authorities chooses an inflation target π^*.

a. Derive the target nominal interest rate in a medium-run equilibrium.
b. Consider medium-run equilibria where potential output does not grow. Derive the relation between money growth and inflation. Explain.
c. Now consider medium-run equilibria where potential output grows at 3% per year. Now derive the relation between money growth and inflation. Do you expect inflation to be higher or lower than money growth? Explain.
d. Consider Figure 23-1. Look first at the period ending in roughly 1995. How do your results in parts b and c relate to it?
e. Focus on the case where all money is currency. We can then think of money demand as being the demand for currency (you can refer back to the appendix to Chapter 4 if needed). Over the past 50 years:
   i. Automatic tellers have allowed cash to be dispensed outside of regular banking hours.
   ii. The use of credit cards for purchases has greatly expanded.
   iii. The use of debit cards for purchases has greatly expanded.
   iv. Most recently, technology has allowed for small purchases by credit and debit cards by waving the card over a payment terminal near the cash register.

   How would each of these innovations affect the demand for currency?
f. The FRED database at the Federal Reserve Bank of St. Louis has an annual series for currency (CURRVALALL). Download this series and the annual series for nominal GDP (GDPA). Construct a ratio of currency to nominal GDP. How does this series behave from 1993 to 2018? Are you surprised?

3. The inflation target: Nearly every major central bank has chosen an inflation target of 2%.

a. Why might a central bank choose a lower inflation target, for example, zero inflation?

b. Why might a central bank choose a higher inflation target, for example, 4% inflation?

4. Indexed bonds and inflation uncertainty

In Chapter 14, in the Focus Box titled "The Vocabulary of Bond Markets," the concept of an inflation-indexed bond was introduced. Although such bonds are typically long in maturity, the example that follows compares a standard one-year Treasury bill with an inflation-indexed one-year Treasury bill.

a. A standard one-year $100 Treasury bill promises to pay $100 in one year and sells for $\$P_B$ (notation is from Chapter 4) today. What is the nominal interest rate on the Treasury bill?
b. Suppose that the price level is P today and $P(t+1)$ next year and the bill sells for $\$P_B$ today. What is the real interest rate on the Treasury bill?
c. An indexed Treasury bill pays a larger payment next year to compensate for inflation between the date of issue and the date of payment. If the bill is issued today when the price index is 100, what will be the payment next year if the price index has risen to 110? What is the real interest rate on an indexed Treasury bill that sells for $\$P_B$ today?
d. If you are an investor, will you want to hold indexed or non-indexed bonds?

5. Unwinding unconventional monetary policy

It was noted in the text that the Federal Reserve purchased, in addition to Treasury bills, large amounts of mortgage-backed securities and long-term government bonds as part of quantitative easing. Figure 23-2 shows that in 2015, there were about $4 trillion of assets in the monetary base. By 2018, total assets fell to $3.5 trillion. The table below presents some further detail on three types of Federal Reserve assets in billions of dollars.

Asset category	End of 2015	End of 2017
Treasury securities less than one year to maturity	216.1	443.7
Treasury securities with more than one year to maturity	2245.4	2010.5
Mortgage-backed securities	1747.5	1764.9

Source: Annual Reports of the Federal Reserve Board, Table 2, Federal Reserve Bank holdings of US Treasury and federal agency securities

a. Why did the Federal Reserve Board buy the mortgage-backed securities?
b. Why did the Federal Reserve Board buy the long-term Treasury bonds?
c. What would you predict as the consequences of the following operation by the Federal Reserve Board: selling $0.5 trillion in mortgage-backed securities and buying $0.5 trillion in Treasury securities with less than one year to maturity?
d. What would you predict as the consequences of the following operation by the Federal Reserve Board: selling $0.5 trillion in Treasury securities with maturity longer than one year and buying $0.5 trillion in Treasury securities with less than one year to maturity?
e. How did the Fed rearrange its balance sheet between 2015 and 2017? Is there evidence of unwinding of quantitative easing?

6. The maximum loan-to-value ratio

Most home-buyers purchase their home with a combination of a cash down payment and a mortgage. The loan-to-value ratio is a rule that establishes the maximum mortgage loan allowed on a home purchase.

a. If a home costs $300,000 and the maximum loan-to-value ratio is 80% as in Denmark, what is the minimum down payment?
b. If the maximum loan-to-value is reduced, how will this affect the demand for homes?
c. In Chapter 14 you were referred to *The Economist House Price Index*. Find that index and look at the behavior of house prices in Canada and the United States from 1970 to the latest date available. On December 10, 2015, the Canadian Minister of Finance announced an increase in the minimum down payment on any portion of a mortgage more than $500,000. (The announcement can be found at www.fin.gc.ca/n15/15-088-eng.asp.) Why was this action taken? Do you see an effect on house prices in Canada? What do you conclude?

7. Bank behavior in the interest rate corridor

The United States (unlike other countries) has two types of bank-like financial institutions. Member banks can borrow from the Federal Reserve at the discount rate and must keep currency in their vaults or deposits at the Federal Reserve earning the deposit rate as reserves. These rates are shown in Figure 22-3. Non-member banks can hold their reserves as currency or deposits at member banks. The federal funds rate is the interest rate on one-day loans in financial markets. It is determined by the demand for and the supply of these funds by both member and non-member banks. Chapter 4 introduced both reserve requirements and balance sheets.

a. If the reserve ratio is 10% and a member bank that can place funds on deposit at the Federal Reserve has the balance sheet below, does this bank have excess reserves? If the deposit rate is 0.5% and the federal funds rate is 0.4%, what is the profit-maximizing choice for the overnight placement of excess reserves for the member bank?

Bank	
Assets	Liablities
Currency 60 Deposits at Fed 50 Loans 600 Bonds 290	Checkable deposits 1000

b. If the reserve ratio is 10% and a member bank that can place funds on deposit at the Federal Reserve has the balance sheet below, does this bank have excess reserves? If the discount rate is 0.75% and the federal funds rate is 0.8%, how should the profit-maximizing member bank borrow overnight to meet reserve requirements?

Banks	
Assets	Liablities
Currency 30 Deposits at Fed 50 Loans 600 Bonds 320	Checkable deposits 1000

c. If all banks in America were member banks, explain why the Federal Reserve could be certain that the federal funds rate fell between the deposit rate and the discount rate?

d. In Figure 22-3, the federal funds rate falls slightly below the deposit rate. Explain, using the balance sheet below for a non-member bank, how this would happen. Non-member banks hold part of their required reserves as deposits at member banks.

Non-member bank	
Assets	Liablities
Currency 50 Deposits at member bank 20 Loans 330 Bonds 100	Checkable deposits 500

DIG DEEPER

8. Taxes, inflation, and home ownership

In this chapter, we discussed the effect of inflation on the effective capital-gains tax rate on the sale of a home. In this question, we explore the effect of inflation on another feature of the tax code—the deductibility of mortgage interest.

Suppose you have a mortgage of $50,000. Expected inflation is π^e, and the nominal interest rate on your mortgage is i. Consider two cases:

i. $\pi^e = 0\%;\ \ i = 4\%$
ii. $\pi^e = 10\%;\ \ i = 14\%$

a. What is the real interest rate you are paying on your mortgage in each case?

b. Suppose you can deduct nominal mortgage interest payments from your income before paying income tax. Assume that the tax rate is 25%. So, for each dollar you pay in mortgage interest, you pay 25 cents less in taxes, in effect getting a subsidy from the government for your mortgage costs. Compute, in each case, the real interest rate you are paying on your mortgage, taking this subsidy into account.

c. Considering only the deductibility of mortgage interest (and not capital-gains taxation), is inflation good for homeowners

9. Suppose you have been elected to Congress. One day, one of your colleagues makes the following statement:

> *The Fed chair is the most powerful economic policy maker in the United States. We should not turn over the keys to the economy to someone who was not elected and therefore has no accountability. Congress should impose an explicit Taylor rule on the Fed. Congress should choose not only the target inflation rate but the relative weight on the inflation and unemployment targets. Why should the preferences of an individual substitute for the will of the people, as expressed through the democratic and legislative processes?*

Do you agree with your colleague? Discuss the advantages and disadvantages of imposing an explicit Taylor rule on the Fed.

EXPLORE FURTHER

10. The frequency of the zero lower bound around the world

Use the FRED database at the Federal Reserve Bank of St. Louis to find the monthly average nominal policy interest rates for four major central banks. The series for these rates are: United States, federal funds (FEDFUNDS); United Kingdom, (INTDSRGBM193N); European Central Bank (covering the Euro-zone countries including Italy, France, and Germany), immediate rate on Euro (IRSTCI01EZM156N); Bank of Japan, immediate rate on yen (IRSTCI01JPM156N); Bank of Canada, immediate rate (IRSTCB01CAM156N).

Which of these central banks has spent a significant period of time at the zero lower bound since 2000?

11. Current monetary policy

Problem 10 in Chapter 4 asked you to consider the current stance of monetary policy. Here, you are asked to do so again, but with the additional understanding of monetary policy you have gained in this and previous chapters.

Go to the Web site of the Federal Reserve Board of Governors (www.federalreserve.gov) and download either the press release you considered in Chapter 4 (if you did Problem 10) or the most recent press release of the Federal Open Market Committee (FOMC).

a. What is the stance of monetary policy as described in the press release?

b. Is there evidence that the FOMC considers both inflation and unemployment when setting interest rate policy as would be implied by the Taylor rule?

c. Does the language make specific reference to a target for inflation?

d. Does the language make specific reference to the natural or target real rate of interest?

e. Does the language raise any issues related to macroprudential regulation of financial institutions?

FURTHER READINGS

- For an early statement of inflation targeting, read "Inflation Targeting: A New Framework for Monetary Policy?" by Ben Bernanke and Frederic Mishkin, *Journal of Economic Perspectives*, 1997, Vol. 11 (Spring): pp. 97–116. (This article was written by Ben Bernanke before he became Chairman of the Fed.)
- For more institutional details on how the Fed actually functions, see www.federalreserve.gov/aboutthefed/default.htm.
- A time frame giving financial developments and the actions of the Fed from 2008 to 2011 is given at www.nytimes.com/interactive/2008/09/27/business/economy/20080927_WEEKS_TIMELINE.html.
- A great long read: The description of the problems in the financial sector and of US monetary policy during the crisis by the Fed Chair himself, in *The Courage to Act: A Memoir of a Crisis and Its Aftermath* by Ben Bernanke, W. W. Norton & Co., Inc., 2015.

24

Epilogue: The Story of Macroeconomics

We have spent 23 chapters presenting the framework that most economists use to think about macroeconomic issues, the major conclusions they draw, and the issues on which they disagree. How this framework has been built over time is a fascinating story. It is the story I want to tell in this chapter.

Section 24-1 starts at the beginning of modern macroeconomics—with Keynes and the Great Depression.

Section 24-2 turns to the *neoclassical synthesis*, a synthesis of Keynes's ideas with those of earlier economists—a synthesis that dominated macroeconomics until the early 1970s.

Section 24-3 describes the *rational expectations critique*, the strong attack on the neoclassical synthesis that led to a complete overhaul of macroeconomics starting in the 1970s.

Section 24-4 gives you a sense of the main lines of research in macroeconomics up to the crisis.

Section 24-5 takes a first pass at assessing the effects of the crisis on macroeconomics.

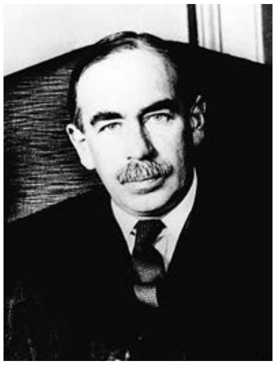
If you remember one basic message from this chapter, it should be: Modern macroeconomics is the result of a long and rich process of construction, crises, and reconstruction. ▶▶▶

24-1 KEYNES AND THE GREAT DEPRESSION

Be warned: The picture gallery is overwhelmingly one of white males. This unfortunately reflects the history of the field. The good news is that things are changing, and women play an increasing role. The American Economic Association Prize for the best economist under 40 was given, in 2019, to a woman, Emi Nakamura, a macroeconomist at Berkeley.

The history of modern macroeconomics starts in 1936, with the publication of John Maynard Keynes's *General Theory of Employment, Interest, and Money*. As he was writing the *General Theory*, Keynes confided to a friend: "I believe myself to be writing a book on economic theory which will largely revolutionize—not, I suppose at once but in the course of the next ten years, the way the world thinks about economic problems."[1]

Keynes was right. The book's timing was one of the reasons for its immediate success. The Great Depression was not only an economic catastrophe but also an intellectual failure for the economists working on **business cycle theory**—as macroeconomics was then called. Few economists had a coherent explanation for the Depression, either its depth or its length. The economic measures taken by the Roosevelt administration as part of the New Deal had been based on instinct rather than on economic theory. The *General Theory* offered an interpretation of events, an intellectual framework, and a clear argument for government intervention.



John Maynard Keynes

The *General Theory* emphasized **effective demand**—what we now call *aggregate demand.* In the short run, Keynes argued, effective demand determines output. Even if output eventually returns to potential, the process is slow at best. One of Keynes's most famous quotes is: "In the long run, we are all dead."

In the process of deriving effective demand, Keynes introduced many of the building blocks of modern macroeconomics:

- The relation of consumption to income, and the multiplier, which explains how shocks to demand can be amplified and lead to larger shifts in output.
- **Liquidity preference**, the term Keynes gave to the demand for money, which explains how monetary policy can affect interest rates and aggregate demand.
- The importance of expectations in affecting consumption and investment; and the idea that *animal spirits* (shifts in expectations) are a major factor behind shifts in demand and output.

The *General Theory* was more than a treatise for economists. It offered clear policy recommendations, and they were in tune with the times: Waiting for the economy to recover by itself was irresponsible. In the midst of a depression, trying to balance the budget was not only stupid, it was dangerous. Active use of fiscal policy was essential to return the country to high employment.

24-2 THE NEOCLASSICAL SYNTHESIS

Within a few years, the *General Theory* had transformed macroeconomics. Not everyone was converted, and few agreed with it all. But most discussions became organized around it.

Paul Samuelson

By the early 1950s a large consensus had emerged, based on an integration of many of Keynes's ideas and the ideas of earlier economists. This consensus was called the **neoclassical synthesis**. To quote from Paul Samuelson, in the 1955 edition of his text *Economics*, the first modern economics text:

> "In recent years, 90 percent of American economists have stopped being 'Keynesian economists' or 'Anti-Keynesian economists.' Instead, they have worked toward a synthesis of whatever is valuable in older economics and in modern theories of income determination. The result might be called neo-classical economics and is accepted, in its broad outlines, by all but about five percent of extreme left-wing and right-wing writers."[2]

[1]Source: J. M. Keynes, *The General Theory of Employment, Interest, and Money,* Palgrave Macmillan, 1936.

[2]Source: Paul A. Samuelson, *Economics,* Tata McGraw-Hill Education, 2010.

The neoclassical synthesis was to remain the dominant view for another 20 years. Progress was astonishing, leading many to call the period from the early 1940s to the early 1970s the golden age of macroeconomics.

Franco Modigliani

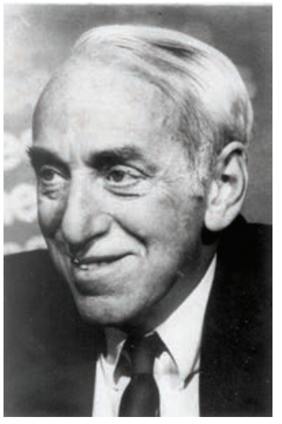
Progress on All Fronts

The first order of business after the publication of the *General Theory* was to formalize mathematically what Keynes meant. Although Keynes knew mathematics, he had avoided using it in the *General Theory*. One result was endless controversies about what Keynes meant and whether there were logical flaws in some of his arguments.

The IS-LM Model

A number of formalizations of Keynes's ideas were offered. The most influential one was the IS-LM model, developed by John Hicks and Alvin Hansen in the 1930s and early 1940s. The initial version of the IS-LM model—which was actually close to the version presented in Chapter 5 of this book—was criticized for emasculating many of Keynes's insights. Expectations played no explicit role, and the adjustment of prices and wages was altogether absent. Yet the IS-LM model provided a basis from which to start building, and as such it was immensely successful. Discussions became organized around the slopes of the IS and LM curves, what variables were missing from the two relations, what equations for prices and wages should be added to the model, and so on.

Theories of Consumption, Investment, and Money Demand



James Tobin

Keynes had emphasized the importance of consumption and investment behavior, and of the choice between money and other financial assets. Major progress was soon made along all three fronts.

In the 1950s, Franco Modigliani (then at Carnegie Mellon, later at MIT) and Milton Friedman (at the University of Chicago) independently developed the theory of consumption we saw in Chapter 15. Both insisted on the importance of expectations in determining current consumption decisions.

James Tobin, at Yale, developed the theory of investment, based on the relation between the present value of profits and investment. The theory was further developed and tested by Dale Jorgenson, at Harvard. You saw this theory in Chapter 15.

Tobin also developed the theory of the demand for money and, more generally, the theory of the choice between different assets based on liquidity, return, and risk. His work has become the basis not only for an improved treatment of financial markets in macroeconomics but also for finance theory in general.

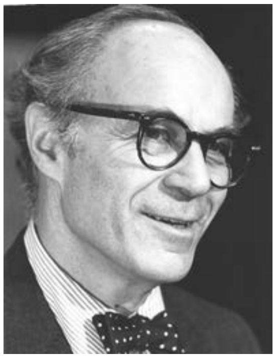
Growth Theory

Robert Solow

In parallel with the work on fluctuations, there was a renewed focus on growth. In contrast to the stagnation in the pre–World War II era, most countries were experiencing rapid growth in the 1950s and 1960s. Even if they experienced fluctuations, their standard of living was increasing rapidly. The growth model developed by MIT's Robert Solow in 1956, which we saw in Chapters 11 and 12, provided a framework to think about the determinants of growth. It was followed by an explosion of work on the roles of saving and technological progress in determining growth.

Macroeconometric Models

All these contributions were integrated in larger and larger macroeconometric models. The first US macroeconometric model, developed by Lawrence Klein at the University of Pennsylvania in the early 1950s, was an extended IS relation, with 16 equations. With the development of the National Income and Product Accounts (making available

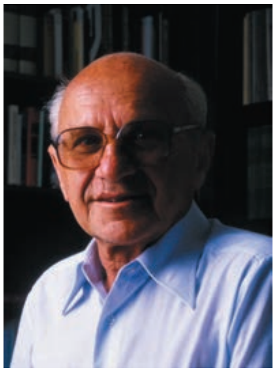


Lawrence Klein

better data) and of both econometrics and computers, the models quickly grew in size. The most impressive effort was the construction of the MPS model (MPS stands for MIT-Penn-SSRC, for the two universities and the research institution—the Social Science Research Council—involved in its construction), developed during the 1960s by a group led by Modigliani. Its structure was an expanded version of the IS-LM model, plus a Phillips curve mechanism. But its components—consumption, investment, and money demand—all reflected the tremendous theoretical and empirical progress since Keynes.

Keynesians versus Monetarists

With such rapid progress, many macroeconomists—those who defined themselves as **Keynesians**—came to believe that the future was bright. The nature of fluctuations was becoming increasingly well understood; the development of models allowed policy decisions to be made more effectively. The time when the economy could be fine-tuned, and recessions all but eliminated, seemed not far in the future.



Milton Friedman

This optimism was met with skepticism by a small but influential minority, the **monetarists**. The intellectual leader of the monetarists was Milton Friedman. Although Friedman saw much progress being made—and was himself the father of one of the major contributions to macroeconomics, the theory of consumption—he did not share in the general enthusiasm. He believed that the understanding of the economy remained very limited. He questioned the motives of governments as well as the notion that they actually knew enough to improve macroeconomic outcomes.

In the 1960s, debates between Keynesians and monetarists dominated the economic headlines. The debates centered around three issues: (1) the effectiveness of monetary policy versus fiscal policy, (2) the Phillips curve, and (3) the role of policy.

Monetary Policy versus Fiscal Policy

Keynes had emphasized *fiscal* rather than *monetary* policy as the key to fighting recessions, and this had remained the prevailing wisdom. The IS curve, many argued, was quite steep: Changes in the interest rate had little effect on demand and output. Thus, monetary policy did not work very well. Fiscal policy, which affects demand directly, could affect output faster and more reliably.

Anna Schwartz

Friedman strongly challenged this conclusion. In their 1963 book *A Monetary History of the United States, 1867–1960*, Friedman and Anna Schwartz painstakingly reviewed the evidence on monetary policy and the relation between money and output in the United States over a century. Their conclusion was not only that monetary policy was powerful but that movements in money explained most of the fluctuations in output. They interpreted the Great Depression as the result of a major mistake in monetary policy, a decrease in the money supply as a result of bank failures—a decrease that the Fed could have avoided by increasing the monetary base.

Friedman and Schwartz's challenge was followed by a vigorous debate and by intense research on the respective effects of fiscal policy and monetary policy. In the end, an intellectual consensus was reached. Both fiscal policy and monetary policy clearly affected the economy. And if policymakers cared about not only the level but also the composition of output, the best policy was typically a mix of the two.

The Phillips Curve

The second debate focused on the Phillips curve. The Phillips curve was not part of the initial Keynesian model, but because it provided such a convenient (and apparently reliable) way of explaining the movement of wages and prices over time, it had become part of the neoclassical synthesis. In the 1960s, based on the empirical evidence up until then, many Keynesian economists believed that there was a reliable trade-off between unemployment and inflation, even in the long run.

Milton Friedman and Edmund Phelps (at Columbia University) strongly disagreed. They argued that the existence of such a long-run trade-off flew in the face of basic economic theory. They argued that the apparent trade-off would quickly vanish if policymakers actually tried to exploit it—that is, if they tried to achieve low unemployment by accepting higher inflation. As we saw in Chapter 8 when we studied the evolution of the Phillips curve, Friedman and Phelps were definitely right. By the mid-1970s, the consensus was indeed that there was no long-run trade-off between inflation and unemployment.

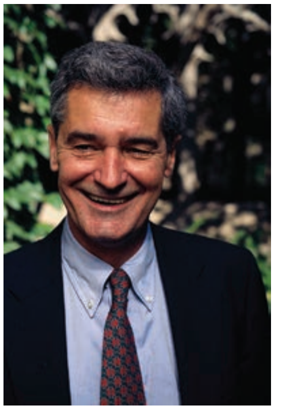
Edmund Phelps

The Role of Policy

The third debate centered on the role of policy. Skeptical that economists knew enough to stabilize output and that policymakers could be trusted to do the right thing, Friedman called for the use of simple rules, such as steady money growth (discussed in Chapter 23). Here is what he said in 1958:

> "A steady rate of growth in the money supply will not mean perfect stability even though it would prevent the kind of wide fluctuations that we have experienced from time to time in the past. It is tempting to try to go farther and to use monetary changes to offset other factors making for expansion and contraction.... The available evidence casts grave doubts on the possibility of producing any fine adjustments in economic activity by fine adjustments in monetary policy—at least in the present state of knowledge. There are thus serious limitations to the possibility of a discretionary monetary policy and much danger that such a policy may make matters worse rather than better.
>
> Political pressures to 'do something' in the face of either relatively mild price rises or relatively mild price and employment declines are clearly very strong indeed in the existing state of public attitudes. The main moral to be drawn from the two preceding points is that yielding to these pressures may frequently do more harm than good."[3]

As we saw in Chapter 21, this debate on the role of macroeconomic policy has not been settled. The nature of the arguments has changed a bit, but they are still with us today.



Robert Lucas

24-3 THE RATIONAL EXPECTATIONS CRITIQUE

Despite the battles between Keynesians and monetarists, macroeconomics in about 1970 looked like a successful and mature field. It appeared to successfully explain events and guide policy choices. Most debates were framed within a common intellectual framework. But within a few years, the field was in crisis. This crisis had two sources.

One was events. By the mid-1970s, most countries were experiencing *stagflation*, a word created at the time to denote the simultaneous existence of high unemployment (stagnation) and high inflation. Macroeconomists had not predicted stagflation. After the fact and after a few years of research, a convincing explanation was provided, based on the effects of adverse supply shocks on both inflation and output. (We discussed the effects of such shocks in Chapter 9.) But it was too late to undo the damage to the discipline's image.

The other was ideas. In the early 1970s, a small group of economists—Robert Lucas at Chicago; Thomas Sargent, then at Minnesota and now at New York University; and Robert Barro, then at Chicago and now at Harvard—led a strong attack against mainstream macroeconomics. They did not mince words. In a 1978 paper, Lucas and Sargent stated:

Thomas Sargent

[3]Source: Milton Friedman. *The Optimum Quantity Of Money,* Aldine Transaction; Revised edition (October 1, 2005). P. 185. ["The Supply of Money and Changes in Prices and Output," Testimony to Congress, 1958].

Robert Barro

"That the predictions [of Keynesian economics] were wildly incorrect, and that the doctrine on which they were based was fundamentally flawed, are now simple matters of fact, involving no subtleties in economic theory. The task which faces contemporary students of the business cycle is that of sorting through the wreckage, determining what features of that remarkable intellectual event called the Keynesian Revolution can be salvaged and put to good use, and which others must be discarded."[4]

The Three Implications of Rational Expectations

Lucas and Sargent's main argument was that Keynesian economics had ignored the full implications of the effect of expectations on behavior. The way to proceed, they argued, was to assume that people formed expectations as rationally as they could, based on the information they had. Thinking of people as having *rational expectations* had three major implications, all highly damaging to Keynesian macroeconomics.

The Lucas Critique

The first implication was that existing macroeconomic models could not be used to help design policy. Although these models recognized that expectations affect behavior, they did not incorporate expectations explicitly. All variables were assumed to depend on current and past values of other variables, including policy variables. Thus, what the models captured was the set of relations between economic variables as they had held in the past, under past policies. Were these policies to change, Lucas argued, the way people formed expectations would change as well, making estimated relations—and, by implication, simulations generated using existing macroeconometric models—poor guides to what would happen under these new policies. This critique of macroeconometric models became known as the **Lucas critique**. To take again the history of the Phillips curve as an example, the data up to the early 1970s had suggested a trade-off between unemployment and inflation. As policymakers tried to exploit that trade-off, it disappeared.

Rational Expectations and the Phillips Curve

The second implication was that when rational expectations were introduced in Keynesian models, these models actually delivered very un-Keynesian conclusions. For example, the models implied that deviations of output from its natural level were short-lived, much more so than Keynesian economists claimed.

This argument was based on a reexamination of the Phillips curve relation. In Keynesian models, the slow return of output to the natural level of output came from the slow adjustment of prices and wages through the Phillips curve mechanism. An increase in money, for example, led first to higher output and lower unemployment. Lower unemployment then led to higher nominal wages and higher prices. The adjustment continued until wages and prices had increased in the same proportion as nominal money, until unemployment and output were both back at their natural levels.

But this adjustment, Lucas pointed out, was highly dependent on wage setters' backward-looking expectations of inflation. In the MPS model, for example, wages responded only to current and past inflation and to current unemployment. But once the assumption was made that wage setters had rational expectations, the adjustment was likely to be much faster. Changes in money, to the extent that they were anticipated, might have no effect on output. For example, anticipating an increase in money of 5% over the coming year, wage setters would increase the nominal wages set in contracts for

[4]Source: Robert E. Lucas, Thomas J. Sargent, *Rational Expectations and Econometric Practice,* Volume 2, University of Minnesota Press, 1981.

the coming year by 5%. Firms would in turn increase prices by 5%. The result would be no change in the real money stock, and no change in demand or output.

In the logic of the Keynesian models, Lucas therefore argued, only *unanticipated changes in money* should affect output. Predictable movements in money should have no effect on activity. More generally, if wage setters had rational expectations, shifts in demand were likely to have effects on output for only as long as nominal wages were set—a year or so. Even on its own terms, the Keynesian model did not deliver a convincing theory of the long-lasting effects of demand on output.

Optimal Control versus Game Theory

The third implication was that if people and firms had rational expectations, it was wrong to think of policy as the control of a complicated but passive system. Rather, the right way was to think of policy as a game between policymakers and the economy. The right tool was not *optimal control* but *game theory*. And game theory led to a different vision of policy. A striking example was the issue of *time inconsistency* discussed by Finn Kydland (then at Carnegie Mellon, now at the University of California–Santa Barbara) and Edward Prescott (then at Carnegie Mellon, now at Arizona State University), an issue that we discussed in Chapter 21. Good intentions on the part of policymakers could actually lead to disaster.

To summarize: When rational expectations were introduced, Keynesian models could not be used to determine policy; Keynesian models could not explain long-lasting deviations of output from the natural level of output; the theory of policy had to be redesigned, using the tools of game theory.

The Integration of Rational Expectations

As you might have guessed from the tone of Lucas and Sargent's quote, the intellectual atmosphere in macroeconomics was tense in the early 1970s. But within a few years, a process of integration (of ideas, not people, because tempers remained high) had begun, and it was to dominate the 1970s and the 1980s.

Fairly quickly, the idea that rational expectations was the right working assumption gained wide acceptance. This was not because macroeconomists believed that people, firms, and participants in financial markets always form expectations rationally. They don't. But rational expectations appeared to be a natural benchmark—at least until economists make more progress in understanding whether, when, and how actual expectations systematically differ from rational expectations.

Work then started on the challenges raised by Lucas and Sargent.

The Implications of Rational Expectations

First, there was a systematic exploration of the role and implications of rational expectations in goods markets, financial markets, and labor markets. Much of what was discovered has been presented in this book. For example:

- Robert Hall, then at MIT and now at Stanford, showed that if consumers are foresighted (in the sense defined in Chapter 15), then changes in consumption should be unpredictable. The best forecast of consumption next year would be consumption this year! Put another way, changes in consumption should be hard to predict. This result came as a surprise to most macroeconomists at the time, but it is in fact based on a simple intuition. If consumers are foresighted, they will change their consumption only when they learn something new about the future. But by definition, such news cannot be predicted. This consumption behavior, known as the **random walk of consumption**, became the benchmark in consumption research thereafter.

Robert Hall

Rudiger Dornbusch

- Rudiger Dornbusch at MIT showed that the large swings in exchange rates under flexible exchange rates, which had previously been thought of as the result of speculation by irrational investors, were fully consistent with rationality. His argument—which we saw in Chapter 20—was that changes in monetary policy can lead to long-lasting changes in nominal interest rates; changes in current and expected nominal interest rates lead in turn to large changes in the exchange rate. Dornbusch's model, known as the *overshooting* model of exchange rates, became the benchmark in discussions of exchange rate movements.

Wage and Price Setting

Second, there was a systematic exploration of the determination of wages and prices, going far beyond the Phillips curve relation. Two important contributions were made by Stanley Fischer, then at MIT, and John Taylor, then at Columbia University and now at Stanford. Both showed that the adjustment of prices and wages in response to changes in unemployment can be slow *even under rational expectations*.

Stanley Fischer

Fischer and Taylor pointed out an important characteristic of both wage and price setting, the **staggering of wage and price decisions**. In contrast to the simple story we told previously, where all wages and prices increased simultaneously in anticipation of an increase in money, actual wage and price decisions are staggered over time. So there is not one sudden synchronized adjustment of all wages and prices to an increase in money. Rather, the adjustment is likely to be slow, with wages and prices adjusting over time to the new level of money through a process of leapfrogging. Fischer and Taylor thus showed that the second issue raised by the rational expectations critique could be resolved, that a slow return of output to the natural level of output can be consistent with rational expectations in the labor market.

The Theory of Policy

Third, thinking about policy in terms of game theory led to an explosion of research on the nature of the games being played, not only between policymakers and the economy but also among policymakers—between political parties, or between the central bank and the government, or between governments of different countries. One of the major achievements of this research was the development of a more rigorous way of thinking about fuzzy notions such as "credibility," "reputation," and "commitment." At the same time, there was a distinct shift in focus from "what governments should do" to "what governments actually do," an increasing awareness of the political constraints that economists should take into account when advising policymakers.

John Taylor

In short: By the end of the 1980s, the challenges raised by the rational expectations critique had led to a complete overhaul of macroeconomics. The basic structure had been extended to take into account the implications of rational expectations or, more generally, of forward-looking behavior by people and firms. As you have seen, these themes play a central role in this book.

24-4 DEVELOPMENTS IN MACROECONOMICS UP TO THE 2009 CRISIS

From the late 1980s to the crisis, three groups dominated the research headlines: the new classicals, the new Keynesians, and the new growth theorists. (Note the generous use of the word *new*. Unlike producers of laundry detergents, economists stop short of using "new and improved." But the subliminal message is the same.)

New Classical Economics and Real Business Cycle Theory

Edward Prescott

The rational expectations critique was more than just a critique of Keynesian economics. It also offered its own interpretation of fluctuations. Lucas argued that instead of relying on imperfections in labor markets, the slow adjustment of wages and prices, and so on to explain fluctuations, macroeconomists should see how far they could go in explaining fluctuations as the effects of shocks in competitive markets with fully flexible prices and wages.

This research agenda was taken up by the **new classicals**. The intellectual leader was Edward Prescott, and the models he and his followers developed are known as **real business cycle (RBC) models**. Their approach was based on two premises.

The first was methodological. Lucas had argued that, to avoid earlier pitfalls, macroeconomic models should be constructed from explicit microfoundations (i.e., utility maximization by workers, profit maximization by firms, and rational expectations). Before the development of computers, this was hard, if not impossible, to achieve. Models constructed in this way would have been too complex to solve analytically. Indeed, much of the art of macroeconomics was in finding simple shortcuts to capture the essence of a model while keeping the model simple enough to solve (it remains the art of writing a good textbook). The development of computing power made it possible to solve such models numerically, and an important contribution of RBC theory was the development of more and more powerful numerical methods of solution, which allowed for the development of richer and richer models.

The second was conceptual. Until the 1970s, most fluctuations had been seen as the result of imperfections, of deviations of actual output from a slowly moving potential level of output. Following up on Lucas's suggestion, Prescott argued in a series of influential contributions that fluctuations could indeed be interpreted as coming from the effects of technological shocks in competitive markets with fully flexible prices and wages. In other words, he argued that movements in actual output could be seen as movements in—rather than as deviations from—the potential level of output. As new discoveries are made, he argued, productivity increases, leading to an increase in output. The increase in productivity leads to an increase in the wage, which makes it more attractive to work, leading workers to work more. Productivity increases therefore lead to increases in both output and employment, just as we observe in the real world. Fluctuations are desirable features of the economy, not something policymakers should try to reduce.

Not surprisingly, this radical view of fluctuations was criticized on many fronts. As we discussed in Chapter 12, technological progress is the result of many innovations, each taking a long time to diffuse throughout the economy. It is hard to see how this process could generate anything like the large short-run fluctuations in output that we observe in practice. It is also hard to think of recessions as times of technological *regress*, when productivity and output both decrease. Finally, as we have seen, there is strong evidence that changes in money, which have no effect on output in RBC models, in fact have strong effects on output in the real world. Still, the conceptual RBC approach proved influential and useful. It made an important point, that not all fluctuations in output are deviations of output from its natural level, but movements in the natural level itself.

New Keynesian Economics

The term **new Keynesians** denotes a loosely connected group of researchers who shared a belief that the synthesis that emerged in response to the rational expectations critique was basically correct. But they also shared the belief that much remained to be learned about the nature of imperfections in different markets and about the implications of those imperfections for macroeconomic fluctuations.

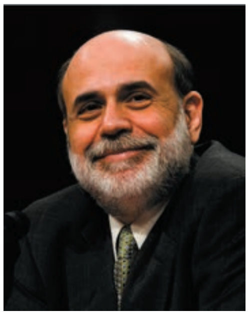
George Akerlof

There was further work on the nature of **nominal rigidities**. As we saw earlier in this chapter, Fischer and Taylor had shown that with the staggering of wage or price decisions, output can deviate from its natural level for a long time. This conclusion raised a number of questions. If the staggering of decisions is responsible, at least in part, for fluctuations, why don't wage setters/price setters synchronize decisions? Why aren't prices and wages adjusted more often? Why aren't all prices and all wages changed, say, on the first day of each week? In tackling these issues, George Akerlof (then at Berkeley, now at Georgetown University), Janet Yellen (then at Berkeley, then Chair of the Federal Reserve Board, and now at the Brookings Institution), and N. Gregory Mankiw (at Harvard University) derived a surprising and important result, often referred to as the **menu cost** explanation of output fluctuations.

Each wage setter or price setter is largely indifferent as to when and how often he changes his own wage or price (for a retailer, changing the prices on the shelf every day versus every week does not make much of a difference to the store's overall profits). Therefore, even small costs of changing prices—like the costs involved in printing a new menu, for example—can lead to infrequent and staggered price adjustment. This staggering leads to slow adjustment of the price level and to large aggregate output fluctuations in response to movements in aggregate demand. In short, decisions that do not matter much at the individual level (how often to change prices or wages) lead to large aggregate effects (slow adjustment of the price level, and shifts in aggregate demand that have a large effect on output).

Janet Yellen

Another line of research focused on imperfections in the labor market. We discussed in Chapter 7 the notion of *efficiency wages*—the idea that wages, if perceived by workers as being too low, may lead to shirking by workers on the job, problems of morale in a firm, difficulties in recruiting or keeping good workers, and so on. One influential researcher in this area was Akerlof, who explored the role of "norms," the rules that develop in any organization—in this case, a firm—to assess what is fair or unfair. This research led him and others to explore issues previously left to research in sociology and psychology, and to examine their macroeconomic implications. In another direction, Peter Diamond (at MIT), Dale Mortensen (at Cornell), and Christopher Pissarides (at the London School of Economics) looked at the labor market as characterized by constant reallocation, large flows, and bargaining between workers and firms, a characterization that has proven extremely useful and that we relied on in Chapter 7.



Ben Bernanke

Yet another line of research, which turned out to be very useful when the crisis occurred, explored the role of imperfections in credit markets. Most macro models assumed that monetary policy worked through interest rates, and that firms could borrow as much as they wanted at the market interest rate. In practice, many firms can borrow only from banks. And banks often turn down potential borrowers, despite the willingness of these borrowers to pay the interest rate charged by the bank. Why this happens, and how it affects our understanding of how monetary policy works, was the focus of research by, in particular, Ben Bernanke (then at Princeton, and then Chair of the Fed, now at the Brookings Institution) and Mark Gertler (at New York University).

New Growth Theory

Paul Romer

After being one of the most active topics of research in the 1960s, growth theory went into an intellectual slump. Since the late 1980s, however, it has made a strong comeback, under the name of **new growth theory**.

Two economists, Robert Lucas (the same Lucas who spearheaded the rational expectations critique) and Paul Romer (then at Berkeley, now at New York University), played an important role in defining the issues. When growth theory faded in the late 1960s, two major issues were left largely unresolved. One was the role of increasing returns to scale—whether, say, doubling capital and labor can actually cause output to more than double. The other was the determinants of technological progress. These are the two major issues on which new growth theory concentrated.

The discussions of the effects of research and development (R&D) on technological progress in Chapter 12 and of the interaction between technological progress and unemployment in Chapter 13 both reflect some of the advances made on this front. An important contribution here was the work of Philippe Aghion (then at Harvard University, now at the Collège de France) and Peter Howitt (then at Brown University), who developed a theme first explored by Joseph Schumpeter in the 1930s: the notion that growth is a process of *creative destruction* in which new products are constantly introduced, making old ones obsolete. Institutions that slow this process of reallocation (for example, by making it harder to create new firms or by making it more expensive for firms to lay off workers) may slow the rate of technological progress and thus decrease growth.


Philippe Aghion

Research also tried to identify the precise role of specific institutions in determining growth. Andrei Shleifer (at Harvard University) explored the role of different legal systems in affecting the organization of the economy, from financial markets to labor markets, and, through these channels, the effects of legal systems on growth. Daron Acemoglu (at MIT) explored how to go from correlations between institutions and growth—democratic countries are on average richer—to causality from institutions to growth. Does the correlation tell us that democracy leads to higher output per person, or does it tell us that higher output per person leads to democracy, or that some other factor leads to both more democracy and higher output per person? Examining the history of former colonies, Acemoglu argued that their growth performance has been shaped by the type of institutions put in place by their colonizers, thus showing a strong causal role of institutions in economic performance.


Peter Howitt

Toward an Integration

In the 1980s and 1990s, discussions between these three groups, and in particular between new classicals and new Keynesians, were often heated. New Keynesians would accuse new classicals of relying on an implausible explanation of fluctuations and ignoring obvious imperfections; new classicals would in turn point to the ad hocery of some of the new Keynesian models. From the outside—and indeed sometimes from the inside—macroeconomics looked like a battlefield rather than a research field.


Andrei Shleifer

By the 2000s, however, a synthesis appeared to be emerging. Methodologically, it built on the RBC approach and its careful description of the optimization problems of people and firms. Conceptually, it recognized the potential importance, emphasized by the RBC and the new growth theory, of changes in the pace of technological progress. But it also allowed for many of the imperfections emphasized by the new Keynesians, from the role of bargaining in the determination of wages, to the role of imperfect information in credit and financial markets, to the role of nominal rigidities in creating a role for aggregate demand to affect output. There was no convergence on a single model or a single list of important imperfections, but there was broad agreement on the framework and on the way to proceed.

A good example of this convergence was the work of Michael Woodford (at Columbia) and Jordi Galí (at Pompeu Fabra). Woodford, Galí, and a number of coauthors developed a model, known as the *new Keynesian model*, that embodies utility and profit maximization, rational expectations, and nominal rigidities. You can think of it as a high-tech version of the model presented in Chapter 16. This model proved extremely useful and influential in the redesign of monetary policy—from the focus on inflation targeting to the reliance on interest rate rules—that we described in Chapter 23. It led to the development of a class of larger models that build on its simple structure but allow for a longer


Daron Acemoglu

menu of imperfections and thus must be solved numerically. These models, which are now standard workhorses in most central banks, are known as "*dynamic stochastic general equilibrium*" (DSGE) models.


Michael Woodford

24-5 FIRST LESSONS FOR MACROECONOMICS AFTER THE CRISIS

Just when a new synthesis appeared to be in sight and macroeconomists felt that they had the tools to understand the economy and design policy, the financial crisis started, which led to the Great Recession. We saw in Section 24-1 that the Great Depression led to a dramatic reassessment of macroeconomics and started the Keynesian revolution. You may wonder, Will this crisis have the same effect on macroeconomics, leading to yet another revolution? It is too early to say, but my guess is probably not a revolution but a major reassessment.


Jordi Galí

There is no question that the crisis reflects a major intellectual failure on the part of macroeconomics. The failure was in not realizing that such a large crisis *could happen*, that the characteristics of the economy were such that a relatively small shock, in this case the decrease in US housing prices, could lead to a major global financial and macroeconomic crisis. The source of the failure, in turn, was insufficient focus on the role of the financial institutions in the economy. (To be fair, a few macroeconomists, who were looking more closely at the financial system, sounded the alarm; best known among them are Nouriel Roubini, at New York University, and the economists at the Bank for International Settlements in Basel, whose job it is to closely follow financial developments.)


Bengt Holmström

By and large, the financial system, and the complex role of banks and other financial institutions in the intermediation of funds between lenders and borrowers, was ignored in most macroeconomic models. There were exceptions. Work by Doug Diamond (at Chicago) and Philip Dybvig (at Washington University in St. Louis) in the 1980s had clarified the nature of bank runs (which we examined in Chapter 6). Illiquid assets and liquid liabilities created a risk of runs even for solvent banks. The problem could be avoided only by the provision of liquidity by the central bank if and when needed. Work by Bengt Holmström (at MIT) and Jean Tirole (then at MIT, and now at Toulouse) had shown that liquidity issues were endemic to a modern economy. Not only banks but also firms could well find themselves in a position where they were solvent but illiquid, unable to raise the additional cash to finish a project or unable to repay investors when they wanted repayment. An important paper by Andrei Shleifer (at Harvard) and Robert Vishny (at Chicago) called "The Limits of Arbitrage" had shown that, after a decline in an asset price below its fundamental value, investors might not be able to take advantage of the arbitrage opportunity; indeed, they may themselves be forced to sell the asset, leading to a further decline in the price and a further deviation from fundamentals. Behavioral economists (for example, Richard Thaler, at Chicago) had pointed to the way individuals differ from the rational individual model typically used in economics, and had drawn implications for financial markets.


Jean Tirole

Thus, most of the elements needed to understand the crisis were available. Much of the work, however, was done outside macroeconomics, in the fields of finance or corporate finance. The elements were not integrated in a consistent macroeconomic model, and their interactions were poorly understood. Leverage, complexity, and liquidity, the factors that, as we saw in Chapter 6, combined to create the crisis, were nearly fully absent from the macroeconomic models used by central banks.

Now more than a decade since the start of the crisis, things have changed dramatically. Not surprisingly, researchers have turned their attention to the financial system and the nature of macrofinancial linkages. Further work is taking place on the various pieces, and these pieces are starting to be integrated into the large macroeconomic models. Lessons for policy are also being drawn, be it on the use of *macroprudential* tools or the dangers of high public debt. There is still a long way to go, but, in the end, macroeconomic models will be richer, with a better understanding of the financial system. Yet one has to be realistic. If history is any guide, the economy will be hit by another type of shock macroeconomists have not thought of.

The lessons from the crisis probably go beyond adding the financial sector to macroeconomic models and analysis. The Great Depression had, rightly, led most economists to question the macroeconomic properties of a market economy and to suggest a larger role for government intervention. The Great Recession is raising similar questions. Both the new classical and new Keynesian models had in common the belief that, in the medium run at least, the economy naturally returned to its natural level. The new classicals took the extreme position that output was always at its natural level. The new Keynesians took the view that, in the short run, output would likely deviate from its natural level but that, in the medium run, natural forces would eventually return the economy to potential. The Great Depression and the long slump in Japan were well known; they were seen, however, as aberrations and thought to be caused by substantial policy mistakes that could have been avoided. Many economists today believe that this optimism was excessive. The seven years spent in the liquidity trap in the United States made clear that the usual adjustment mechanism—namely, a decrease in interest rates in response to low output—could break down and actually prevent a return to normal.

If there is a consensus, it might be that, with respect to small shocks and normal fluctuations, the adjustment process works; but that, in response to large, exceptional shocks, the normal adjustment process may fail, the room for policy may be limited, and it may take a long time for the economy to repair itself. For the moment, the priority for researchers is to better understand what has happened, and for policymakers to use, as best they can, the monetary and fiscal policy tools they have, to steer the world economy back to health.

Let me end the book with a cartoon. Macroeconomics is complex. I hope the book has helped you understand it better.

Andrews McMeel Universal/GoComics.com

SUMMARY

- The history of modern macroeconomics starts in 1936, with the publication of Keynes's *General Theory of Employment, Interest, and Money*. Keynes's contribution was formalized in the IS-LM model by John Hicks and Alvin Hansen in the 1930s and early 1940s.
- The period from the early 1940s to the early 1970s can be called the golden age of macroeconomics. Among the major developments were the theories of consumption, investment, money demand, and portfolio choice; growth theory; and large macroeconometric models.
- The main debate during the 1960s was between Keynesians and monetarists. Keynesians believed developments in macroeconomic theory allowed for better control of the economy. Monetarists, led by Milton Friedman, were more skeptical of the ability of governments to help stabilize the economy.
- In the 1970s, macroeconomics experienced a crisis for two reasons: One was the appearance of stagflation, which came as a surprise to most economists. The other was a theoretical attack led by Robert Lucas. Lucas and his followers showed that when rational expectations were introduced (1) Keynesian models could not be used to determine policy, (2) Keynesian models could not explain long-lasting deviations of output from its natural level, and (3) the theory of policy needed to be redesigned using the tools of game theory.
- Much of the 1970s and 1980s was spent integrating rational expectations into macroeconomics. As is reflected in this text, macroeconomists are now much more aware of the role of expectations in determining the effects of shocks and policy and of the complexity of policy than they were two decades ago.
- Recent research in macroeconomic theory, up to the crisis, proceeded along three lines: New classical economists explored the extent to which fluctuations can be explained as movements in the natural level of output, as opposed to movements away from the natural level of output. New Keynesian economists explored more formally the role of market imperfections in fluctuations. New growth theorists explored the determinants of technological progress. These lines were increasingly overlapping, and, on the eve of the crisis, a new synthesis appeared to be emerging.
- The crisis reflects a major intellectual failure on the part of macroeconomics: the failure to understand the macroeconomic importance of the financial system. Although many of the elements needed to understand the crisis had been developed before the crisis, they were not central to macroeconomic thinking and were not integrated in large macroeconomic models. Much research is now focused on macrofinancial linkages.
- The crisis also raised a larger issue, about the adjustment process through which output returns to its natural level. If there is a consensus, it might be that with respect to small shocks and normal fluctuations, the adjustment process works and policy can accelerate this return; but that, in response to large, exceptional shocks, the normal adjustment process may fail, the room for policy may be limited, and it may take a long time for the economy to repair itself.

KEY TERMS

business cycle theory, 508
effective demand, 508
liquidity preference, 508
neoclassical synthesis, 508
Keynesians, 510
monetarists, 510
Lucas critique, 512
random walk of consumption, 513
staggering (of wage and price decisions), 514
new classicals, 515
real business cycles, 515
RBC models, 515
new Keynesians, 515
nominal rigidities, 516
menu costs, 516
new growth theory, 516

FURTHER READINGS

- Two classics are John Maynard Keynes, *The General Theory of Employment, Interest, and Money* (Palgrave Macmillan, 1936) and Milton Friedman and Anna Jacobson Schwartz, *A Monetary History of the United States,* (Princeton University Press, 1963). The first makes for hard reading, and the second is a heavy volume.
- For a more accessible description of Keynes' work and influence, read *Keynes: Useful Economics for the World Economy*, by Peter Temin and David Vines (MIT Press, 2014).
- For an account of macroeconomics in textbooks since the 1940s, read Paul Samuelson's "Credo of a Lucky Textbook Author," *Journal of Economic Perspectives*, 1997, Vol. 11 (Spring): pp. 153–160.
- In the introduction to *Studies in Business Cycle Theory* (MIT Press, 1981), Robert Lucas develops his approach to macroeconomics and gives a guide to his contributions.
- The paper that launched real business cycle theory is Edward Prescott, "Theory Ahead of Business Cycle

Measurement," *Federal Reserve Bank of Minneapolis Review*, 1986 (Fall): pp. 9–22. It is not easy reading.

- For more on new Keynesian economics, read David Romer, "The New Keynesian Synthesis," *Journal of Economic Perspectives*, 1993, Vol. 7 (Winter): pp. 5–22.
- For more on new growth theory, read Paul Romer, "The Origins of Endogenous Growth," *Journal of Economic Perspectives*, 1994, Vol. 8 (Winter): pp. 3–22.
- For a detailed look at the history of macroeconomic ideas, with in-depth interviews of most of the major researchers, read Brian Snowdon and Howard Vane, *Modern Macroeconomics: Its Origins, Development and Current State* (Edward Elgar Publishing Ltd., 2005).
- For two points of view on the precrisis state of macroeconomics, read V. V. Chari and Patrick Kehoe, "Macroeconomics in Practice: How Theory Is Shaping Policy," *Journal of Economic Perspectives,* 2006, 20(4): pp. 3–28; and N. Gregory Mankiw, "The Macroeconomist as Scientist and Engineer," *Journal of Economic Perspectives,* 20(4): pp. 29–46.
- For a skeptical view of financial markets and the contributions of Thaler and Shleifer among others, read *The Myth of the Rational Market: A History of Risk, Reward, and Delusion on Wall Street*, by Justin Fox (Harper Collins Publishers, 2009).
- For a discussion of the fiscal and monetary policy challenges facing policymakers today: *Rethinking Stabilization Policy: Evolution or Revolution?*, by Olivier Blanchard and Lawrence Summers, www.nber.org/papers/w24179, 2018.

If you want to learn more about macroeconomic issues and theory:

- For more on the history of economic thought in general (not just macroeconomics), a nice blog site is "The Undercover Historian," https://beatricecherrier.wordpress.com/
- Most economics journals are heavy on mathematics and hard to read, but a few make an effort to be more reader-friendly. The *Journal of Economic Perspectives*, in particular, has nontechnical articles on current economic research and issues. The *Brookings Papers on Economic Activity*, published twice a year, analyze current macroeconomic problems; so does *Economic Policy*, published in Europe, which focuses more on European issues.
- Most regional Federal Reserve Banks also publish reviews with easy-to-read articles; these reviews are available free of charge. Among these are the *Economic Review*, published by the Cleveland Fed; the *Economic Review*, published by the Kansas City Fed; the *New England Economic Review*, published by the Boston Fed; and the *Quarterly Review*, published by the Minneapolis Fed.
- More advanced treatments of current macroeconomic theory—roughly at the level of a first graduate course in macroeconomics—are given by David Romer, *Advanced Macroeconomics*, 5th ed. (McGraw-Hill, 2018) and by Olivier Blanchard and Stanley Fischer, *Lectures on Macroeconomics* (MIT Press, 1989).

Appendices

APPENDIX 1 An Introduction to the National Income and Product Accounts

This appendix introduces the basic structure and the terms used in the national income and product accounts. The basic measure of aggregate activity is gross domestic product (GDP). The **national income and product accounts (NIPA)**, or simply **national accounts**, are organized around two decompositions of GDP.

One decomposes GDP from the *income side*: Who receives what?

The other decomposes GDP from the *production side* (called the *product side* in the national accounts): What is produced, and who buys it?

The Income Side

Table A1-1 looks at the income side of GDP—who receives what.

The top part of the table (lines 1–8) goes from GDP to national income—the sum of the incomes received by the different factors of production:

- The starting point, in line 1, is **gross domestic product (GDP)**. GDP is defined as *the market value of the goods and services produced by labor and property located in the United States.*
- The next three lines take us from GDP to **gross national product (GNP)** (line 4). GNP is an alternative measure of aggregate output. It is defined as *the market value of the goods and services produced by labor and property supplied by US residents.*

Until the 1990s, most countries used GNP rather than GDP as the main measure of aggregate activity. The emphasis in the US national accounts shifted from GNP to GDP in 1991. The difference between the two comes from the distinction between "located in the United States" (used for GDP) and "supplied by US residents" (used for GNP).

Table A1-1 GDP: The Income Side, 2018 (billions of dollars)

From gross domestic product to national income:		
1 Gross domestic product (GDP)	20,494	
2 Plus: Receipts of factor income from the rest of the world		1,076
3 Minus: Payments of factor income to the rest of the world		–815
4 Equals: Gross national product (GNP)	20,755	
5 Minus: Consumption of fixed capital		–3,273
6 Equals: Net national product (NNP)	17,481	
7 Minus: Statistical discrepancy		–62
8 Equals: National income	17,544	
Decomposition of national income:		
9 Indirect taxes	1,429	
10 Compensation of employees	10,856	
11 Wages and salaries		8,834
12 Supplements to wages and salaries		2,021
13 Corporate profits and business transfers	2,263	
14 Net interest	577	
15 Proprietors' income	1,579	
16 Rental income of persons	760	

Source: Survey of Current Business, April 2019, Tables 1-7-5 and 1-12

For example, profit from a US-owned plant in Japan is not included in US GDP, but is included in US GNP.

To go from GDP to GNP, we must first add **receipts of factor income from the rest of the world**, which is income from US capital or US residents abroad (line 2), then subtract **payments of factor income to the rest of the world**, which is income received by foreign capital and foreign residents in the United States (line 3).

In 2018, payments from the rest of the world exceeded receipts to the rest of the world by $261 billion, so GNP was larger than GDP by $261 billion.

- The next step takes us from GNP to **net national product (NNP)** (line 6). The difference between GNP and NNP is the depreciation of capital, called **consumption of fixed capital** in the national accounts.
- Finally, lines 7 and 8 take us from NNP to **national income** (line 8). National income is defined as the *income that originates in the production of goods and services supplied by residents of the United States*. In theory, national income and NNP should be equal. In practice, they typically differ, because they are constructed in different ways.

NNP is constructed from the top down, starting from GDP and going through the steps shown in Table A1-1. National income is constructed instead from the bottom up, by adding the different components of factor income (compensation of employees, corporate profits, and so on). If we could measure everything exactly, the two measures should be equal. In practice, the two measures differ, and the difference between the two is called the *statistical discrepancy*. In 2018, national income computed from the bottom up (the number in line 8) was larger than the NNP computed from the top down (the number in line 6) by $62 billion. The statistical discrepancy is a useful reminder of the statistical problems involved in constructing the national income accounts. Although $62 billion seems like a large error, as a percentage of GDP, it is about 0.3 percentage point.

The bottom part of the table (lines 9–15) decomposes national income into different types of income.

- **Indirect taxes** (line 9). Some of the national income goes directly to the state in the form of sales taxes. (Indirect taxes are just another name for sales taxes.)

  The rest of national income goes either to employees or to firms:
- **Compensation of employees** (line 10), or labor income, is what goes to employees. It is the largest component of national income, accounting for 61% of national income. Labor income is the sum of wages and salaries (line 11) and of supplements to wages and salaries (line 12). These range from employer contributions for social insurance (by far the largest item) to such exotic items as employer contributions to marriage fees for justices of the peace.
- **Corporate profits and business transfers** (line 13). Profits are revenues minus costs (including interest payments) and minus depreciation. (Business transfers, which account for $159 billion out of $2,263 billion, are items such as liability payments for personal injury and corporate contributions to nonprofit organizations.)
- **Net interest** (line 14) is the interest paid by firms minus the interest received by firms, plus interest received from the rest of the world minus interest paid to the rest of the world. In 2018, most of net interest represented net interest paid by firms: The United States received about as much in interest from the rest of the world as it paid to the rest to the world. So the sum of corporate profits plus net interest paid by firms was approximately $\$2{,}263 \text{ billion} + \$577 \text{ billion} = \$2{,}840 \text{ billion}$ or about 16% of national income.
- **Proprietors' income** (line 15) is the income received by persons who are self-employed. It is defined as *the income of sole proprietorships, partnerships, and tax-exempt cooperatives.*
- **Rental income of persons** (line 16) is the income from the rental of real property, minus depreciation on this real property. Houses produce housing services; rental income measures the income received for these services.

  If the national accounts counted only actual rents, rental income would depend on the proportion of apartments and houses that were rented versus those that were owner occupied. For example, if everybody became the owner of the apartment or the house in which he or she lived, rental income would go to zero, and thus measured GDP would drop. To avoid this problem, national accounts treat houses and apartments as if they were all rented out. So, rental income is constructed as actual rents plus *imputed* rents on houses and apartments that are owner occupied.

Before we move to the product side, Table A1-2 shows how we can go from national income to personal disposable income, which is the income available to persons after they have received transfers and paid taxes.

- Not all national income (line 1) is distributed to persons.

Some of the income goes to the state in the form of indirect taxes, so the first step is to subtract indirect taxes. (Line 2 in Table A1-2 is equal to line 9 in Table A1-1.)

Table A1-2	From National Income to Personal Disposable Income, 2018 (billions of dollars)		
1	**National income**	**17,544**	
2	**Minus: Indirect taxes**		**−1,429**
3	**Minus: Corporate profits and business transfers**		**−2,263**
4	**Minus: Net interest**		**−577**
5	**Plus: Income from assets**		**2,768**
6	**Plus: Personal transfers**		**2,980**
7	**Minus: Contributions for social insurance**		**−1,361**
8	**Equals: Personal income**	**17,582**	
9	**Minus: Personal tax and non-tax payments**		**−2,050**
10	**Equals: Personal disposable income**	**15,532**	

Source: Survey of Current Business, April 2019, Tables 1-7-5, 1-12, and 2-1

Some of the corporate profits are retained by firms. Some of the interest payments by firms go to banks or go abroad. So the second step is to subtract all corporate profits and business transfers (line 3—equal to line 13 in Table A1-1) and all net interest payments (line 4—equal to line 14 in Table A1-1), and add back all income from assets (dividends and interest payments) received by persons (line 5).

- People receive income not only from production but also from public transfers (line 6). Transfers accounted for $2,980 billion in 2018. From these transfers must be subtracted personal contributions for social insurance, $1,361 billion (line 7).
- The net result of these adjustments is **personal income**, the income actually received by persons (line 8). **Personal disposable income** (line 10) is equal to personal income minus personal tax and nontax payments (line 9). In 2018, personal disposable income was $15,532 billion, or about 75% of GDP.

The Product Side

Table A1-3 looks at the product side of the national accounts—what is produced, and who buys it.

Start with the three components of domestic demand: consumption, investment, and government spending.

- Consumption, called **personal consumption expenditures** (line 2), is by far the largest component of demand. It is defined as *the sum of goods and services purchased by persons resident in the United States.*

  In the same way that national accounts include imputed rental income on the income side, they include imputed housing services as part of consumption. Owners of a house are assumed to consume housing services for a price equal to the imputed rental income of that house.

  Consumption is disaggregated into three components: purchases of **durable goods** (line 3), **nondurable goods** (line 4), and **services** (line 5). Durable goods are commodities that can be stored and have an average life of at least three years; automobile purchases are the largest item here. Nondurable goods are commodities that can be stored but have an average life of less than three years. Services are commodities that cannot be stored and must be consumed at the place and time of purchase.
- Investment, called **gross private domestic fixed investment** (line 6), is the sum of two very different components.

  Nonresidential investment (line 7) is the purchase of new capital goods by firms. These may be either **structures** (line 8)—mostly new plants—or **equipment and software** (line 9)—such as machines, computers, or office equipment.

  Residential investment (line 10) is the purchase of new houses or apartments by persons.
- **Government purchases** (line 11) equal the purchases of goods by the government plus the compensation of government employees. (The government is thought of as buying the services of the government employees.)

Government purchases equal the sum of purchases by the federal government (line 12), which themselves can be disaggregated between spending on national defense (line 13) and nondefense spending (line 14), and purchases by state and local governments (line 15).

Table A1-3	GDP: The Product Side, 2018 (billions of dollars)			
1	Gross domestic product	20,494		
2	Personal consumption expenditures	13,949		
3	Durable goods		1,459	
4	Nondurable goods		2,879	
5	Services		9,610	
6	Gross private domestic fixed investment	3,650		
7	Nonresidential		2,799	
8	Structures			637
9	Equipment and software			2,161
10	Residential		794	
11	Government purchases	3,520		
12	Federal		1,320	
13	National defense			779
14	Nondefense			541
15	State and local		2,201	
16	Net exports	−625		
17	Exports		2,531	
18	Imports		−3,156	
19	Change in business inventories	56		

Source: Survey of Current Business, April 2019, Table 1-1-5

Note that government purchases do not include transfers from the government or interest payments on government debt. These do not correspond to purchases of either goods or services and so are not included here. This means that the number for government purchases shown in Table A1-3 is substantially smaller than the number we typically hear for government spending, which includes transfers and interest payments.

- The sum of consumption, investment, and government purchases gives the demand for goods by US firms, US persons, and the US government. If the United States were a closed economy, this would be the same as the demand for US goods. But because the US economy is open, the two numbers are different. To get to the demand for US goods, we must make two adjustments. First, we must add the foreign purchases of US goods, **exports** (line 17). Second, we must subtract US purchases of foreign goods, **imports** (line 18). In 2014, exports were smaller than imports by $625 billion. Thus, **net exports** (or, equivalently, the **trade balance**), was equal to −$625 billion (line 16).
- Adding consumption, investment, government purchases, and net exports gives the *total purchases of US goods.* Production may, however, be less than purchases if firms satisfy the difference by decreasing inventories. Or production may be greater than purchases, in which case firms are accumulating inventories. The last line of Table A1-3 gives **the change in business inventories** (line 19), also sometimes called (rather misleadingly) "inventory investment." It is defined as the *change in the volume of inventories held by businesses*. The change in business inventories can be positive or negative. In 2014, it was small and positive; US production was higher than total purchases of US goods by $56 billion.

The Federal Government in the National Income Accounts

Table A1-4 presents the basic numbers describing federal government economic activity in 2018, using NIPA numbers. (The numbers are sometimes presented using fiscal rather than calendar years. The reason for doing so is that budget projections are typically framed in terms of fiscal year rather than calendar year numbers. The fiscal year runs from October 1 of the previous calendar year to September 30 of the current calendar year.)

The reason for using the NIPA numbers rather than the official budget numbers is that they are economically more meaningful—they are a better representation of what the government is doing in the economy than the

Table A1-4	US Federal Budget Revenues and Expenditures, 2018 (billions of dollars)			
1	**Revenues**	**3,500**		
2	**Personal taxes**		**1,614**	
3	**Corporate profit taxes**		**158**	
4	**Indirect taxes**		**160**	
5	**Social insurance contributions**		**1,345**	
6	**Other**		**223**	
7	**Expenditures, excluding net interest payments**	**3,937**		
8	**Consumption expenditures**		**1,032**	
9	**Defense**			**778**
10	**Nondefense**			**254**
11	**Transfers to persons**		**2,180**	
12	**Grants to state/local governments**		**578**	
13	**Other**		**147**	
14	**Primary balance (+ sign: surplus, – sign: deficit)**	**–437**		
15	**Net interest payments**	**545**		
16	**Real interest payments**		**258**	
17	**Inflation component**		**287**	
18	**Official surplus: (1) minus (7) minus (15)**	**–982**		
19	**Inflation-adjusted surplus: (18) plus (17)**	**–695**		

Source: Survey of Current Business, April 2019, Table 3-2. Inflation adjustment calculated using debt from Table B-45, Economic Report of the President.

numbers presented in the various budget documents. Budget numbers presented by the government need not follow the national income accounting conventions and sometimes involve creative accounting.

In 2018, federal revenues were $3,500 billion (line 1). Of those, personal taxes (also called *income taxes*) accounted for $1,614 billion, or 46% of revenues; social insurance contributions (also called *payroll taxes*) accounted for $1,345 billion, or 38% of revenues.

Expenditures excluding interest payments but including transfer payments to individuals were $3,937 billion (line 7). Consumption expenditures (mostly wages and salaries of public employees and depreciation of capital) accounted for $1,032 billion, or 26% of expenditures. Excluding defense, consumption expenditures were only $254 billion. **Transfers to persons** (also called *entitlement programs*, mostly unemployment, retirement, and health benefits) were a much larger $2,180 billion.

The federal government was therefore running a primary deficit of $437 billion (line 1 minus line 7, here recorded as a negative value of the **primary balance** in line 14).

Net interest payments on the debt held by the public totaled $545 billion (line 15). The **official deficit** was therefore equal to $982 billion (line 14 plus line 15). We know, however, that this measure is incorrect (see the Focus Box "Inflation Accounting and the Measurement of Deficits" in Chapter 22). It is appropriate to correct the official deficit measure for the role of inflation in reducing the real value of the public debt. The correct measure, the **inflation-adjusted deficit**—the sum of the official deficit plus *real* interest payments—was $695 billion (line 19), or 3.4% of GDP.

Warning

National accounts give an internally consistent description of aggregate activity. But underlying these accounts are many choices of what to include and what not to include, where to put some types of income or spending, and so on. Here are five examples.

- Work within the home is not counted in GDP. If, for example, two neighbors decide to babysit each other's child and pay each other for the babysitting services, measured GDP will go up, whereas true GDP clearly does not change. The solution would be to count work within the home in GDP, just as rents are imputed for owner-occupied housing. But, so far, this has not been done.

- The purchase of a house is treated as an investment, and housing services are then treated as part of consumption. Contrast this with the treatment of automobiles. Despite the fact that they provide services for a long time—although not as long a time as houses do—purchases of automobiles are not treated as investment. They are treated as consumption and appear in the national accounts only in the year in which they are bought.
- Firms' purchases of machines are treated as investment. The purchase of education is treated as consumption of education services. But education is clearly in part an investment; people acquire education in part to increase their future income.
- Many government purchases have to be valued in the national accounts in the absence of a market transaction. How do we value the work of teachers in teaching children to read when that transaction is mandated by the state as part of compulsory education? The rule used is to value it at cost, so using the salaries of teachers.
- The correct calculation of the government's deficit (and debt) is a challenging task. Here is one aspect of the problem: Suppose teachers are paid partly with cash and partly with the promise of a future retirement pension. There is an important sense that the pension is just like government debt (i.e., a future liability of taxpayers). However, these liabilities are not counted in the deficit measure in Table A1-4 or in standard measures of public debt. Another problem lies in the treatment of private sector debt guarantees by federal or state government. Should such contingent liabilities be counted as part of public debt?

The list could go on. However, the point of these examples is not to make you conclude that national accounts are wrong. Most of the accounting decisions you just saw were made for good reasons, often because of data availability or for simplicity. The point is that to use national accounts best, you should understand their logic, but also understand the choices that have been made and thus their limitations.

Key Terms

- national income and product accounts (NIPA), A-1
- national accounts, A-1
- gross domestic product (GDP), A-1
- gross national product (GNP), A-1
- receipts of factor income from the rest of the world, A-2
- payments of factor income to the rest of the world, A-2
- net national product (NNP), A-2
- consumption of fixed capital, A-2
- national income, A-2
- indirect taxes, A-2
- compensation of employees, A-2
- corporate profits, A-2
- net interest, A-2
- proprietors' income, A-2
- rental income of persons, A-2
- personal income, A-3
- personal disposable income, A-3
- personal consumption expenditures, A-3
- durable goods, A-3
- nondurable goods, A-3
- services, A-3
- gross private domestic fixed investment, A-3
- nonresidential investment, A-3
- structures, A-3
- equipment and software, A-3
- residential investment, A-3
- government purchases, A-3
- exports, A-4
- imports, A-4
- net exports, A-4
- trade balance, A-4
- changes in business inventories, A-4
- transfers to persons, A-5
- primary surplus, A-5
- official deficit, A-5
- inflation-adjusted deficit, A-5

Further Readings

For more details, read "A Guide to the National Income and Product Accounts of the United States," May 2019 (www.bea.gov/national/pdf/nipaguid.pdf).

APPENDIX 2 A Math Refresher

This appendix presents the mathematical tools and mathematical results used in this text.

Geometric Series

Definition. A **geometric series** is a sum of numbers of the form:

$$1 + x + x^2 + \cdots + x^n$$

where x is a number that may be greater or smaller than one, and x^n denotes x to the power n; that is, x times itself n times.

Examples of such series are:

- The sum of spending in each round of the multiplier (Chapter 3). If c is the marginal propensity to consume, then the sum of increases in spending after $n + 1$ rounds is given by:

$$1 + c + c^2 + \cdots + c^n$$

- The present discounted value of a sequence of payments of \$1 each year for n years (Chapter 14), when the interest rate is equal to i:

$$1 + \frac{1}{1 + i} + \frac{1}{(1 + i)^2} + \cdots + \frac{1}{(1 + i)^{n-1}}$$

We usually have two questions we want to answer with such a series:

1. What is the sum?
2. Does the sum explode as we let n increase, or does it reach a finite limit (and, if so, what is that limit)?

The following propositions tell you what you need to know to answer these questions.

Proposition 1 tells you how to compute the sum:

Proposition 1:

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x} \qquad \text{(A2.1)}$$

Here is the proof: Multiply the sum by $(1 - x)$, and use the fact that $x^a x^b = x^{a+b}$ (that is, you must add exponents when multiplying):

$$\begin{aligned}(1 + x + x^2 + \cdots + x^n)(1 - x) &= 1 + x + x^2 + \cdots + x^n \\ &\quad - x - x^2 - \cdots - x^n - x^{n+1} \\ &= 1 \qquad\qquad\qquad\qquad - x^{n+1}\end{aligned}$$

All the terms on the right except for the first and the last cancel. Dividing both sides by $(1 - x)$ gives equation (A2.1).

This formula can be used for any x and any n. If, for example, x is 0.9 and n is 10, then the sum is 6.86. If x is 1.2 and n is 10, then the sum is 32.15.

Proposition 2 tells you what happens as n gets large:

Proposition 2: If x is less than one, the sum goes to $1/(1 - x)$ as n gets large. If x is equal to or greater than one, the sum explodes as n gets large.

Here is the proof: If x is less than one, then x^n goes to zero as n gets large. Thus, from equation (A2.1), the sum goes to $1/(1 - x)$. If x is greater than one, then x^n becomes larger and larger as x^n increases, $1 - x^n$ becomes a larger and larger negative number, and the ratio $(1 - x^n)/(1 - x)$ becomes a larger and larger positive number. Thus, the sum explodes as n gets large.

Application from Chapter 14: Consider the present value of a payment of \$1 forever, starting next year, when the interest rate is i. The present value is given by:

$$\frac{1}{(1 + i)} + \frac{1}{(1 + i)^2} + \cdots \qquad \text{(A2.2)}$$

Factoring out $1/(1 + i)$, rewrite this present value as:

$$\frac{1}{(1 + i)}\left[1 + \frac{1}{(1 + i)} + \cdots\right]$$

The term in brackets is a geometric series, with $x = 1/(1 + i)$. As the interest rate i is positive, x is less than 1. Applying Proposition 2, when n gets large, the term in brackets equals

$$\frac{1}{1 - \frac{1}{(1 + i)}} = \frac{(1 + i)}{(1 + i - 1)} = \frac{(1 + i)}{i}$$

Replacing the term in brackets in the previous equation by $(1 + i)/i$ gives:

$$\frac{1}{(1 + i)}\left[\frac{(1 + i)}{i}\right] = \frac{1}{i}$$

The present value of a sequence of payments of \$1 a year forever, starting next year, is equal to \$1 divided by the interest rate. If i is equal to 5% per year, the present value equals $\$1/0.05 = \$20$.

Useful Approximations

Throughout this text, we use a number of approximations that make computations easier. These approximations are most reliable when the variables x, y, and z are small, say between 0 and 10%. The numerical examples in Propositions 3–10 are based on the values $x = 0.05$ and $y = 0.03$.

Proposition 3:

$$(1+x)(1+y) \approx (1+x+y) \tag{A2.3}$$

Here is the proof. Expanding $(1+x)(1+y)$ gives $(1+x)(1+y) = 1+x+y+xy$. If x and y are small, then the product xy is very small and can be ignored as an approximation (for example, if $x = 0.05$ and $y = 0.03$, then $xy = 0.0015$). So, $(1+x)(1+y)$is approximately equal to $(1+x+y)$.The approximation (on the right side) gives 1.08 compared to an exact value (on the left) of 1.0815.

Proposition 4:

$$(1+x)^2 \approx 1+2x \tag{A2.4}$$

The proof follows directly from Proposition 3, with $y = x$. For the value of $x = 0.05$, the approximation gives 1.10, compared to an exact value of 1.1025.

Application from Chapter 14: From arbitrage, the relation between the two-year interest rate and the current and the expected one-year interest rates is given by:

$$(1+i_{2t})^2 = (1+i_{1t})(1+i^e_{1t+1})$$

Using Proposition 4 for the left side of the equation gives:

$$(1+i_{2t})^2 \approx 1+2i_{2t}$$

Using Proposition 3 for the right side of the equation gives:

$$(1+i_{1t})(1+i^e_{1t+1}) \approx 1+i_{1t}+i^e_{1t+1}$$

Using this expression to replace $(1+i_{1t})(1+i^e_{1t+1})$ in the original arbitrage relation gives:

$$1+2i_{2t} = 1+i_{1t}+i^e_{1t+1}$$

Or, reorganizing:

$$i_{2t} = \frac{(i_{1t}+i^e_{1t+1})}{2}$$

The two-year interest rate is approximately equal to the average of the current and the expected one-year interest rates.

Proposition 5:

$$(1+x)^n \approx 1+nx \tag{A2.5}$$

The proof follows by repeated application of Propositions 3 and 4. For example, $(1+x)^3 = (1+x)^2(1+x) \approx (1+2x)(1+x)$ by Proposition 4, $\approx (1+2x+x) = 1+3x$ by Proposition 3.

The approximation becomes worse as n increases, however. For example, for $x = 0.05$ and $n = 5$, the approximation gives 1.25, compared to an exact value of 1.2763. For $n = 10$, the approximation gives 1.50, compared to an exact value of 1.63.

Proposition 6:

$$\frac{(1+x)}{(1+y)} \approx (1+x-y) \tag{A2.6}$$

Here is the proof: Consider the product of $(1+x-y)(1+y)$. Expanding this product gives $(1+x-y)(1+y) = 1+x+xy-y^2$. If both x and y are small, then xy and y^2 are very small, so $(1+x-y)(1+y) \approx (1+x)$. Dividing both sides of this approximation by $(1+y)$ gives the preceding proposition.

For the values of $x = 0.05$ and $y = 0.03$, the approximation gives 1.02, while the correct value is nearly the same, 1.019.

Application from Chapter 14: The real interest rate is defined by:

$$(1+r_t) = \frac{(1+i_t)}{(1+\pi^e_{t+1})}$$

Using Proposition 6 gives

$$(1+r_t) \approx (1+i_t-\pi^e_{t+1})$$

Simplifying:

$$r_t \approx i_t - \pi^e_{t+1}$$

This gives us the approximation we use at many points in this text. The real interest rate is approximately equal to the nominal interest rate minus expected inflation.

These approximations are also convenient when dealing with growth rates. Define the rate of growth of x by $g_x \equiv \Delta x/x$, and similarly for z, g_z, and y, g_y. The numerical examples below are based on the values $g_x = 0.05$ and $g_y = 0.03$.

Proposition 7: If $z = xy$ then:

$$g_z \approx g_x + g_y \tag{A2.7}$$

Here is the proof: Let Δz be the increase in z when x increases by Δx and y increases by Δy. Then, by definition:

$$z+\Delta z = (x+\Delta x)(y+\Delta y)$$

Divide both sides by z.
The left side becomes:

$$\frac{(z+\Delta z)}{z} = \left(1+\frac{\Delta z}{z}\right)$$

The right-hand side becomes

$$\frac{(x+\Delta x)(y+\Delta y)}{z} = \frac{(x+\Delta x)}{x}\frac{(y+\Delta y)}{y} = \left(1+\frac{\Delta x}{x}\right)\left(1+\frac{\Delta y}{y}\right)$$

where the first equality follows from the fact that $z = xy$, the second equality from simplifying each of the two fractions.

Using the expressions for the left and right sides gives:

$$\left(1 + \frac{\Delta z}{z}\right) = \left(1 + \frac{\Delta x}{x}\right)\left(1 + \frac{\Delta y}{y}\right)$$

Or, equivalently,

$$(1 + g_z) = (1 + g_x)(1 + g_y)$$

From Proposition 3, $(1 + g_z) \approx (1 + g_x + g_y)$,or, equivalently,

$$g_z \approx g_x + g_y$$

For $g_x = 0.05$ and $g_y = 0.03$, the approximation gives $g_z = 8\%$, while the correct value is 8.15%.

Application from Chapter 13: Let the production function be of the form $Y = NA$, where Y is production, N is employment, and A is productivity. Denoting the growth rates of Y, N, and A by g_Y, g_N, and g_A respectively, Proposition 7 implies

$$g_Y \approx g_N + g_A$$

The rate of output growth is approximately equal to the rate of employment growth plus the rate of productivity growth.

Proposition 8: If $z = x/y$, then

$$g_z \approx g_x - g_y \qquad \text{(A2.8)}$$

Here is the proof: Let Δz be the increase in z, when x increases by Δx and y increases by Δy. Then, by definition:

$$z + \Delta z = \frac{x + \Delta x}{y + \Delta y}$$

Divide both sides by z.
The left side becomes:

$$\left(\frac{z + \Delta z}{z}\right) = \left(1 + \frac{\Delta z}{z}\right)$$

The right side becomes:

$$\frac{(x + \Delta x)}{(y + \Delta y)}\frac{1}{z} = \frac{(x + \Delta x)}{(y + \Delta y)}\frac{y}{x} = \frac{(x + \Delta x)/x}{(y + \Delta y)/y} = \frac{1 + (\Delta x/x)}{1 + (\Delta y/y)}$$

where the first equality comes from the fact that $z = x/y$, the second equality comes from rearranging terms, and the third equality comes from simplifying.

Using the expressions for the left and right sides gives:

$$1 + \Delta z/z = \frac{1 + (\Delta x/x)}{1 + (\Delta y/y)}$$

Or, substituting:

$$1 + g_z = \frac{1 + g_x}{1 + g_y}$$

From Proposition 8, $(1 + g_z) \approx (1 + g_x - g_y)$,or, equivalently,

$$g_z \approx g_x - g_y$$

For $g_x = 0.05$ and $g_y = 0.03$, the approximation gives $g_z = 2\%$, while the correct value is 1.9%.

Application from Chapter 9: Let M be nominal money, P be the price level. It follows that the rate of growth of the real money stock M/P is given by:

$$g_{M/P} \approx g_M - \pi$$

where π is the rate of growth of prices or, equivalently, the rate of inflation.

Functions

We use functions informally in this text, as a way of denoting how a variable depends on one or more other variables.

In some cases, we look at how a variable Y moves with a variable X. We write this relation as

$$Y = \underset{+}{f(X)}$$

A plus sign below X indicates a positive relation; an increase in X leads to an increase in Y. A minus sign below X indicates a negative relation; an increase in X leads to a decrease in Y.

In some cases, we allow the variable Y to depend on more than one variable. For example, we allow Y to depend on X and Z:

$$Y = \underset{(+,-)}{f(X, Z)}$$

The signs indicate that an increase in X leads to an increase in Y, and that an increase in Z leads to a decrease in Y.

An example of such a function is the investment function (5.1) in Chapter 5:

$$I = \underset{(+,-)}{I(Y, i)}$$

This equation says that investment, I, increases with production, Y, and decreases with the interest rate, i.

In some cases, it is reasonable to assume that the relation between two or more variables is a **linear relation**. A given increase in X always leads to the same increase in Y. In that case, the function is given by:

$$Y = a + bX$$

This relation can be represented by a line giving Y for any value of X.

The parameter a gives the value of Y when X is equal to zero. It is called the **intercept** because it gives the value of Y when the line representing the relation "intercepts" (crosses) the vertical axis.

The parameter b tells us by how much Y increases when X increases by one unit. It is called the **slope** because it is equal to the slope of the line representing the relation.

A simple linear relation is the relation $Y = X$, which is represented by the 45-degree line and has a slope of 1. Another example of a linear relation is the consumption function (3.2) in Chapter 3:

$$C = c_0 + c_1 Y_D$$

where C is consumption and Y_D is disposable income. c_0 tells us what consumption would be if disposable income were equal to zero. c_1 tells us by how much consumption increases when income increases by 1 unit; c_1 is called the *marginal propensity to consume*.

Logarithmic Scales

A variable that grows at a constant growth rate increases by larger and larger increments over time. Take a variable X that grows over time at a constant growth rate of, say, 3% per year.

- Start in year 0 and assume $X = 2$. A 3% increase in X represents an increase of $0.06(0.03 \times 2)$.
- Go to year 20. X is now equal to $2(1.03)^{20} = 3.61$. A 3% increase now represents an increase of $0.11(0.03 \times 3.61)$.
- Go to year 100. X is equal to $2(1.03)^{100} = 38.4$. A 3% increase represents an increase of $1.15(0.03 \times 38.4)$, so an increase about 20 times larger than in year 0.

If we plot X against time using a standard (linear) vertical scale, the plot looks like Figure A2-1(a). The increases in X become larger and larger over time (0.06 in year 0, 0.11 in year 20, 1.15 in year 100). The curve representing X against time becomes steeper and steeper.

Another way of representing the evolution of X is to use a *logarithmic scale* to measure X on the vertical axis. The property of a logarithmic scale is that the same *proportional* increase in this variable is represented by the same vertical distance on the scale. So the behavior of a variable such as X that increases by the same proportional increase (3%) each year is now represented by a line. Figure A2-1(b) represents the behavior of X, this time using a logarithmic scale on the vertical axis. The fact that the relation is represented by a

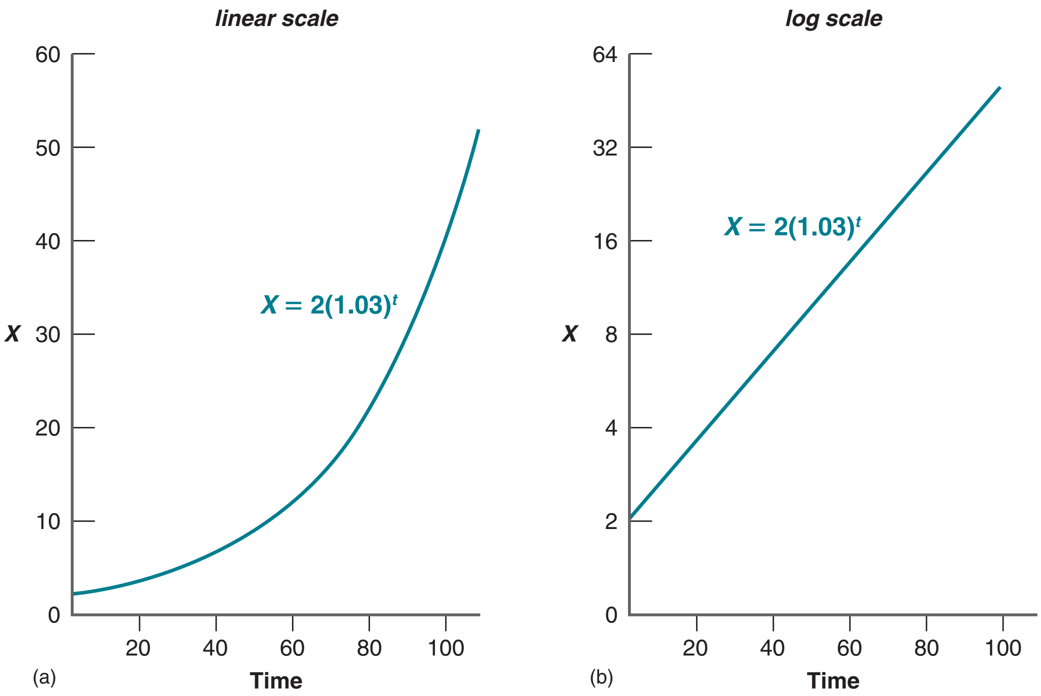


Figure A2-1

(a) The evolution of X (using a linear scale) (b) The evolution of X (using a logarithmic scale)

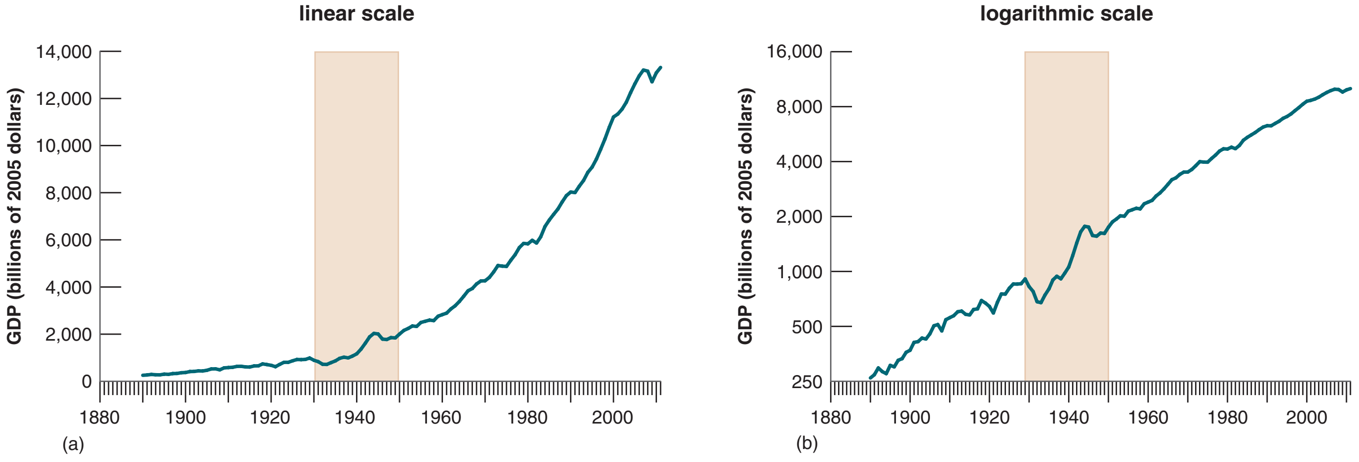


Figure A2-2

(a) US GDP since 1890 (using a linear scale) (b) US GDP since 1890 (using a logarithmic scale)

Source: 1890–1928: Historical Statistics of the United States, Table F1-5, adjusted for level to be consistent with the post-1929 series. 1929–2011 BEA, billions of chained 2005 dollars. www.bea.gov/national/index.htm#gdp.

line indicates that *X* is growing at a constant rate over time. The higher the rate of growth, the steeper the line.

In contrast to *X*, economic variables such as GDP do not grow at a constant growth rate every year. Their growth rate may be higher in some decades, lower in others; a recession may lead to a few years of negative growth. When looking at their evolution over time, it is often more informative to use a logarithmic scale rather than a linear scale. Let's see why.

Figure A2-2(a) plots real US GDP from 1890 to 2011 using a standard (linear) scale. Because real US GDP is about 51 times bigger in 2011 than in 1890, the same proportional increase in GDP is 51 times bigger in 2011 than in 1890. So the curve representing the evolution of GDP over time becomes steeper and steeper over time. It is difficult to see from the figure whether the US economy is growing faster or slower than it was 50 years or 100 years ago.

Figure A2-2(b) plots US GDP from 1890 to 2011 using a logarithmic scale. If the growth rate of GDP was the same every year—so the proportional increase in GDP was the same every year—the evolution of GDP would be represented by a line, the same way the evolution of *X* was represented by a line in Figure A2-1(b). Because the growth rate of GDP is not constant from year to year—so the proportional increase in GDP is not the same every year—the evolution of GDP is no longer represented by a line. Unlike in Figure A2-2(a), GDP does not explode over time, and the graph is more informative. Here are two examples.

- If, in Figure A2-2(b), we were to draw a line to fit the curve from 1890 to 1929, and another line to fit the curve from 1950 to 2011 (the two periods are separated by the shaded area in Figure A2-2(b)), the two lines would have roughly the same slope. What this tells us is that the average growth rate was roughly the same during the two periods.
- The decline in output from 1929 to 1933 is visible in Figure A2-2(b). (By contrast, the Great Recession looks small relative to the Great Depression.) So is the strong recovery of output that follows. By the 1950s, output appears to be back to its old trend line. This suggests that the Great Depression was not associated with a permanently lower level of output.

Note in both cases that you could not have derived these conclusions by looking at Figure A2-2(a), but you can derive them by looking at Figure A2-2(b). This shows the usefulness of using a logarithmic scale.

Key Terms

- linear relation, A-9
- intercept, A-10
- slope, A-10

APPENDIX 3 An Introduction to Econometrics

How do we know that consumption depends on income? How do we know the value of the propensity to consume?

To answer these questions and, more generally, to estimate behavioral relations and find out the values of the relevant parameters, economists use *econometrics*—the set of statistical techniques designed for use in economics. Econometrics can get very technical, but I shall outline in this appendix the basic principles behind it, using as an example the consumption function introduced in Chapter 3. We'll concentrate on estimating c_1, the propensity to consume out of income.

Changes in Consumption and in Disposable Income

The propensity to consume tells us by how much consumption changes for a given change in income. A natural first step is simply to plot changes in consumption versus changes in income and see how the relation between the two looks. (For simplicity, I shall use GDP as a measure of income. Clearly, a better specification would use disposable personal income, as well as other variables such as wealth, and allow consumption to depend not only on current income but on past income as well. I ignore these complications here.)

You can see the relation in Figure A3-1.

The vertical axis in Figure A3-1 measures the annual change in consumption minus the average annual change in consumption, for each year from 1970 to 2018. More precisely:

Let C_t denote consumption in year t. Let ΔC_t denote $C_t - C_{t-1}$, the change in consumption from year $t - 1$ to year t. Let $\overline{\Delta C}$ denote the average annual change in consumption since 1970. The variable measured on the vertical axis is constructed as $\Delta C_t - \overline{\Delta C}$. A positive value of the variable represents an increase in consumption larger than average, whereas a negative value represents an increase in consumption smaller than average.

Similarly, the horizontal axis measures the annual change in income, minus the average annual change in income since 1970, $\Delta Y_t - \overline{\Delta Y}$.

A particular dot in the figure gives the deviations of the change in consumption and income from their respective means for a particular year between 1970 and 2018. In 2018, for example, the change in consumption was higher than average by \$127 billion, and the change in income was higher than average by \$237 billion. (The point corresponding to 2018 is indicated by the red dot. For our purposes, it is not important to know which year each dot refers to, just what the set of points in the diagram looks like. So, except for 2018, the years are not indicated in Figure A3-1.)

Figure A3-1 suggests two main conclusions.

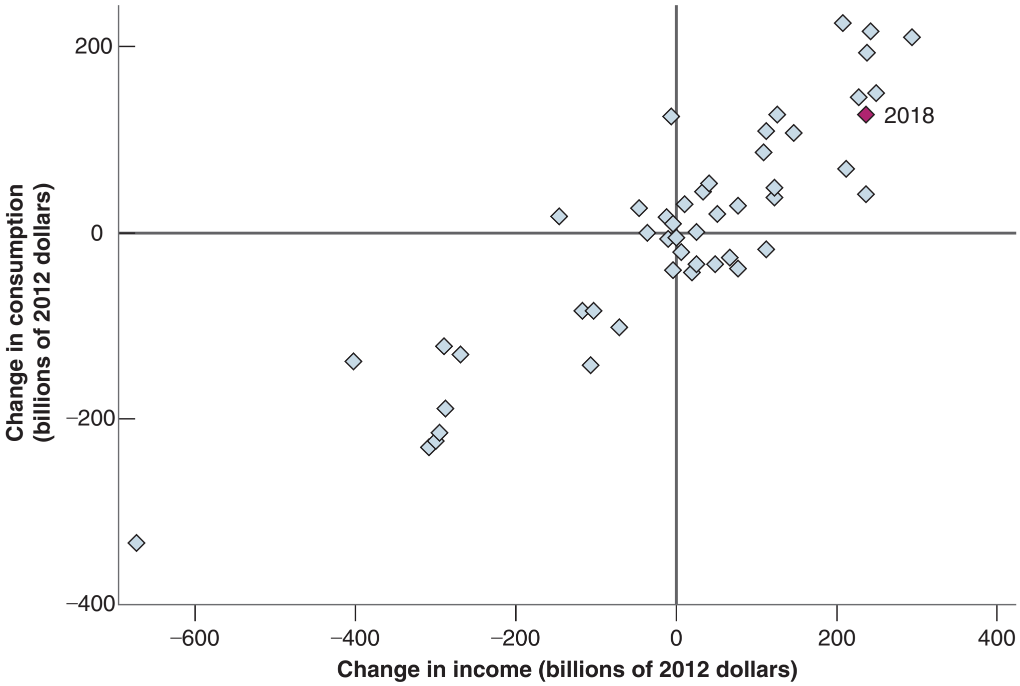
- One, there is a clear positive relation between changes in consumption and changes in income. Most of the

Figure A3-1

Changes in Consumption versus Changes in Income, 1970–2018

There is a clear positive relation between changes in consumption and changes in income.

Source: FRED: PCECCA, GDPC1



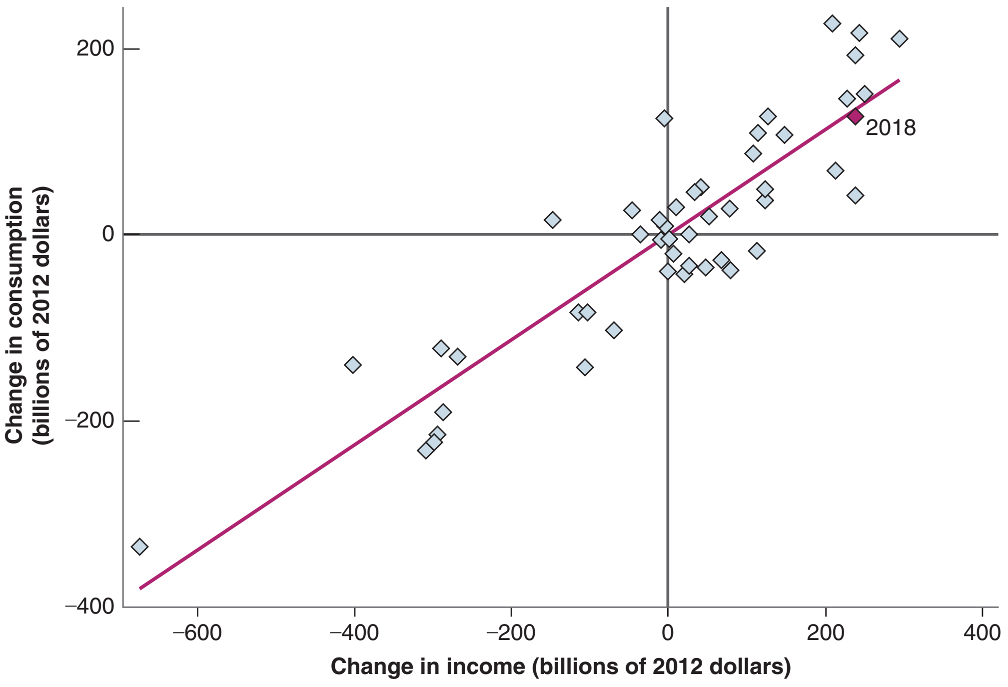


Figure A3-2

Changes in Consumption and Changes in Income: The Regression Line

The regression line is the line that best fits the scatter of points.

points lie in the upper-right and lower-left quadrants of the figure. When income increases by more than average, consumption also typically increases by more than average; when income increases by less than average, so typically does consumption.

- Two, the relation between the two variables is good but not perfect. In particular, some points lie in the upper-left quadrant; these points correspond to years when smaller-than-average changes in income were associated with higher-than-average changes in consumption.

Econometrics allows us to state these two conclusions more precisely and to get an estimate of the propensity to consume. Using an econometrics software package, we can find the line that best fits the cloud of points in Figure A3-1. This line-fitting process is called **ordinary least squares (OLS)**.[1] The estimated equation corresponding to the line is called a **regression**, and the line itself is called the **regression line**.

In this case, the estimated equation is given by

$$(\Delta C_t - \overline{\Delta C}) = 0.56(\Delta Y_t - \overline{\Delta Y}) + \text{residual}$$

$$\overline{R}^2 = 0.81 \qquad \text{(A3.1)}$$

The regression line corresponding to this estimated equation is drawn in Figure A3-2. Equation (A3.1) reports two important numbers. (Econometrics packages give more information than that reported above; a typical printout, together with further explanations, is given in the Focus Box "A Guide to Understanding Econometric Results.")

- The first important number is the estimated propensity to consume. The equation tells us that an increase in income of $1 billion above normal is typically associated with an increase in consumption of $0.56 billion above normal. In other words, the estimated propensity to consume is 0.56. It is positive but smaller than 1.
- The second important number is $\overline{R}^2$, which is a measure of how well the regression line fits.

Having estimated the effect of income on consumption, we can decompose the change in consumption for each year into the part that is due to the change in income—the first term on the right in equation (A3.1)—and the rest, which is called the **residual**. For example, the residual for 2018 is indicated in Figure A3-2 by the vertical distance from the point representing 2018 to the regression line. It turns out that the point for 2018 is nearly on the regression line (this is just a coincidence), so the residual is very small. If all the points in Figure A3-2 were exactly on the estimated line, all residuals would be zero; all changes in consumption would be explained by changes in income. As you can see, however, this is not the case. $\overline{R}^2$ is a statistic that tells us how well the line fits. $\overline{R}^2$ is always between 0 and 1. A value of 1 would imply that the relation between the two variables is perfect, that all points are exactly on the regression line. A value of

[1]The term *least squares* comes from the fact that the line minimizes the sum of the squared distances of the points to the line—thus it gives the "least squares." The word *ordinary* comes from the fact that this is the simplest method used in econometrics.

FOCUS

A Guide to Understanding Econometric Results

In your readings, you may run across results of estimation using econometrics. Here is a guide, which uses the slightly simplified but otherwise untouched computer output for equation (A3.1).

$\bar{R}^2$ is a measure of fit. The closer to 1, the better the fit of the regression line. A value of 0.81 indicates that much but not all of the movement in the dependent variable can be explained by movements in the independent variables.

The period of estimation includes all years from 1970 to 2018. There are therefore 49 **usable observations** used in the regression. **Degrees of freedom** is the number of observations minus the number of parameters to be estimated. There is one estimated parameter here: the coefficient on ΔY. Thus, there are $49 - 1 = 48$ degrees of freedom. A simple rule is that one needs at least as many observations as parameters to be estimated, and preferably many more; put another way, degrees of freedom must be positive, and the larger the better.

The variable we are trying to explain is called the **dependent variable.** Here the dependent variable is $\Delta \bar{C}_t$—the annual change in consumption minus its mean.

Dependent Variable $\Delta C_t - \Delta \bar{C}$—Estimation by Least Squares
Annual Data from 1970 to 2018
Usable Observations: 49
Degrees of Freedom: 48
$R^2 = 0.81$

Variable	Coefficient	*t*-Statistic
$\Delta Y_t - \Delta \bar{Y}$	**0.56**	**14.1**

The variables that we use to explain the dependent variable are called the **independent variables.** Here there is only one independent variable, $\Delta Y_t - \Delta \bar{Y}$—the annual change in income minus its mean.

For each independent variable, the computer gives the estimated coefficient, as well as a ***t*-statistic.** The *t*-statistic associated with each estimated coefficient tells us how confident we can be that the true coefficient is different from zero. A *t*-statistic greater than 2 indicates that we can be at least 95% sure that the true coefficient is different from zero. A *t*-statistic of 14.1 (on the coefficient associated with income) is so high that we can be nearly completely sure (more than 99.99% sure) that the true coefficient is different from zero.

zero would imply that the computer can see no relation between the two variables. The value of $\bar{R}^2$ of 0.81 in equation (A3.1) is high, confirming the message from Figure A3-2: Movements in income clearly affect consumption, but there is still quite a bit of movement in consumption that cannot be explained by movements in income.

Correlation versus Causality

We have established that consumption and income typically move together. More formally, we have seen that there is a positive **correlation**—the technical term for *co-relation*—between annual changes in consumption and annual changes in income. And we have interpreted this relation as showing **causality**—that an increase in income causes an increase in consumption.

We need to think again about this interpretation. A positive relation between consumption and income may reflect the effect of income on consumption. But it may also reflect the effect of consumption on income. Indeed, the model we developed in Chapter 3 tells us that if, for any reason, consumers decide to spend more, then output and therefore income will increase. If part of the relation between consumption and income comes from the effect of consumption on income, interpreting equation (A3.1) as telling us about the effect of income on consumption is not right.

An example will help here. Suppose consumption

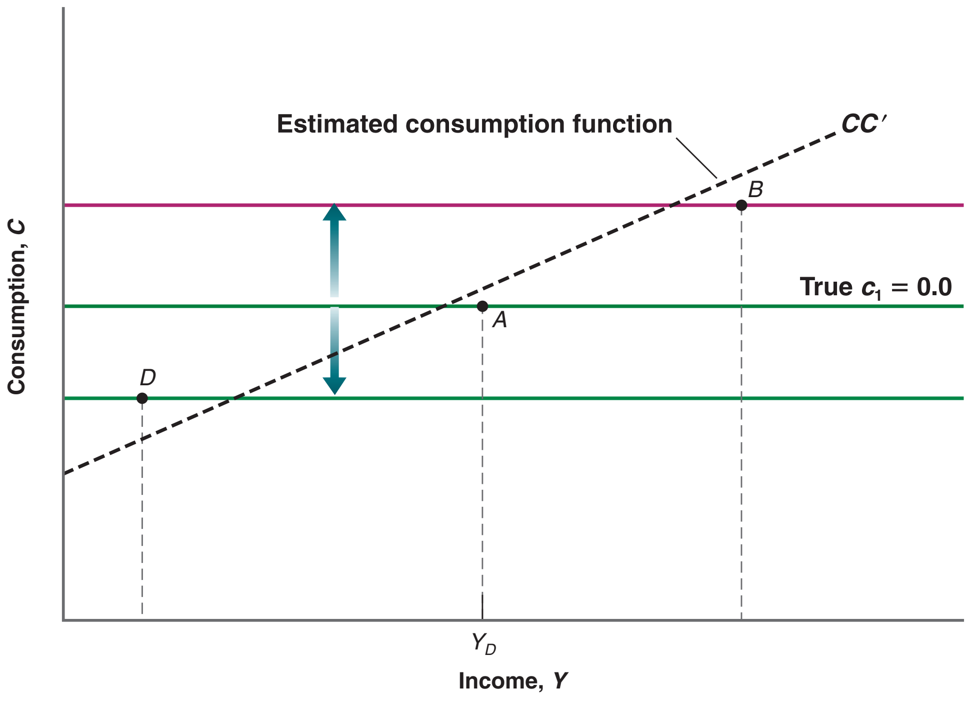


Figure A3-3

A Misleading Regression

The relation between income and consumption comes from the effect of consumption on income rather than from the effect of income on consumption.

does not depend on income, so that the true value of c_1 is zero. (This is not realistic, but it will make the point most clearly.) So draw the consumption function as a horizontal line (a line with a zero slope) in Figure A3-3. Next, suppose income equal to Y, so that the initial combination of consumption and income is given by point A.

Now suppose that, because of improved confidence, consumers increase their consumption, so the consumption line shifts up. If demand affects output, output and therefore income will increase, so that the new combination of consumption and income will be given by, say, point B. If, instead, consumers become more pessimistic, the consumption line will shift down, and so will output, leading to a combination of consumption and income given by, say, point D.

If we look at the economy described in the previous two paragraphs, we observe points A, B, and D. If, as we did previously, we draw the best-fitting line through these points, we shall estimate an upward-sloping line, such as CC', and thus a positive value for propensity to consume, c_1. Remember, however, that the true value of c_1 is zero. Why do we get the wrong answer—a positive value for c_1—when the true value is zero? Because we interpret the positive relation between income and consumption as showing the effect of income on consumption, where, in fact, the relation reflects the effect of consumption on income; higher consumption leads to higher demand, higher output, and so higher income.

There is an important lesson here: *the difference between correlation and causality*. The fact that two variables move together does not imply that movements in the first variable cause movements in the second variable. Perhaps the causality runs the other way: movements in the second variable cause movements in the first variable. Or perhaps, as is likely to be the case here, the causality runs both ways: income affects consumption, *and* consumption affects income.

Is there a way out of the correlation-versus-causality problem? If we are interested—and we are—in the effect of income on consumption, can we still learn that from the data? The answer: yes, but only by using more information.

Suppose we *knew* that a specific change in income was not caused by a change in consumption. Then, by looking at the reaction of consumption to *this* change in income, we could learn how consumption responds to income; we could estimate the propensity to consume.

This answer would seem to simply assume away the problem. How can we tell that a change in income is not due to a change in consumption? In fact, sometimes, we can tell. Suppose, for example, that there is an increase in exports due to strong foreign demand, leading to an increase in output and therefore an increase in income. In that case, if both income and consumption increase, it is likely that the movement in consumption reflects the effect of income on consumption, and we can then estimate the propensity to consume.

This example suggests a general strategy.

- Find exogenous variables—that is, variables that affect income but are not in turn affected by it. Exports is a decent candidate in this case. Another one might be defense spending.
- Look at the change in consumption in response not to all changes in income—as we did in the previous regression—but in response to the changes in income that can be explained by changes in these exogenous variables.

By following this strategy, we can be reasonably confident that what we are estimating is the effect of income on consumption, and not the other way around.

The problem of finding such exogenous variables is known as the **identification problem** in econometrics. These exogenous variables, when they can be found, are called **instruments**. Methods of estimation that rely on the use of such instruments are called **instrumental variable methods**.

When equation (A3.1) is estimated using an instrumental variable method—using exports as the instrument—rather than OLS as we did previously, the estimated equation becomes

$$(\Delta C_t - \overline{\Delta C}) = 0.41(\Delta Y_t - \overline{\Delta Y})$$

Note that the coefficient on income, 0.41, is smaller than the 0.56 in equation (A3.1). This decrease in the estimated propensity to consume is what we would expect. Our previous estimate in equation (A3.1) reflected not only the effect of income on consumption but also the effect of consumption back on income. The use of an instrument eliminates this second effect, which is why we find a smaller estimated effect of income on consumption.

This short introduction to econometrics is no substitute for a course in econometrics. But it gives you a sense of how economists use data to estimate relations and parameters and to identify causal relations between economic variables.

Key Terms

- ordinary least squares (OLS), A-13
- regression, A-13
- regression line, A-13
- residual, A-13
- $\overline{R}^2$, A-13
- usable observations, A-14
- degrees of freedom, A-14
- dependent variable, A-14
- independent variables, A-14
- *t*-statistic, A-14
- correlation, A-14
- causality, A-14
- identification problem, A-16
- instruments, A-16
- instrumental variable methods, A-16

Glossary

above the line, below the line In the balance of payments, the items in the current account are above the line that divides them from the items in the financial account, which appear below the line.

accelerationist Phillips curve See *modified Phillips curve*.

adaptive expectations A backward-looking method of forming expectations by adjusting for past mistakes.

aggregate output The total amount of output produced in the economy.

aggregate private spending The sum of all nongovernment spending. Also called *private spending*.

aggregate production function The relation between the quantity of aggregate output produced and the quantities of inputs used in production.

American Recovery and Reinvestment Act (ARRA) The fiscal stimulus program introduced in February 2009 by the US administration.

anchored Inflation expectations are said to be anchored if they do not respond to actual inflation (related: de-anchored, re-anchored, unanchored).

animal spirits A term introduced by Keynes to refer to movements in consumption or investment that could not be explained by movements in current variables.

appreciation (nominal) An increase in the value of domestic currency in terms of foreign currency. Corresponds to an increase in the exchange rate E, as defined in this text.

appropriability (of research results) The extent to which firms benefit from the results of their research and development efforts.

arbitrage The proposition that the expected rates of return on two financial assets must be equal. Also called *risky arbitrage* to distinguish it from riskless arbitrage, the proposition that the actual rates of return on two financial assets must be the same.

automatic stabilizer The fact that a decrease in output leads, under given tax and spending rules, to an increase in the budget deficit. This increase in the budget deficit in turn increases demand and thus stabilizes output.

autonomous spending The component of the demand for goods that does not depend on the level of output.

backloading A policy is backloaded if it is to be implemented in the future rather than in the present.

balance of payments A set of accounts that summarize a country's transactions with the rest of the world.

balanced budget A budget in which taxes are equal to government spending.

balanced growth The situation in which output, capital, and effective labor all grow at the same rate.

bands Upper and lower limits on exchange rate movements.

bank reserves Holdings of central bank money by banks. The difference between what banks receive from depositors and what they lend to firms or hold as bonds.

bank run Simultaneous attempts by depositors to withdraw their funds from a bank.

bargaining power The relative strength of each side in a negotiation or a dispute.

base year When constructing real GDP by evaluating quantities in different years using a given set of prices, the year to which this given set of prices corresponds.

Basel II, Basel III International accords about the regulation of the banking sector.

basis points A basis point is a hundredth of a percent. An increase of the interest rate by 100 basis points is a 1% increase in the interest rate.

behavioral equation An equation that captures some aspect of behavior.

bilateral exchange rate The real exchange rate between two countries.

bond A financial asset that promises a stream of known payments over some period of time.

bond rating The assessment of a bond based on its default risk.

borrowing rate The rate at which consumers or firms can borrow from a financial institution.

bracket creep The increase in the marginal tax rate faced by individuals as their nominal income goes up and tax brackets remain unchanged in nominal terms.

budget deficit The excess of government expenditures over government revenues.

budget surplus See *public saving*.

business cycle theory The study of macroeconomic fluctuations.

business cycles See *output fluctuations*.

capital account Net capital transfers to and from the rest of the world.

capital accumulation Increase in the capital stock.

capital controls Restrictions on the foreign assets domestic residents can hold and on the domestic assets foreigners can hold.

capital ratio Ratio of the capital of a bank to its assets.

cash flow The net flow of cash a firm is receiving.

causality A relation between cause and effect.

central bank money Money issued by the central bank. Also known as the *monetary base* and *high-powered money*.

central parity The reference value around which the exchange rate is allowed to move under a fixed exchange rate system. The center of the band.

change in business inventories In the national income and product accounts, the change in the volume of inventories held by businesses.

checkable deposits Deposits at banks and other financial institutions against which checks can be written.

churn The concept that new goods make old goods obsolete, that new production techniques make older techniques and worker skills obsolete, and so on (related: churning).

Cobb-Douglas production function A production function giving output as a weighted geometric average of labor and capital.

collateral The asset pledged in order to get a loan. In case of default, the asset goes to the lender.

collateralized debt obligation (CDO) Security based on an underlying portfolio of assets.

collective bargaining Wage bargaining between unions and firms.

common currency The currency used in the countries that are members of a common currency area.

compensation of employees In the national income and product accounts, the sum of wages and salaries and of supplements to wages and salaries.

confidence band When estimating the dynamic effect of one variable on another, the range of values where we can be confident the true dynamic effect lies.

Congressional Budget Office (CBO) An office of Congress in charge of constructing and publishing budget projections.

constant returns to scale The proposition that a proportional increase (or decrease) of all inputs leads to the same proportional increase (or decrease) in output.

consumer price index (CPI) The cost of a given list of goods and services consumed by a typical urban dweller.

consumption (*C*) Goods and services purchased by consumers.

consumption function A function that relates consumption to its determinants.

consumption of fixed capital Depreciation of capital.

contractionary open market operation An open market operation in which the central bank sells bonds to decrease the money supply.

conventional monetary policy The use of the policy rate as the main instrument to affect economic activity.

convergence The tendency for countries with lower output per capita to grow faster, leading to convergence of output per capita across countries.

core inflation rate Inflation rate excluding volatile prices, such as the prices of food and energy.

corporate bond A bond issued by a corporation.

corporate profits and business transfers In the national income and product accounts, firms' revenues minus costs (including interest payments) and minus depreciation.

correlation A measure of the way two variables move together. A positive correlation indicates that the two variables tend to move in the same direction. A negative correlation indicates that the two variables tend to move in opposite directions. A correlation of zero indicates that there is no apparent relation between the two variables.

corridor system A monetary arrangement where the central bank sets two rates: a rate at which it lends to banks, and a lower rate at which banks can lend to the central bank.

cost of living The average price of a consumption bundle.

coupon bond A bond that promises multiple payments before maturity and one payment at maturity.

coupon payments The payments before maturity on a coupon bond.

coupon rate The ratio of the coupon payment to the face value of a coupon bond.

crawling peg An exchange rate mechanism in which the exchange rate is allowed to move over time according to a prespecified formula.

creative destruction The proposition that growth simultaneously creates and destroys jobs.

credibility The degree to which people and markets believe that a policy announcement will actually be implemented and followed through.

credit easing Monetary policy measures aimed at increasing the supply of credit by banks.

currency Coins and bills.

currency board An exchange rate system in which (i) the central bank stands ready to buy or sell foreign currency at the official exchange rate; (ii) the central bank cannot engage in open market operations, that is, buy or sell government bonds.

current account In the balance of payments, the summary of a country's payments to and from the rest of the world.

current account balance The sum of net exports, net income, and net transfers from the rest of the world.

current account deficit A negative current account balance.

current account surplus A positive current account balance.

Current Population Survey (CPS) A large monthly survey of US households used, in particular, to compute the unemployment rate.

current yield The ratio of the coupon payment to the price of a coupon bond.

cyclically adjusted deficit A measure of what the government deficit would be under existing tax and spending rules, if output were at its natural level. Also called a *full-employment deficit, mid-cycle deficit, standardized employment deficit*, or *structural deficit*.

debt finance Financing based on loans or the issuance of bonds.

debt monetization The printing of money by the central bank to finance a deficit.

debt ratio See *debt-to-GDP ratio*.

debt rescheduling The rescheduling of interest payments or payment of principal, typically to decrease current payments.

debt restructuring A decrease in the value of a debt, through a decrease in the value of the principal, or a decrease in interest payments.

debt-to-GDP ratio The ratio of debt to gross domestic product. Also called simply the *debt ratio*.

decreasing returns to capital The property that increases in capital lead to smaller and smaller increases in output as the level of capital increases.

decreasing returns to labor The property that increases in labor lead to smaller and smaller increases in output as the level of labor increases.

default risk The risk that debt will not be repaid in full.

deflation Sustained decline in the price level; negative inflation.

deflation spiral A mechanism through which deflation increases the real interest rate, which in turn leads to lower activity, and leads to further deflation, a further increase in the real interest rate, and so on.

deflation trap The situation of a country subject to a deflation spiral.

degrees of freedom The number of usable observations in a regression minus the number of parameters to be estimated.

demand for domestic goods The demand for domestic goods by people, firms, and governments, both domestic and foreign. Equal to the domestic demand for goods plus net exports.

demand deposits A bank account that allows depositors to write checks or get cash on demand, up to an amount equal to the account balance. See *checkable deposits*.

dependent variable A variable whose value is determined by one or more other variables.

depreciation (nominal) A decrease in the value of domestic currency in terms of a foreign currency. Corresponds to a decrease in the exchange rate *E*, as defined in this text.

devaluation A decrease in the exchange rate (*E*) in a fixed exchange rate system.

direct finance Financing through markets, through the issuance of bonds or equities.

discount bond A bond that promises a single payment at maturity.

discount factor The value today of a dollar (or other national currency unit) at some time in the future.

discount rate (i) The interest rate used to discount a sequence of future payments. Equal to the nominal interest rate when discounting future nominal payments and to the real interest rate when discounting future real payments. (ii) The interest rate at which the Fed lends to banks.

discouraged worker A person who has given up looking for employment.

disposable income The income that remains once consumers have received transfers from the government and paid their taxes.

dividends The portion of a corporation's profits that the firm pays out each period to its shareholders.

divine coincidence The proposition that, if inflation remains stable, this is a signal that output is equal to potential output.

dollar GDP See *nominal GDP*.

dollarization The use of dollars in domestic transactions in a country other than the United States.

domestic demand for goods The sum of consumption, investment, and government spending.

durable goods Commodities that can be stored and have an average life of at least three years.

duration of unemployment The period of time during which a worker is unemployed.

dynamics Movements of one or more economic variables over time.

Easterlin paradox The proposition that higher income in a country is not associated with higher levels of happiness.

econometrics Statistical methods applied to economics.

effective demand Synonym for *aggregate demand*.

effective labor The number of workers in an economy times the state of technology.

effective real exchange rate See *multilateral exchange rate*.

efficiency wage theory A theory that argues that a higher wage may lead workers to be more engaged and more productive.

employment The number of people employed.

employment protection The set of regulations determining the conditions under which a firm can lay off a worker.

employment rate The ratio of employment to the labor force.

endogenous variable A variable that depends on other variables in a model and is thus explained within the model.

equilibrium Equality between demand and supply (production).

equilibrium condition The condition that supply be equal to demand.

equilibrium in the goods market The condition that the supply of goods be equal to the demand for goods.

equipment and software investment The purchase of machines and software by firms.

equity finance Financing based on the issuance of shares.

equity premium Risk premium required by investors to hold stocks rather than short-term bonds.

euro A European currency that replaced national currencies in 11 countries in 2002 and is now used in 19 countries.

euro area The set of countries that share the euro as a common currency.

European Central Bank (ECB) The central bank, located in Frankfurt, in charge of determining monetary policy in the euro area.

European Monetary System (EMS) A series of rules that implemented bands for bilateral exchange rates between member countries in Europe from 1979 to 1998.

European Union A political and economic organization of 28 European nations. Formerly called the European Community.

ex-dividend price The price of the stock just after the dividend has been paid.

exogenous variable A variable that is not explained within a model, but rather is taken as given.

expansion A period of positive GDP growth.

expansionary open market operation An open market operation in which the central bank buys bonds to increase the money supply.

expectations hypothesis The hypothesis that financial investors are risk neutral, which implies that expected returns on all financial assets have to be equal.

expectations-augmented Phillips curve See *modified Phillips curve*.

expected present discount value The value today of current and expected future payments.

expected present discounted value The value today of an expected sequence of future payments; also called *present discounted value* or *present value*.

exports (*X*) The purchases of domestic goods and services by foreigners.

extension agreements Agreements to extend the result of negotiations between a set of unions and firms to all firms in a given sector.

external finance Financing of firms through external funds (as opposed to retained earnings).

face value (on a bond) The single payment at maturity promised by a discount bond.

fad A period of time during which, for reasons of fashion or overoptimism, financial investors are willing to pay more for a stock than its fundamental value.

federal deposit insurance Insurance provided by the US government that protects each bank depositor up to $250,000 per account.

federal funds market The market where banks that have excess reserves at the end of the day lend them to banks that have insufficient reserves.

federal funds rate The interest rate determined by equilibrium in the federal funds market. The interest rate affected most directly by changes in monetary policy.

Federal Reserve Bank (Fed) The US central bank.

fertility of research The degree to which spending on research and development translates into new ideas and new products.

final good A good that is used directly for consumption or investment (as opposed to intermediate goods, which are used in the process of production).

financial account The account showing the financial transactions of a country with the rest of the world.

financial account balance The difference between what a country borrows from the rest of the world and what it lends to the rest of the world. Also, the difference between the change in domestic holdings of foreign assets, and the change in foreign holdings of domestic assets (related: financial account deficit, financial account surplus).

financial account deficit A negative financial account balance.

financial account surplus A positive financial account balance. The country borrows more from the rest of the world than it lends to the rest of the world. A financial account surplus corresponds to a current account deficit.

financial intermediary A financial institution that receives funds from people, firms, or other financial institutions and uses these funds to make loans or buy financial assets.

financial investment The purchase of financial assets.

financial wealth The value of all of one's financial assets minus all financial liabilities. Sometimes called *wealth*, for short.

fine-tuning A macroeconomic policy aimed at precisely hitting a given target, such as constant unemployment or constant output growth.

fire sale prices Very low asset prices, reflecting the need for sellers to sell, and the absence of sufficient buyers, because of liquidity constraints.

fiscal austerity A reduction in public spending or an increase in taxes, aimed at reducing the budget deficit.

fiscal consolidation See *fiscal contraction*.

fiscal contraction A policy aimed at reducing the budget deficit through a decrease in government spending or an increase in taxation. Also called *fiscal consolidation*.

fiscal dominance A situation in which monetary policy becomes subordinated to fiscal policy; for example, when the central bank issues money to finance the deficit.

fiscal expansion An increase in government spending or a decrease in taxation, which leads to an increase in the budget deficit.

fiscal multiplier The size of the effect of government spending on output.

fiscal policy A government's choice of taxes and spending.

fixed exchange rate An exchange rate between the currencies of two or more countries that is fixed at some level and adjusted only infrequently.

fixed investment The purchase of equipment and structures (as opposed to inventory investment). See *investment*

flexible inflation targeting A way of conducting monetary policy to return inflation to target inflation over time.

float The exchange rate is said to float when it is determined in the foreign exchange market, without central bank intervention.

flow A variable that can be expressed as a quantity per unit of time (such as income).

force of compounding The large effects of sustained growth on the level of a variable.

foreign direct investment The purchase of existing firms or assets, or the development of new firms by foreign investors.

foreign exchange Foreign currency; all currencies other than the domestic currency of a given country.

foreign exchange reserves Foreign assets held by the central bank.

four tigers The four Asian economies of Singapore, Taiwan, Hong Kong, and South Korea.

full-employment deficit See *cyclically adjusted deficit*.

fully funded social security system A retirement system in which the contributions of current workers are invested in financial assets, with the proceeds (principal and interest) given back to the workers when they retire.

fundamental value (of a stock) The present value of expected dividends.

G20 The group of 20 countries, representing about 85% of world production, which met regularly during the crisis and served as a forum for coordination of economic policies.

game Strategic interactions between players.

game theory The prediction of outcomes from games.

GDP adjusted for inflation See *real GDP*.

GDP deflator The ratio of nominal GDP to real GDP; a measure of the overall price level. Gives the average price of the final goods produced in the economy.

GDP growth The growth rate of real GDP in year t; equal to $(Y_t - Y_{t-1})/Y_{t-1}$.

GDP in chained (2012) dollars See *real GDP*.

GDP in constant dollars See *real GDP*.

GDP in current dollars See *nominal GDP*.

GDP in terms of goods See *real GDP*.

general-purpose technologies Technologies that have applications across many sectors.

geometric series A mathematical sequence in which the ratio of one term to the preceding term remains the same. A sequence of the form $1 + c + c^2 + \cdots + c^n$.

Gini coefficient A measure of inequality. A Gini coefficient of zero corresponds to complete equality. A Gini coefficient of 1 corresponds to complete inequality (all the income going to one individual).

GNP See *gross national product*.

gold standard A system in which a country fixed the price of its currency in terms of gold and stood ready to exchange gold for currency at the stated parity.

golden-rule level of capital The level of capital at which steady-state consumption is maximized.

government bond A bond issued by a government or a government agency.

government budget constraint The budget constraint faced by the government. The constraint implies that an excess of spending over revenues must be financed by borrowing, and thus leads to an increase in debt.

government purchases In the national income and product accounts, the sum of the purchases of goods by the government plus compensation of government employees.

government spending (*G*) The goods and services purchased by federal, state, and local governments.

government transfers Payments made by the government to individuals that are not in exchange for goods or services (e.g., Social Security payments).

Great Financial Crisis The financial crisis which led to a worldwide recession in 2008 and 2009.

Great Moderation The period from the mid-1980s to the mid-2000s when the volatility of output and the volatility of inflation both declined.

Great Recession The worldwide recession, triggered by the financial crisis, which started in 2008.

gross debt The value of the government's financial liabilities (as opposed to net debt, the value of the government's financial liabilities minus the value of the government's financial assets).

gross domestic product (GDP) A measure of aggregate output in the national income and product accounts; the market value of the goods and services produced by labor and property located in the United States.

gross domestic product (GDP) (versus gross national product (GNP)) Gross domestic product measures value added domestically. Gross national product measures value added by domestic factors of production.

gross national product (GNP) A measure of aggregate output in the national income and product accounts; the market value of the goods and services produced by labor and property supplied by US residents.

gross private domestic fixed investment In the national income and product accounts, the sum of nonresidential investment and residential investment.

growth The steady increase in aggregate output over time.

haircut A reduction in the nominal value of debt.

hard peg A fixed exchange rate regime, with a strong commitment of the central bank to maintain the exchange rate fixed.

hedonic pricing An approach to calculating real GDP that treats goods as providing a collection of characteristics, each with an implicit price.

high-powered money See *central bank money*.

hires Workers newly employed by firms.

housing wealth The value of the housing stock.

human capital The skills possessed by the workers in an economy.

human wealth The labor-income component of wealth.

hyperinflation Very high inflation.

hysteresis A permanent effect of temporary shocks; for example, the permanent effects of a recession on labor force participation.

identification problem In econometrics, the problem of finding whether correlation between variables X and Y indicates a causal relation from X to Y, or from Y to X, or both. This problem is solved by finding exogenous variables, called instruments, that affect X and do not affect Y directly, or affect Y and do not affect X directly.

identity An equation that holds by definition, denoted by the $\equiv$ sign.

import compression The decrease in imports due to a decrease in domestic demand.

imports (Q) The purchases of foreign goods and services by domestic consumers, firms, and the government.

inclusive growth Growth that benefits all.

income The flow of revenue from work, rental income, interest, and dividends.

independent central bank A central bank that makes decisions independently of the government.

independent variable A variable that is taken as given in a relation or in a model.

index number A number, such as the GDP deflator, that has no natural level and is thus set to equal some value (typically 1 or 100) in a given period.

indexed bond A bond that promises payments adjusted for inflation.

indirect taxes Taxes on goods and services; in the United States, primarily sales taxes.

inflation A sustained rise in the general level of prices.

inflation rate The rate at which the price level increases over time.

inflation targeting The conduct of monetary policy to achieve a given inflation rate over time.

inflation-adjusted deficit The correct economic measure of the budget deficit: The sum of the primary deficit and real interest payments (as opposed to the official deficit, which is the sum of the primary deficit and nominal interest payments).

insolvency The inability of a debtor, be it a firm, a person, or the government, to repay its debt.

instrumental variable methods In econometrics, methods of estimation that use instruments to estimate causal relations between different variables.

instruments In econometrics, the exogenous variables that allow the identification problem to be solved.

intercept In a linear relation between two variables, the value of the first variable when the second variable is equal to zero.

interest parity condition See *uncovered interest parity*.

interest rate rule A monetary policy rule in which the interest rate is adjusted in response to output and to inflation.

intermediate good A good used in the production of a final good.

internal finance Financing of firms through internal funds (retained earnings).

International Monetary Fund (IMF) The principal international economic organization; publishes the *World Economic Outlook* annually and the *International Financial Statistics* (IFS) monthly.

inventory investment The difference between production and sales.

investment (I) Purchases of new houses and apartments by people, and purchases of new capital goods (machines and plants) by firms.

IS curve A downward-sloping curve relating output to the interest rate. The curve corresponding to the IS relation, the equilibrium condition for the goods market.

IS relation An equilibrium condition stating that the demand for goods must be equal to the supply of goods, or equivalently that investment must be equal to saving; the equilibrium condition for the goods market.

IS-LM model A model based on equilibrium in the goods and the financial markets in a closed economy.

J-curve A curve depicting the initial deterioration in the trade balance caused by a real depreciation, followed by an improvement in the trade balance.

junior securities Securities that are repaid after senior securities in case of insolvency.

junk bond A bond with a high risk of default.

labor force The number of people who are either working or looking for work.

labor hoarding The decision by firms to keep some excess workers in response to a decrease in sales.

labor in efficiency units See *effective labor*.

labor market rigidities Restrictions on firms' ability to adjust their level of employment.

labor productivity The ratio of output to the number of workers.

layoffs Workers who lose their jobs either temporarily or permanently.

lender of last resort If a solvent bank cannot finance itself, it can borrow from the central bank, which acts as a lender of last resort.

leverage ratio Ratio of the assets of the bank to its capital (the inverse of the capital ratio).

life (of a bond) The length of time during which the bond pays interest, which ends with the repayment of principal.

life-cycle theory of consumption The theory of consumption, developed initially by Franco Modigliani, that emphasizes that the planning horizon of consumers is their lifetime.

linear relation A relation between two variables such that a one-unit increase in one variable always leads to an increase of n units in the other variable.

liquidity An asset is liquid if it can be sold quickly. A financial institution is liquid if it can sell its assets quickly.

liquidity facilities The specific ways in which a central bank can lend to financial institutions.

liquidity preference The term introduced by Keynes to denote the demand for money.

liquidity provision The provision of liquidity to banks by the central bank.

liquidity trap The case where nominal interest rates are equal to zero, and monetary policy cannot, therefore, decrease them further.

LM curve An upward-sloping curve relating the interest rate to output; the curve corresponding to the *LM* relation, the equilibrium condition for financial markets.

LM relation An equilibrium condition stating that the demand for money must be equal to the supply of money.

loan-to-value (LTV) ratio The ratio of the loan that people can take as a proportion of the value of the house or apartment they buy.

logarithmic scale A scale in which the same proportional increase is represented by the same distance on the scale, so that a variable that grows at a constant rate is represented by a straight line.

long run A period of time extending over decades.

long-term interest rate The interest rate on long-term bonds.

Lucas critique The proposition, put forth by Robert Lucas, that existing relations between economic variables may change when policy changes. An example is the apparent trade-off between inflation and unemployment, which may disappear if policymakers try to exploit it.

M1 The sum of currency, traveler's checks, and checkable deposits—assets that can be used directly in transactions; also called *narrow money*.

Maastricht treaty A treaty signed in 1991 that defined the steps involved in the transition to a common currency for the European Union.

macroprudential tools The instruments used to regulate the financial system, such as loan-to-value ratios or capital ratio requirements.

Malthusian trap The case of an economy where increases in productivity lead to a decrease in mortality and an increase in population, leaving income per person unchanged.

market income inequality Inequality of revenues before transfers and taxes (as opposed to disposable income inequality, inequality of revenues after transfers and taxes).

markup The ratio of the price to the cost of production.

Marshall-Lerner condition The condition under which a real depreciation leads to an increase in net exports.

maturity The length of time over which a financial asset (typically a bond) promises to make payments to the holder.

medium run The period of time between the short run and the long run; roughly, what happens between three and thirty years.

menu cost The cost of changing a price.

mid-cycle deficit See *cyclically adjusted deficit*.

models of endogenous growth Models in which accumulation of physical and human capital can sustain growth even in the absence of technological progress.

modified Phillips curve The curve that plots the change in the inflation rate against the unemployment rate; also called *accelerationist Phillips curve*.

monetarism, monetarists A group of economists in the 1960s, led by Milton Friedman, who argued that monetary policy had powerful effects on activity.

monetary base Money issued by the central bank (currency and central bank reserves).

monetary contraction A change in monetary policy that leads to an increase in the interest rate; also called *monetary tightening*.

monetary expansion A change in monetary policy that leads to a decrease in the interest rate.

monetary-fiscal policy mix The combination of monetary and fiscal policies in effect at a given time.

monetary tightening See *monetary contraction*.

money Financial assets (currency and checkable deposits) that can be used directly to buy goods.

money finance Financing of the budget deficit through money creation.

money illusion The proposition that people make systematic mistakes in assessing nominal versus real changes in incomes and interest rates.

money market funds Non-bank financial institutions that receive funds from people and use them to buy short-term bonds.

mortgage-based security (MBS) A security based on an underlying portfolio of mortgages.

mortgage lenders The institutions that make housing loans to households.

multilateral exchange rate (multilateral real exchange rate) The real exchange rate between a country and its trading partners, computed as a weighted average of bilateral real exchange rates; also called the *trade-weighted real exchange rate* or *effective real exchange rate*.

multiplier The ratio of the change in an endogenous variable to the change in an exogenous variable (e.g., the ratio of the change in output to a change in autonomous spending).

Mundell-Fleming model A model of simultaneous equilibrium in both goods and financial markets for an open economy.

narrow banking Restrictions on banks that would require them to hold only short-term government bonds.

national accounts See *national income and product accounts*.

national income In the United States, the income that originates in the production of goods and services supplied by residents of the United States.

national income and product accounts (NIPA) The system of accounts used to describe the evolution of the sum, composition, and distribution of aggregate output.

natural rate of interest The rate of interest consistent with a level of demand for goods equal to potential output.

natural rate of unemployment The unemployment rate at which price and wage decisions are consistent.

neoclassical synthesis A consensus in macroeconomics, developed in the early 1950s, based on an integration of Keynes' ideas and the ideas of earlier economists.

net capital flows Capital flows from the rest of the world to the domestic economy, minus capital flows to the rest of the world from the domestic economy (also, financial account balance).

net debt Value of government financial liabilities, minus the value of financial assets held by the government.

net exports The difference between exports and imports; also called the *trade balance*.

net interest In the national income and product accounts, the interest paid by firms minus the interest received by firms, plus interest received from the rest of the world minus interest paid to the rest of the world.

net national product (NNP) Gross national product minus capital depreciation.

net transfers received In the current account, the net value of foreign aid received minus foreign aid given.

neutral rate of interest See *natural rate of interest*.

new classicals A group of economists who interpret fluctuations as the effects of shocks in competitive markets with fully flexible prices and wages.

new growth theory Developments in growth theory since the late 1980s that explore the determinants of technological progress and the role of increasing returns to scale in growth.

new Keynesians A group of economists who believe in the importance of nominal rigidities in fluctuations, and who are exploring the role of market imperfections in explaining fluctuations.

nominal exchange rate The price of domestic currency in terms of foreign currency. The number of units of foreign currency you can get for one unit of domestic currency.

nominal GDP The sum of the quantities of final goods produced in an economy times their current price; also known as *dollar GDP* and *GDP in current dollars*.

nominal interest rate The interest rate in terms of the national currency (in terms of dollars in the United States). It tells us how many dollars one has to repay in the future in exchange for borrowing one dollar today.

nominal rigidities The slow adjustment of nominal wages and prices to changes in economic activity.

nondurable goods Commodities that can be stored but have an average life of less than three years.

nonhuman wealth The financial and housing component of wealth.

noninstitutional civilian population The number of people potentially available for civilian employment.

nonresidential investment The purchase of new capital goods by firms: structures and producer durable equipment.

North American Free Trade Agreement (NAFTA) An agreement signed by the United States, Canada, and Mexico in which the three countries agreed to establish all of North America as a free trade zone.

not in the labor force The number of people who are neither employed nor looking for employment.

***n*-year interest rate** See *yield to maturity*.

official deficit The difference between public spending, including nominal interest payments, and public revenues.

Okun coefficient The effect of a change in the rate of growth of output on the change in the unemployment rate.

Okun's law The relation between GDP growth and the change in the unemployment rate.

open market operation The purchase or sale of government bonds by the central bank for the purpose of increasing or decreasing the money supply.

openness in factor markets The opportunity for people to choose where to live and work, and for firms to invest at home or abroad.

openness in financial markets The opportunity for financial investors to choose between domestic and foreign financial assets.

openness in goods markets The opportunity for consumers and firms to choose between domestic and foreign goods.

optimal control The control of a system (a machine, a rocket, an economy) by means of mathematical methods.

optimal control theory The set of mathematical methods used for optimal control.

optimal currency area The properties of a common currency area needed for it to function smoothly.

option value of waiting In the case of irreversible decisions, the option of waiting for uncertainty to be resolved so as to make a more informed decision.

ordinary least squares (OLS) A statistical method to find the best-fitting relation between two or more variables.

Organization for Economic Cooperation and Development (OECD) An international organization that collects and studies economic data for many countries. Most of the world's rich countries belong to the OECD.

Organization of Petroleum Exporting Countries (OPEC) A set of petroleum-producing countries, which long acted as a production cartel.

original Phillips curve A relation between the unemployment rate and the rate of inflation (as opposed to the accelerationist Phillips curve, a relation between the unemployment rate and the change in the rate of inflation).

out of the labor force People of working age not working (in the market economy) and not looking for a job.

output fluctuations Movements in output around its trend; also called *business cycles*.

output gap The difference between actual output and potential output.

output per person A country's gross domestic product divided by its population.

panel dataset A dataset that gives the values of one or more variables for many individuals or many firms over some period of time.

paradox of saving The result that an attempt by people to save more may lead both to a decline in output and to either unchanged or even lower saving. Also called the *paradox of thrift*.

paradox of thrift See *paradox of saving.*

parameter A coefficient in a behavioral equation.

participation rate The ratio of the labor force to the total population of working age.

patent The legal right granted to a person or firm to exclude anyone else from the production or use of a new product or technique for a certain period of time.

pay-as-you-go Social Security system A retirement system in which the contributions of current workers are used to pay benefits to current retirees.

PAYGO (pay-as-you-go) rule A budget rule requiring any new spending to be financed by additional revenues.

payments of factor income to the rest of the world In the United States, income received by foreign capital and foreign residents.

peg The exchange rate to which a country commits under a fixed exchange rate system.

permanent income theory of consumption The theory of consumption, developed by Milton Friedman, that emphasizes that people make consumption decisions based not on current income but on their notion of permanent income.

personal consumption expenditures Spending on consumption goods and services by households.

personal disposable income The income available to consumers after they have received transfers and paid taxes.

personal income The income actually received by persons.

Phillips curve The curve that plots the relation between movements in inflation and unemployment.

players The participants in a game. Depending on the context, players may be people, firms, governments, and so on.

policy coordination (of macroeconomic policies between two countries) The joint design of macroeconomic policies to improve the economic situation in two countries.

policy mix See *monetary-fiscal policy mix*.

policy rate The interest rate set by the central bank.

political business cycle Fluctuations in economic activity caused by the manipulation of the economy for electoral gain.

potential output The level of output associated with the unemployment rate being equal to the natural unemployment rate.

present discounted value See *expected present discounted value*.

present value See *expected present discounted value*.

price level The general level of prices in an economy.

price risk The risk that an asset price falls, leading to a loss for the holder.

price-setting relation The relation between the price chosen by firms, the nominal wage, and the markup.

primary deficit Government spending, excluding interest payments on the debt, minus government revenues (the negative of the primary surplus).

primary surplus Government revenues minus government spending, excluding interest payments on the debt.

private saving (S) Saving by consumers and by firms.

private sector involvement A reduction in the value of the debt held by the private sector in case of debt rescheduling or debt restructuring.

private spending See *aggregate private spending*.

production function The relation between the quantity of output and the quantities of inputs used in production.

profitability The expected present discounted value of profits.

propagation mechanism The dynamic effects of a shock on output and its components.

propensity to consume (c_1) The effect of an additional dollar of disposable income on consumption.

propensity to save The effect of an additional dollar of disposable income on saving (equal to one minus the propensity to consume).

property rights The legal rights given to property owners.

proprietors' income In the national income and product accounts, the income of sole proprietorships, partnerships, and tax-exempt cooperatives.

public saving Saving by the government; equal to government revenues minus government spending; also called the *budget surplus*. (A budget deficit represents public dissaving.)

purchasing power Income in terms of goods.

purchasing power parity (PPP) A method of adjustment used to allow for international comparisons of GDP.

QE1, QE2, QE3 The first, second, and third instances of unconventional monetary policy (quantitative easing) in the United States during the financial crisis.

quantitative easing Purchases of financial assets by the central bank at the zero lower bound, leading to an increase in the balance sheet of the central bank.

quits Workers who leave their jobs for what they perceive as a better alternative.

quotas Restrictions on the quantities of goods that can be imported.

R^2 A measure of fit, between zero and one, from a regression. An R^2 of zero implies that there is no apparent relation between the variables under consideration. An R^2 of one implies a perfect fit: all the residuals are equal to zero.

random walk The path of a variable whose changes over time are unpredictable.

random walk of consumption The proposition that, if consumers are foresighted, changes in their consumption should be unpredictable.

rate of growth of total factor productivity See *Solow residual*.

rating agencies Firms that assess the creditworthiness of various debt securities and debt issuers.

rational expectations The formation of expectations based on rational forecasts, rather than on simple extrapolations of the past.

rational speculative bubble An increase in asset prices based on the rational expectation of further increases in prices in the future.

real appreciation An increase in the relative price of domestic goods in terms of foreign goods; an increase in the real exchange rate.

real business cycle (RBC) models Economic models that assume that output is always at its natural level. Thus, all output fluctuations are movements of the natural level of output as opposed to movements away from the natural level of output.

real depreciation A decrease in the relative price of domestic goods in terms of foreign goods; a decrease in the real exchange rate.

real exchange rate The relative price of domestic goods in terms of foreign goods.

real GDP A measure of aggregate output. The sum of quantities produced in an economy times their price in a base year. Also known as *GDP in terms of goods, GDP in constant dollars*, or *GDP adjusted for inflation*. The current measure of real GDP in the United States is called *GDP in (chained) 2012 dollars*.

real GDP in chained (2012) dollars See *real GDP*.

real GDP per person Ratio of real GDP to population.

real interest rate The interest rate in terms of goods. It tells us how many goods one has to repay in the future in exchange for borrowing the equivalent one good today.

receipts of factor income from the rest of the world In the United States, income received from abroad by US capital or US residents.

recession A period of negative GDP growth. Usually refers to at least two consecutive quarters of negative GDP growth.

regression The output of ordinary least squares; gives the equation corresponding to the estimated relation between variables, together with information about the degree of fit and the relative importance of the different variables.

regression line The best-fitting line corresponding to the equation obtained by using ordinary least squares.

rental cost of capital See *user cost of capital*.

rental income of persons In the national income and product accounts, the income from the rental of real property, minus depreciation on this property.

research and development (R&D) Activity aimed at discovering and developing new ideas and products.

reservation wage The wage that would make a worker indifferent between working and being unemployed.

reserve ratio The ratio of bank reserves to checkable deposits.

reserves See *bank reserves*.

residential investment The purchase of new houses or apartments by people.

residual The difference between the actual value of a variable and the value implied by the regression line. Small residuals indicate a good fit.

revaluation An increase in the exchange rate (E) in a fixed exchange rate system.

Ricardian equivalence The proposition that neither government deficits nor government debt have an effect on economic activity; also called the *Ricardo-Barro proposition*.

Ricardo-Barro proposition See *Ricardian equivalence*.

risk aversion A person's preference to receive a given amount for sure instead of an uncertain amount with the same expected value.

risk premium The difference between the expected rate of return on a risky asset and the safe interest.

risk premium on bonds The additional interest rate a bond has to pay, reflecting the risk of default on the bond.

safe haven A country that is considered safe by financial investors.

saving The sum of private and public saving

saving rate The ratio of saving to GDP.

securitization The issuance of securities, based on an underlying portfolio of assets, such as mortgages, or commercial paper.

seignorage The revenues from the creation of money.

senior securities Securities repaid before junior securities in case of insolvency.

separations Workers who are leaving or losing their jobs.

services Products that cannot be stored and must be consumed at the place and time of purchase.

shadow banking system The set of nonbank financial institutions, from structured investment vehicles to hedge funds.

share A financial asset issued by a firm that promises to pay a sequence of payments, called *dividends*, in the future; also called *stock*.

shocks Movements in the factors that affect aggregate demand and/or aggregate supply.

shoe-leather costs The costs of going to the bank to take money out of a checking account.

short run A period of time extending over a few years at most.

short-term interest rate The interest rate on a short-term bond (typically a year or less).

skill-biased technological progress The proposition that new machines and new methods of production require skilled workers to a greater degree than in the past.

slope In a linear relation between two variables, the amount by which the first variable increases when the second increases by one unit.

Social Security trust fund The funds accumulated by the US Social Security system as a result of surpluses in the past.

Solow residual The excess of actual output growth over what can be accounted for by the growth in capital and labor.

spending caps Legislative limits on public spending.

spread The difference between the interest rate on a risky bond and the interest rate on a safe bond.

Stability and Growth Pact (SGP) A set of rules governing public spending, deficits, and debt in the European Union.

stagflation The combination of stagnation and inflation.

staggering of wage and price decisions The fact that different wages are adjusted at different times, making it impossible to achieve a synchronized decrease in nominal wage inflation.

standard of living Real GDP per person.

standardized employment deficit See *cyclically adjusted deficit*.

state of technology The degree of technological development in a country or industry.

static expectations The formation of expectations, on the assumption that the future will be like the past.

statistical discrepancy A difference between two numbers that should be equal, coming from differences in sources or methods of construction for the two numbers.

steady state In an economy without technological progress, the state of the economy where output and capital per worker are no longer changing; in an economy with technological progress, the state of the economy where output and capital per effective worker are no longer changing.

stock A variable that can be expressed as a quantity at a point in time (such as wealth); also a synonym for *share*.

stocks An alternative term for *inventories*; also, an alternative term for *shares*.

strategic interactions An environment in which the actions of one player depend on and affect the actions of another player.

structural change A change in the economic structure of the economy, typically associated with growth.

structural deficit See *cyclically adjusted deficit*.

structural rate of unemployment See *natural rate of unemployment*.

structured investment vehicle (SIV) Financial intermediaries set up by banks. SIVs borrow from investors, typically in the form of short-term debt, and invest in securities.

structures In the national income and product accounts: plants, factories, office buildings, and hotels.

subprime mortgages Mortgages with a higher risk of default by the borrower.

sudden stops A sudden decrease in the willingness of foreign investors to hold the debt of a particular country.

supply siders A group of economists in the 1980s who believed that tax cuts would increase activity by enough to increase tax revenues.

tariffs Taxes on imported goods.

tax smoothing The principle of keeping tax rates roughly constant, so that the government runs large deficits when government spending is exceptionally high and small surpluses the rest of the time.

Taylor rule A rule, suggested by John Taylor, telling a central bank how to adjust the nominal interest rate in response to deviations of inflation from its target and of the unemployment rate from the natural rate.

technological progress An improvement in the state of technology.

technological unemployment Unemployment brought about by technological progress.

technology frontier The state of technological knowledge.

term premium The difference between the interest rate on a long-term bond and the interest rate on a short-term bond.

term structure of interest rates See *yield curve*.

time inconsistency In game theory, the incentive for one player to deviate from a previously announced course of action once the other player has moved.

Tobin's *q* The ratio of the value of the capital stock, computed by adding the stock market value of firms and the debt of firms and dividing the sum by the replacement cost of capital.

total factor productivity (TFP) growth The rate of technological progress.

total wealth The sum of human wealth and nonhuman wealth.

toxic assets Nonperforming assets, from subprime mortgages to nonperforming loans.

tradable goods Goods that compete with foreign goods in domestic or foreign markets.

trade balance The difference between exports and imports; also called *net exports*.

trade deficit A negative trade balance, that is, imports exceed exports.

trade surplus A positive trade balance, that is, exports exceed imports.

trade-weighted exchange rate See *multilateral exchange rate.*

transfers to persons Unemployment, retirement, health, and other benefits paid by the state.

Treasury bill (T-bill) A US government bond with a maturity of up to one year.

Treasury bond A US government bond with a maturity of 10 years or more.

Treasury Inflation Protected Securities (TIPS) US government bonds that pay the real (rather than the nominal) interest rate.

Treasury note A US government bond with a maturity of one to 10 years.

Troubled Asset Relief Program (TARP) The program introduced in October 2008 by the US administration, aimed at buying toxic assets and, later, providing capital to banks and other financial institutions in trouble.

***t*-statistic** A statistic associated with an estimated coefficient in a regression that indicates how confident one can be that the true coefficient differs from zero.

twin deficits The budget and trade deficits that characterized the United States in the 1980s.

unconventional monetary policy Monetary policy measures used to increase economic activity when the policy rate reached the zero lower bound.

uncovered interest parity (UIP) relation An arbitrage relation stating that domestic and foreign bonds must have the same expected rate of return, expressed in terms of a common currency.

underground economy That part of a nation's economic activity that is not measured in official statistics, either because the activity is illegal or because people and firms are seeking to avoid paying taxes.

underwater A loan is underwater if its value is higher than the value of the collateral it corresponds to. For example, a mortgage is underwater if its value exceeds the price of the corresponding house.

unemployment Number of people not working but looking for a job.

unemployment insurance Unemployment benefits paid by the state to the unemployed.

unemployment rate The ratio of the number of people unemployed to the labor force.

usable observation An observation for which the values of all the variables under consideration are available for regression purposes.

user cost of capital The cost of using capital over a year, or a given period of time. The sum of the real interest rate and the depreciation rate; also called the rental cost of capital.

value added The value a firm adds in the production process, equal to the value of its production minus the value of the intermediate inputs it uses in production.

VIX Index of stock market volatility.

wage indexation A provision that automatically increases wages in response to an increase in prices.

wage-setting relation The relation between the wage chosen by wage setters, the price level, and the unemployment rate.

war of attrition When both parties to an argument hold their grounds, hoping that the other party will give in.

wealth See *financial wealth*.

wholesale funding Financing through the issuance of short-term debt rather than through deposits.

Wicksellian rate of interest See *neutral, or natural, rate of interest*.

yield The ratio of the coupon payment to the value of the bond.

yield curve The relation between yield and maturity for bonds of different maturities; also called the *term structure of interest rates*.

yield to maturity The constant interest rate that makes the price of an *n*-year bond today equal to the present value of future payments; also called the *n-year interest rate*.

zero lower bound The lowest interest rate the central bank can achieve before it becomes more attractive to hold cash than to hold bonds.

Index

A

AAA bond, 114–115, 125
Above the line transactions, 365
Acceptable pay, 276
Account/accounting
  balance in, 365, 390–391
  deficit in, 365, 388–389, 391, 428–429
  inflation in, 464
  national, 20, 509–510, A–1, A–4
  surplus, 365, 428
  transactions, 365
Accumulation of capital, 237
Acemoglu, Daron, 268, 517
Adjustment, dynamics of, 56
"Admiral Van de Eyck," 306
After-tax labor income, 315–316, 470
Aggregate output, 20–25, 34. *See also* gross domestic product (GDP); specific topics
Aggregate private spending, 337
Aggregate production function, 212–213
Aghion, Philippe, 517
Akerlof, George, 516
Alternative employment arrangements, 166
Alternative saving rate. *See also* saving rate
  capital and, 223–225
  output and, 223–230
Alternative saving rates, 223–230
Amazon, 143, 274
AMECO. *See* annual macroeconomic (AMECO) database
American Recovery and Reinvestment Act, 127
Anchored expectations, 160
Animal spirits, 508
Annual macroeconomic (AMECO) database, 18
Apple I, 251
Applied research and development (R&D), 251
Appreciation, 359, 401–402, 408, 434
  nominal, 359, 398–399
  real, 361, 399, 423
Appropriability, 251
Arbitrage, 295–296, 299, 301
Argentina, 434
Asset, 416
  domestic, 364, 367, 369
  foreign, 367, 416
  illiquidity of, 124
  interest-paying, 367, 369
  liquid, 518
  toxic, 124
Asset prices, 306–307, 336, 499
Automatic stabilizer, 472
Autonomous spending, 52

B

Baby Boom generation, 479
Backloading, 344
Balance
  account, 365, 366
  income, 365
  of payment, 365–367
  primary, A–5
  in trade, 47, 379–381, 383, 386, 395, 405, 408, 421, A–4
Balanced budget, 52, 455–457
Balanced growth, 248–249
Balance sheet of central bank, 74
Band, 103, 410
Bank/banking, 76–78, 118
Bank for International Settlements in Basel, 518
Bank runs, 118–119, 518
Bankruptcy, 57–58
Bargaining, collective, 141
Bargaining power, 142
Bargaining rules, 167
Barro, Robert, 470, 511
Basel II, 500
Basel III, 500
Base year, 40
Basis points, 475
BBB bond, 114–115, 125
BEA. *See* Bureau of Economic Analysis (BEA)
Behavioral equation, 48
Below the line transactions, 365
Benartzi, Shlomo, 236
Bernanke, Ben, 516
Bezos, Jeff, 143
Bilateral exchange rates, 363–364, 410
Bitcoins, 80
Black October, 306–307
BlackRock, 275
Blinder, Alan, 451
Bloom, Nick, 254
BLS. *See* Bureau of Labor Statistics (BLS)
Bond, 68–69, 292–299. *See also* specific types of
  AAA, 114–115, 125
  BBB, 114–115, 125
  Brazilian, 370
  corporate, 114–115
  coupon, 294
  discount, 294
  domestic, 399–402, 416, 425
  face value of, 294
  foreign, 399–402, 416–417
  government, 69, 114–115, 293–294, 497
  indexed, 493
  interest rate on, 68
  junk, 294
  maturity of, 114, 292–293
  nominal, 294
  open market for, 74
  present value of, 294–295
  price of, 74–75, 293–296, 299
  risk of, 293, 297–298
  Treasury, 294
  two-year, 294–298
  yield curve in, 298–299
Bond market, 294
Bond ratings, 294
Bond yields, 74–75, 293, 296–297
Booms, 487
Borrowing rate, 120
Bracket creep, 492
Brazilian bond, 370
"The Bretton Woods period," 419
Britain, 423
Brynjolfsson, Eric, 266
Bubbles, 306–308
Budget. *See also* specific topics
  balanced, 52, 455–457
  deficit in, 59, 228, 342, 345–346, 391, 407, 433, 452, 454–455, 463–464, 473
  fiscal policy on constraints in, 463–469
  government, 457, 463, 465
  surplus in, 59, 228, 457
Budget Enforcement Act, 455, 457
Bundesbank, 412
Burden of war, 472–473
Bureau of Economic Analysis (BEA), 41
Bureau of Labor Statistics (BLS), 29
Bush, George, 98
Business cycle, 193, 452, 508
Business inventories, change in, A–4
Business transfers, A–2

C

Capital. *See also* specific types of
  alternative saving rates, dynamics of, 223–225
  decreasing returns to, 213–214
  endogenous growth and, 238
  fixed, A–2
  flows in, 366, 400
  golden-rule level of, 230, 239
  human, 236, 237–238
  output and, 220–223, 237–238
  per effective worker, 245–250
  per worker, 213–214, 221, 230–232, 245–246, 249
  physical, 236, 237–238, 241
  production function of, extending, 236–237
  returns to, 213–214
  technological progress and, 246–249
  transfers in, 366
Capital accumulation, 215, 231
  in France, 226
  human, 237
  investment and, 222–223
  output, effects on, 221
  physical, 237
  in United States, 231–232
Capital controls, 355, 500
Capital income, 21–22
Capital in the 21st Century (Piketty), 274
Capital mobility, 416–417
Capital ratio, 116
Capital stock, 247, 329
Case-Shiller index, 122
Cash flow, 326, 464
Causality, A–14–A–16
Cavallo, Domingo, 434
CDO's. *See* collateralized debt obligations (CDOs)
Central bank, 74, 449–450
Central bank money, 77–78, 416–417, 476. *See also* money
Central parity, 410, 427
Chained (2012) dollars, 24, 41
Chain-type indexes, 40–41
Change in business inventories, A–4
Change in real debt, 464
Checkable deposits, 68, 77
Chicago University, 511
China, 13–15, 258, 262–263
Churchill, Winston, 423
The Churn, 269–270
Civilian unemployment rate, 17
Climate change, 278–280
Clinton, Bill, 101, 451
Closed economy, 355, 403
Cobb, Charles, 241
Cobb-Douglas production function, 241
Coca Cola, 253
Cold War, 457
Collateral, 126
Collateralized debt obligations (CDOs), 124
Collective bargaining, 141
Columbia University, 406, 511, 514
Commitment, 514
Common currency, 431–432, 434. *See also* euro
Common currency area, 9, 431. *See also* euro area
  in exchange rate, 431–433
Common price, 22
Compensation of employees, A–2

Composition of consumption, 376
Compounding, force of, 207
Computer prices, 25
Concord Coalition, 232
Confidence band, 103
Confucius, 258
Conservative central banker, 450–451
Consols, 291
Constant dollars, 24
Constant interest rates, 290–291
Constant returns to scale, 213
Constitution, 443
Consumer, consumption by, 314–315
Consumer Price Index (CPI), 29–31, 192, 299
Consumer spending, 304
Consumption, 46. *See also* investment
  behavior of, 229
  composition of, 376
  by consumer, 314–315
  decision for, 336
  demand for goods, impact on, 48–50
  expectations and, 314–320
  expenditures for personal, A–3
  of fixed capital, A–2
  function of, 48, 57–58
  in goods market, 376–377
  investment and, 329, 473
  level of, 376
  life-cycle theory of, 314
  macroeconomics of, 509
  marginal propensity to, 49
  per person, 204–205
  per worker, 229, 235
  propensity to, 49
  random walk of, 513
  saving rate and, 228–230
  theory of, 314, 328, 509
  unpredictable, 513
  zero, 229
Contract with America, 443–444
Convergence, 208
Core inflation rate, 32
Corporate bond, 114–115
Corporate profits, A–2
Corporate tax, 407
Correlation, A–14–A–16
Corridor system, 497–498
Cost
  implicit, 324
  of inflation, 491–494
  of living, 29
  menu, 516
  rental, 324
  shadow, 324
  shoe-leather, 491
  user, 324
Coupon bonds, 294
Coupon payments, 294
CPI. *See* Consumer Price Index (CPI)
CPS. *See* Current Population Survey (CPS)
Crawling pegs, 409–410
Creative destruction, 269, 517
Credibility, 344, 449–451, 514
Credit easing, 496
Cruzeiro, 370
Cultural Revolution, 258
Culture of entrepreneurship, 251
Currency, 68. *See also* euro; money
  common, 431–432, 434
  foreign, 364, 416, 421
  in households, average, 71
  optimal area of, 431
Currency board, 434–435
Current account, 365, 428–429
Current dollars, 24
Current domestic interest rate, 427
Current interest rates, 429
Current output, 327
Current period, 337
Current Population Survey (CPS), 26, 137–138, 271, 315
Current profit, 324–326
Current real interest rate, 339
Current yield, 294
Cyclically adjusted deficit, 471–472

D
“Dance of the dollar,” 363, 429
Dartmouth University, 318
Davis, Steve, 270
De-anchored expectations, 158–161
Deaton, Angus, 207
Debt
  change in real, 464
  fiscal policy on, 463–465, 474–479
  gross, 479
  net, 479
  public, 519
  real, 464
  rescheduling, 476
  restructuring, 476
  stabilization of, 467
Debt default, 476
Debt finance, 299
Debt monetization, 476
Debt ratio, 468–469, 480–481
Debt-to-GDP ratio (debt ratio), 468–470, 480–481
Decreasing returns to capital, 213–214
Decreasing returns to labor, 213
Default risk, 293, 474–476
Deficit
  account, 365, 388–389, 391, 428–429
  cyclically adjusted, 471–472
  excessive, 456–457
  federal, 479
  in financial account, 366
  fiscal policy on, 407–408, 463–465, 470–474
  full-employment, 471
  government, 464–465, 473, 476, 479–480
  inflation-adjusted, 463, A–5
  measurements of, 464
  midcycle, 471
  official, A–5
  primary, 465
  reduction in, 101, 342–349
  standardized employment, 471
  structural, 471
  in trade, 47, 365, 381–382, 384–385, 388, 407–409, 422, 434
  twin, 407
Deficit finance, 472–473
Deficit spending, 472
Deflation, 27
  IS-LM model for spirals in, 184–185
  nominal interest rate, relation to, 113–114
  Phillips curve and, 169–170
Deflation spiral, 184–185
Deflation trap, 184–186
Degrees of freedom, A–14
Delors, Jacques, 433
Demand. *See also* specific types of
  effective, 508
  foreign, 383
  for goods, 47–50
  for labor, 154
  money, 68–71, 93, 488–489
Demand deposits, 118, 499
Demand for goods, 47–50
  consumption, impact on, 48–50
  domestic, 376, 379, 382–383
  in goods market, 47–50
  government spending, impact on, 50
  investment, impact on, 50
Democrats, 454
Department of Commerce, 25
Dependent variable, A–14
Deposit, 68, 77, 118, 499
Depreciation, 359, 420, 426, 429, 432
  direct effects of, 386
  of the dollar, 385, 429
  in goods market, 386–387
  of investment, 321
  nominal, 359, 421
  real, 361, 385–387, 421, 433
Depression, 186
Deutsche Mark (DM), 412, 426
Devaluation, 359, 410, 423–426, 434, 475
Diamond, Doug, 518
Diamond, Peter, 493, 516
Direct finance, 115
Direct investment, foreign, 500
Discount bonds, 294
Discounted value, 288–289
Discount factor, 288
Discount rate, 288
Discouraged workers, 26, 138
Disposable income, 48, A–3
Disposable income inequality, 275, 276–277
Distortions in tax, 473–474, 492
Dividends, 299
Divine coincidence, 489
DM. *See* Deutsche Mark (DM)
Dollar, 80. *See also* currency; money
  chained, 24, 41
  constant, 24
  current, 24
  depreciation of the, 385, 429
  GDP, 23–24
  international prices on, 204
  value of a, 289
“Dollar cycle,” 363
Dollar GDP, 24. *See also* nominal gross domestic product (GDP)
Dollarization, 434–435
Domestic asset, 364, 367–369
Domestic bond, 399–402, 416, 425
Domestic demand, 376–377, 380–383
Domestic fixed investment, A–3
Domestic goods
  demand for, 376, 379, 382–383
  in goods market, 376
  in open economy, 358
Domestic inflation, 422
Domestic interest-paying assets, 367, 369
Domestic interest rate, 416–417, 422, 424–425, 427–428
Domestic money, 367, 417
Domestic nominal interest rate, 424
Domestic real interest rates, 439–440
Dornbusch, Rudiger, 514
Douglas, Paul, 241
Dow Jones Industrial Average, 305, 307
Draghi, Mario, 476
DSGE. *See* dynamic stochastic general equilibrium (DSGE) model
Durable goods, A–3
Dybvig, Philip, 518
Dynamics of adjustment, 56
Dynamic stochastic general equilibrium (DSGE) model, 518

E
Earnings losses, 270
Easing credit, 496
Easterlin paradox, 206–207
ECB. *See* European Central Bank (ECB)
Econometrics, 55, A–12, A12–A16
The Economic Consequences of Mr. Churchill (Keynes), 423
Economic fluctuations, 193
Economic growth in China, 258
Economic Recovery Tax Act, 407
Economic Report of the President, 17
Economics (Samuelson), 508
Economics/economy. *See also* open economy
  closed, 355, 403
  inflation rate, impact on, 30–31
  new Keynesian, 515–516
  “Robinson Crusoe,” 59
  stock market, 302–303
  unemployment, impact on, 26–27
The Economist, 18
Effective demand, 508
Effective labor, 244–245, 248
Effective worker, 245–250
Efficiency wages, 143, 516
Emotional well-being, 207

Employees, compensation of, A–2. *See also* worker
Employment, 25–26
alternative arrangements for, 166
protection during, 145
rate of, 138
standardized deficit in, 471
EMS. *See* European Monetary System (EMS)
Endogenous growth, 238
Endogenous variables, 50
Entitlement programs, 526
Entrepreneurship, culture of, 251
Equilibrium, 51, 58–60, 91–92, 300. *See also* equilibrium output
in central bank money, 78–79
dynamic stochastic general, 518
exchange rate, 404
in financial markets, 178–179, 399–402, 404
in goods market, 51–60, 178, 358, 379, 398–399, 404
inflation rate, 71–73
interest rate, 71–72
investment, impact on, 58–59
in labor markets, 178–179
medium-run, 182–184, 421–422
output, 51–58, 379–380, 386, 403
saving, impact on, 58–59
short-run, 181–182, 186–187, 190, 421–422
unemployment rate, 148
wages and, 148–149
Equilibrium condition, 51, 71–72, 78–79, 398
Equilibrium output, 51–58
adjusting, 56, 58
algebra for computing, 52
econometrics for describing, 55
in goods market, 379–380
graph for determining, 53–55
Equipment, A–3
Equity finance, 299
Equity premium, 300
EU. *See* European Union (EU)
Euro, 409–410, 427, 433
in euro area, 12–13
Euro area, 9–13, 456–457
Euro Crisis, 6
European Central Bank (ECB), 347, 433, 476
European Commission, 431, 457
European Monetary System (EMS), 410–412, 426–427, 433
European unemployment, 167–168
European Union (EU), 9
Ex-ante real interest rate, 113
Excessive Deficit Procedure, 457
Excessive deficits, 456–457
Exchange rate, 203–204, 428, 514. *See also* exchange rate regime; fixed exchange rates
bilateral, 363, 410
common currency areas in, 431–433
current account and, 428–429
current interest rates and, 429
equilibrium, 404
expected future, 401–402
floating, 424
future, 429–430
in goods market, 387–388
inflation rate, relation between, 427–430
movements in, 427–430
multilateral, 363
nominal, 358, 359, 362, 388, 398, 408, 410, 421–422
in open economy, 369–371
overshooting, 514
real, 358–363, 376–377, 385–386, 399, 408, 420–424, 439–440
US, 363
volatility of, 429–430
Exchange rate crises, 424–427
Exchange rate regime, 419–440. *See also* specific types of
choosing between, 430–435
under fixed exchange rates, 424–427
under flexible exchange rates, 427–430
in medium run, 420–423
Exchange reserve, 416
Ex-dividend price, 300
Exogenous variables, 50, A–16
Expectations, 287–312. *See also* asset; bond; expected present discounted values; rational expectations; stock; specific topics; specific types of
about the future, role of, 343
anchored, 160
consumption and, 314–320
de-anchored, 160
decisions and, 336–339
deficit reduction and, 342–349
in financial market, 287–312
investment and, 320–328
IS relation and, 336–339
policy and, 339–342, 447–451
of profit, 320
rational, 341
re-anchored, 161–162
static, 322, 333
Expectations hypothesis, 295
Expected future exchange rate, 401–402
Expected future interest rate, 344, 428–429
Expected future output, 327, 344
Expected future profits, 324
Expected inflation, 156–158, 175
Expected present discounted values, 288–292
computing, 288–289
constant interest rates and, 290–291
general formula for, 289
nominal interest rates and, 291–292, 312
payments and, 290–291
real interest rates and, 291–292, 312
under static expectations, 333
zero interest rates and, 291
Expected profit, 324–326
Expected rate of return, 368, 399
Expenditures on personal consumption, A–3
Exports, 47, A–4
GDP, exceeding, 358
in goods market, 356–358, 377
net, 379, 386, 398, 406, A–4
in open economy, 356–358
Ex-post real interest rate, 113
External finance, 299

F

Facebook, 274
Face value of bond, 294
Fads, 307
Fed. *See* Federal Reserve Bank
Federal deficit, 479
Federal deposit insurance, 118
Federal funds market, 79–80
Federal funds rate, 8, 79–80, 497–498
Federal Reserve Bank, 4, 67, 349
Federal Reserve Bank of Dallas, 269
Federal Reserve Board, 71
Federal Reserve Economic Database (FRED), 17, 18
Fertility of research, 251
Final goods, 21
Finance. *See also* specific types of
debt, 299
deficit, 472–473
direct, 115
equity, 299
external, 299
internal, 299
money, 476–479
Financial account, 366–367
Financial crisis, 122–128
Financial intermediaries, 76, 115–119
financial market, 115–119
housing problems, role of, 123–125
illiquidity of assets by, 124
leverage of, 116–119, 123–124
liquidity of liabilities by, 125
securitization by, 124
wholesale funding by, 125
Financial investment, 69
Financial market, 67–86. *See also* goods market; investment-saving, liquidity-money (IS-LM) model; specific topics
equilibrium in, 178–179, 399–402, 404
expectations in, 287–312
financial crisis, impact on, 122–128
financial intermediaries in, role of, 115–119
goods market and, combining, 402–403
housing problems, impact on, 122–128
interest rates in, 71–80, 92–93
IS-LM model in, 92–96, 109–131, 178–181
liquidity trap in, 81–83
money demand in, 68–71
open economy in, 364–371
real income in, 92–93
real money in, 92–93
risk in, 114–115
Financial policy, 126. *See also* fiscal policy; monetary policy
Financial shock, 121–122
Financial stability, 498–502
Financial wealth, 69, 314
Fine tuning unemployment, 447
Fire sale prices, 117–119
Fiscal austerity, 102, 347
Fiscal Compact, 457
Fiscal consolidation, 94, 185–188
Fiscal contraction, 94
Fiscal deficit, 407–408
Fiscal dominance of fiscal policy, 476
Fiscal expansion, 94
Fiscal multipliers, 347
Fiscal policy, 50, 407, 446–447, 462–463
on budget constraints, 463–469
on current *vs.* future, 465
on cyclically adjusted deficit, 471–472
on debt, 463–465, 467, 469–470, 474–479
on default risk, 474–476
on deficits, 463–465, 470–474
fiscal dominance of, 476
fixed exchange rates and, 411
on full repayment, 465–467
in goods market, 383–385, 387–388
hyperinflation and, 479
on IS-LM model, 94–96, 127–128
on macroeconomics, 510
monetary policy *vs.*, 510
on money finance, 476–479
in open economy, effects of, 405–406
on output stabilization, 471–472
on Ricardian equivalence, 470–471
today's challenges facing, 479–481
on vicious cycles, 474–476
on wars, 472–474
Fiscal rules on euro area, 456–457
Fiscal stimulus, 384
Fischer, Stanley, 514
Fixed capital, consumption of, A–2
Fixed exchange rates, 359, 401, 409–412, 420, 422, 424, 430–431, 435. *See also* exchange rate
capital mobility and, 416–417
exchange rate crises under, 424–427
exchange rate regime under, 424–427
fiscal policy and, 411
interest rates and, 416–417
IS relation with, 421, 439

monetary policy and, 410–411
in open economy, 409–412
types of, 409–410
Fixed investment, 46
Fixed-money-growth rule, 451
Fleming, Marcus, 397, 406
Flexible exchange rates, 409–411, 420, 423, 425, 430, 435
exchange rate movements under, 427–430
exchange rate regime under, 427–430
Flexible inflation targeting, 489
Floating exchange rate, 424
Ford, Henry, 143
Ford Company, 143
Foreign asset, 367–369
Foreign bond, 399–402, 416–417
Foreign currency, 364, 416, 421
Foreign demand, 382–383
Foreign direct investment, 500
Foreign exchange, 364, 416–417
Foreign goods, 358
Foreign inflation, 422
Foreign interest-paying assets, 367, 416
Foreign interest rate, 416–417, 422, 424, 428
nominal, 424
real, 428, 439–440
Foreign money, 367
Four tigers, 211
Franc, 500
France, 226
Franc parity, 412
Frank, Barney, 384
FRED. *See* Federal Reserve Economic Database (FRED)
Friedman, Milton, 163, 314, 319, 447, 509–511
Full-employment deficit, 471
Full repayment, 465–467
Fully funded social security system, 231
Fundamental value, 306–307
Funds/funding
money market, 68–69
Social Security trust, 231, 479
wholesale, 125
Future
exchange rate in, 401–402, 430
expectations about the, 343
interest rates in, 343–344, 427–430
output in, 327, 343, 344
profits in, expected, 324
of technological progress, 266–267
Future period, 337, 343. *See also* future

G

Gali, Jordi, 517
Gallup World Poll, 206–207
Game theory, 447, 513–514
Garber, Peter, 306
GDP. *See* gross domestic product (GDP)
GDP deflator, 28–30
GDP growth, 24
GDP in dollars, 23–24. *See also* nominal gross domestic product (GDP)
General-purpose technologies, 266
General Theory of Employment, Interest, and Money (Keynes), 58, 87, 341, 508–509
Geometric series, 55, A–7
German reunification, 412
Germany, 412
Gertler, Mark, 516
Gini coefficient, 275–277
Globalization, 357
Global temperature, 278–279
Global warming, 278–280
GNP. *See* gross national product (GNP)
Golden rule, 230, 235–236, 239
Goldin, Claudia, 273
Gold standard, 423
Goods. *See also* specific types of
demand for, 47–50, 376, 383
durable, A–3
final, 21
GDP in terms of, 24
intermediate, 20–21
nondurable, A–3
for trade, 357
Goods market, 45–65. *See also* financial market*;* investment-saving, liquidity-money (IS-LM) model; specific topics
account balance in, 390–391
consumption in, 376–377
demand for goods in, 47–50, 376
depreciation in, 386–387
domestic demand in, 380–382
equilibrium in, 51–60, 178, 358, 379, 398–399, 404
equilibrium output in, 379–380
exchange rate in, 387–388
exports in, 356–358, 377
financial market and, combining, 402–403
fiscal policy in, 383–385, 387–388
foreign demand in, 382–383
GDP in, composition of, 46–47
government, impact on omnipotent, 60–62
growth in, 376–377
imports in, 356–358, 377
interest rates in, 88–89
investment in, 88–89, 376–377, 390–391
IS curve in, 91–92
IS-LM model in, 88–92, 94–96, 178–181
in open economy, 356–364, 375–395
output in, determining, 89–91
sales in, 88–89
saving in, 390–391
trade balance in, 379–380, 386
Google, 274
Gordon, Robert, 266
Government
budgets for, 457, 463, 465
deficits in, 464–465, 473, 476, 479–480
omnipotent, 60–62
purchases by, A–3–A–4
spending by, 47, 50, 405, 473
transfers within, 47
Government bond, 69, 114–115, 293–294, 497
Great Depression, 3, 5–6, 57–58, 119, 125, 169–170, 185, 186, 201, 267, 306, 348, 452, 473, 508, 518, 519
depression in, 186
macroeconomics during, 508
Great Financial Crisis, 3, 67, 81, 109, 307, 388
Great Leap Forward, 258
Great Moderation, 485
Great Recession, 186, 319, 348, 400, 518, 519
Greece, 388–389
Greenhouse effect, 278
Greenhouse gas, 278
Greenspan, Alan, 98
Griliches, Zvi, 252
Gross debt, 479
Gross domestic product (GDP), 17
aggregate output, for measuring, 20–22
defined, 20–22, A–1
dollar, 24, 41
exports, exceeding, 358
goods, in terms of, 24
in goods market, composition of, 46–47
gross national product *vs.*, 368
income side of, 21–22, A–1–A–3
inflation, adjusted for, 24
level of, 24
nominal, 22–24, 29
production side of, 20–22, A–3–A–4
real, 22–25, 29, 40–41
technological progress associated with, 25
Gross national product (GNP), 366–368, A–1–A2
Growth, 201. *See also* output; specific types of
balanced, 248–249
endogenous, 238
in goods market, 376–377
inclusive, 276
nominal money, 488–489
productivity, 267–268
rate of, 3, 7, 215, 227, 234, 446–447, 511
technological progress and, 244–250, 255–258
theory of, 509, 516–517
G20, 384

H

Haircut, 476
Hall, Robert, 513
Hansen, Alvin, 87, 509
Happiness, 28, 206–207
Hard peg, 434–435
Harmonized unemployment rate, 17
Harvard University, 252, 273, 318, 326, 470, 509, 511
Health and Retirement Study, 318
Hedonic pricing, 25
Heston, Alan, 204
Hicks, John, 87, 509
High debt, fiscal policy on, 474–479
High inflation, 166–169
High productivity growth, 267–268
Hires, 136–137
Holland, 306
Holmstrom, Bengt, 518
Hostage takings, 448
Household/housing. *See also* housing problems
currency in, average, 71
prices for, 122–123, 308–309, 499
saving rate of, 236
wealth in, 314
Housing problems, 122–128. *See also* household/housing
financial intermediaries, role of, 123–125
financial market, impact on, 122–128
housing prices, relationship with, 122–123
macroeconomic implications of, 125
policy responses to, 125–128
subprime mortgages, relationship with, 122–123
Howitt, Peter, 517
Human capital, 236, 237–238
Human wealth, 314, 470
Hybrid corn, 252
Hyperinflation, 454, 477–479, 491
Hysteresis, 166

I

IBM, 267
Identity/identification, 48, A–16
Illiquidity of assets, 124
Illusion of money, 492–493, 494
IMF. *See* International Monetary Fund (IMF)
Imitation in research, 254
Implicit cost, 324
Import, 47, A–4
adding, 379
compression of, 389
domestic demand of, 377
in goods market, 356–358, 377
net, 47
in open economy, 356–358
subtracting, 378–379
Inclusive growth, 276
Income, 69. *See also* specific types of
aggregate output, for measuring, 21–22
capital, 21–22
disposable, 48, 275, 276–277
GDP, side of, 21–22, A–1–A–3
inequality in, 275, 276–277
labor, 21–22, 315–316, 470
national, 525–526, A–2
net, 366, 368, 391

nominal, 69–70, 93
permanent, 314
personal, A–3
profit of, 21
proprietors', A–2
real, 92–93
rental, A–2
semantic traps associated with, 69
taxes on, 407, 526
Income balance, 365
Independent central bank, 449–450
Independent variable, A–14
Indexed bond, 493
Index number, 28–29
Indirect taxes, A–2
Industrial R&D, 250–251
Industrial Revolution, 267, 278
Inequality, 275, 276–277. *See also* wage inequality
market income, 275, 276–277
output, impact on, 269–278
Inflation, 27, 421. *See also* inflation rate
accounting for, 464
benefits of, 494–495
costs of, 491–494
domestic, 422
expected, 156–158, 175
foreign, 422
GDP, adjusted for, 24
in medium run, 156–158
policy on, 448–449
pure, 30
targeting, 485, 489–490
unemployment, impact on, 156–158, 175
variability of, 493–494
zero, 448, 448–449
Inflation-adjusted deficit, 463, A–5
Inflation rate, 7, 27–31, 410, 478. *See also* inflation
banks, impact on, 76–79
benefits of, 494–495
core, 32
costs of, 491–494
CPI, for measuring, 29–30
economy, impact on, 30–31
equilibrium, 71–73
exchange rate, relation between, 427–430
in federal funds market, 79–80
GDP deflator, impact on, 27–29
monetary policy on, 73–75
money and, 71–73, 75–76
open market operations, impact on, 73–75
optimal, 491–496
Innovation in research, 254
Insolvency, 116–117
Instrument, A–16
Instrumental variable methods, A–16
Insurance, 118, 145, 167, 499
Intercept, A–10
Interest, 183, 463, A–2. *See also* interest rate; specific types of
Interest parity condition, 368–369, 399–400, 410, 421, 424, 427
Interest-paying assets, 367, 369
Interest rate, 71–80. *See also* real interest rate; specific types of
on bond, 68
constant, 290–291
current, 339, 427, 429
domestic, 416–417, 422, 424–425, 428, 439–440
equilibrium, 71–72
expected future, 344, 428–429
in financial market, 71–80, 92–93
fixed exchange rates and, 416–417
foreign, 416–417, 422, 424, 428
future, 343, 429–430
in Germany, 412
in goods market, 88–89
long-term, 293, 445
nominal, 110–114
n-year, 296
in open economy, 369–371
rules for, 485, 490–491
short-term, 293
term structure of, 293
zero, 291
Intermediate good, 20–21
Internal finance, 299
International dollar prices, 204
International Monetary Fund (IMF), 18, 384, 406, 445
Inventory investment, 47
Investment, 46, 69, 327. *See also* specific types of
capital accumulation and, 222–223
consumption and, 329
decision to, 322, 336
deficit reduction, impact on, 101
demand for goods, impact on, 50
depreciation of, 321
equilibrium, impact on, 58–59
expectations and, 320–328
financial, 69
fixed, 46
foreign direct, 500
in goods market, 88–89, 376–377, 390–391
inventory, 47
macroeconomics of, 509
nonresidential, 46–47, 98, A–3
output and, 221–222
private, 473, A–3
profit from, 320–322, 324–328
residential, 46, A–3
stock market and, 323–324
theory of, 320
volatility of, 328–329
worker, per effective, 246–248
Investment equals Saving (IS relation), 59, 119, 127, 421
expectations and, 336–339
with fixed exchange rates, 421, 439
Investment-saving, liquidity-money (IS-LM) model, 87–107, 119–121, 303–304, 399, 404–406, 509. *See also* financial market; goods market
complications associated with, 183–185
for deflation, spirals in, 184–185
dynamics of, 102–104
in financial market, 92–96, 109–131, 178–181
financial policies and, 121–122, 126 (*See also* fiscal policy; monetary policy)
financial shocks associated with, 121–122
fiscal consolidation and, 185–188
in goods market, 88–92, 94–96, 178–181
macroeconomics in, 509
for medium run, 177–197
for monetary-fiscal policy mix, 96–102, 104
oil prices and, 188–191
risk premium in, 120
for zero lower bound, 184–185
Investment-saving, liquidity-money, Phillips curve (IS-LM-PC) model, 177, 190, 445
Ireland, 345–346
IS curve, 91–92, 94, 338–340, 343–344
IS-LM. *See* investment-saving, liquidity-money (IS-LM) model
IS-LM-PC. *See* investment-saving, liquidity-money, Phillips curve (IS-LM-PC) model
IS relation. *See* Investment equals Saving (IS relation)
It's a Wonderful Life, 118

J

Job destruction, 270
Jorgenson, Dale, 509
Junior securities, 124
Junk bonds, 294

K

Kahneman, Daniel, 207
Katz, Larry, 273
Keyesian Revolution, 512
Keynes, John Maynard, 58, 87, 423, 508–509
Keynesian model, 510–513, 519
new, 515–517
Khitatrakun, Surachai, 318
Klein, Lawrence, 509
Korean War, 256
Kravis, Irving, 204
Kuwait, 368
Kydland, Finn, 513

L

Labor, 249. *See also* specific types of
contracts for, 168–169
decreasing returns to, 213
demand for, 154
effective, 244–245, 248
income resulting from, 21–22, 315–316, 470
productivity of, 145–146
returns to, 213
Labor Department, 138
Labor force, 25–26, 136, 138
Labor market, 135–154
equilibrium in, 178–179
flows of workers in, 136–138
medium run in, 135–154
price determination in, 145–146
rigidities in, 12, 165–166
unemployment in, 139–141, 146–149
wage determination in, 141–145
Labor supply, 154
Lamont, Owen, 326
Layoff, 137, 270
Lehman bankruptcy, 57–58
Lehman Brothers, 4–5, 57, 125, 348
Lender of last resort, 499
Lending/lenders, 117–119, 123
Leverage, 116–119, 123–124
Liability, 125, 518
Life-cycle theory of consumption, 314
Life satisfaction, 207
"The Limits of Arbitrage," 518
Linear relation, 49, A–9–A–10
Liquid assets, 518
Liquidity, 117–119
facilities, 126
of liabilities, 125
monetary policy on, 499
preference of, 508
provision to, 118
Liquidity trap, 81–83, 485, 497–498
Liquid liabilities, 518
Living, cost of, 29
Loan, 77
Loan-to-value (LTV) ratio, 500–501
Logarithmic scale, 201, A–10–A–11
London School of Economics, 254
Long run, 34, 199–284. *See also* output
Long-term interest rates, 293, 445
Loss in earnings, 270
Low productivity growth, 268
LTV. *See* loan-to-value (LTV) ratio
Lucas, Robert, 238, 341, 511–513, 515, 516
Lucas critique, 512
Luddites, 267

M

Maastricht treaty, 433, 456
Macroeconomics. *See also* specific topics
of consumption, 509
following 2009, 518–519
during Great Depression, 508
growth theory and, 509
housing problems, implications of, 125

of investment, 509
IS-LM model in, 509
models of, 509–510
of money demand, 509
neoclassical synthesis and, 508–511
Phillips curve and, 510–511
policy in, role of, 509–511
rational expectations in, 511–514
up to 2009, 514–518
Macroeconomics crisis, 4–6
Macroeconomists, 18
Macroprudential tools, 499–502, 519
Maddison, Angus, 18
Malthus, Thomas Robert, 209
Malthusian era, 209
Malthusian trap, 209
Mao,Chairman, 258
Marginal propensity to consume, 49
Market, 275, 276–277, 294. *See also* financial market; goods market
Markup, 146
Marshall-Lerner condition, 386, 395
Mary Poppins, 118
Massachusetts Institute of Technology (MIT), 212, 266, 268, 314, 318, 447, 493, 509, 513, 514
Maturity, 114, 292–294, 296–299
Mawrodi, Sergei, 306
MBS. *See* mortgage-backed security (MBS)
McDonald's, 141
McTeer, Robert, 269
Medicaid, 479
Medicare, 479
Medium run, 34, 133–197. *See also* specific topics
aggregate output, 34
devaluation in, 422–423
equilibrium in, 182–184, 421–422
exchange rate regime in, 420–423
inflation in, 156–158
IS-LM model for, 177–197
in labor market, 135–154
Phillips curve related to, 158–164
short run *vs.*, 193
unemployment in, 156–158
Menem, Carlos, 434
Menu cost, 516
Midcycle deficit, 471
Minimum wage, 167
MIT. *See* Massachusetts Institute of Technology (MIT)
MIT-Penn-SSRC. *See* MPS (MIT-Penn-SSRC) model
MMM, 306
Model-T, 143
Modigliani, Franco, 314, 319, 447, 509–510
M1, 487–488
Monetarists, 510
Monetary contraction, 96
Monetary expansion, 96, 303–304
Monetary-fiscal policy mix, 96–97, 100–102, 104. *See also* fiscal policy; monetary policy
A Monetary History of the United States, 1867-1960 (Friedman), 510
Monetary policy, 120, 341, 346–347, 404, 407, 411, 429, 431, 433, 443, 446–447
conventional, 496
expectations and, 339–342
financial stability and, 498–502
fiscal policy *vs.*, 510
fixed exchange rates and, 410–411
fixed-money-growth rule in, 451
goals of, 487
on inflation, 73–75, 489, 491–496
on interest rate rule, 490–491
on IS-LM model, 96, 127
on lender of last resort, 499
on liquidity, 497–499
on macroeconomics, 510
on macroprudential tools, 499–502
on money targeting, 487–489
in open economy, 404
Taylor rule and, 490
unconventional, 496, 496–498
zero lower bound and, 481, 485
Monetary tightening, 96
Monetization of debt, 476
Money, 68–69. *See also* currency; specific topics
central bank, 77–78, 416–417, 476
demand for, 68–73, 93, 488–489, 509
domestic, 367, 417
financing for, 476–479
foreign, 367
growth of, 488–489
illusions of, 492–493, 494
inflation rate, choosing between, 75–76
neutrality of, 183
real, 92–93
semantic traps associated with, 69
supply of, 71–73, 93, 488
targeting with, 487–489
Money market funds, 68–69
Monitoring the World Economy (Maddison), 18
Moody's Investors Service, 294
Mortensen, Dale, 516
Mortgage-backed security (MBS), 124
Mortgage lenders, 123
MPS (MIT-Penn-SSRC) model, 510
Multilateral exchange rates, 363–364
Mundell, Robert, 397, 406, 431–432
Mundell-Fleming model, 397, 407
Mutual funds, 68–69

N

NAFTA. *See* North American Free Trade Agreement (NAFTA)
Narrow banking, 118
Nasdaq Composite Index, 305
National accounts. *See* National Income and Product Accounts (NIPA)
National Bureau of Economic Research (NBER), 98
National income, 525–526, A–2
National Income and Product Accounts (NIPA), 20, 509–510, A–1, A–4
Natural level of output, 150
Natural rate of employment, 162–164
Natural rate of interest, 183
Natural rate of unemployment, 146, 148, 162–164, 444–446, 448–449, 472, 490. *See also* unemployment
in labor market, 146–149
oil prices, impact on, 189–191
price-setting relation with, 147–148
unemployment relational to, 148–149
variations of, 165–166
wages and, 146–149
NBER. *See* National Bureau of Economic Research (NBER)
Neoclassical synthesis, 508–511
Net capital flows, 366
Net capital transfers, 366
Net debt, 479
Net export, 379, 386, 398, 406, A–4
Net import, 47
Net income, 366, 368, 391
Net interest, A–2
Net national product (NNP), A–2
Neutrality of money, 183
Neutral rate of interest, 183
New classicals, 515
New Deal, 186, 508
New growth theory, 516–517
New Keynesian model, 515–517
New York University, 511, 518
Nikkei index, 307
NIPA. *See* National Income and Product Accounts (NIPA)
NNP. *See* net national product (NNP)
Nominal appreciation, 359, 398–399
Nominal bonds, 294
Nominal depreciation, 359, 421
Nominal exchange rate, 358–362, 388, 398, 408, 410, 421–422
Nominal gross domestic product (GDP), 22–24, 29
Nominal income, 69–70, 93
Nominal interest rate, 110–114, 288, 302, 312, 320, 411–412
deflation, relation to, 113–114
expected present discounted values and, 291–292, 312
future, 430
unemployment and, 490
in the United States, 112–113
zero lower bound, relation to, 113–114
Nominal money growth, 488–489
Nominal payments, 312
Nominal rigidities, 516
Nondurable goods, A–3
Nondurable services, A–3
Nonhuman wealth, 314, 319
Noninstitutional civilian population, 136
Nonresidental investment, A–3
Nonresidental structures, A–3
Nonresidential investment, 46–47, 98
Nordhaus, William, 280
North American Free Trade Agreement (NAFTA), 355
North Korea, 256–257
Northwestern University, 266
N-year interest rate, 296

O

Obama, President, 457
OECD. *See* Organization for Economic Cooperation and Development (OECD)
Official deficit, A–5
Oil prices, 188–191, 192
Okun coefficient, 181
Okun's law, 31–32, 180–181, 446
OLS. *See* ordinary least squares (OLS)
One-time devaluations, 422
One-year bond, 294–298
One-year interest rate, 424, 427
OPEC. *See* Organization of Petroleum Exporting Countries (OPEC)
Open economy, 356–364. *See also* financial market; goods market
assets in, 367–369
balance of payments in, 365–367
exchange rates in, 358–364, 369–371, 409–412
exports in, 356–358
in financial market, 364–371
goods in, 358
in goods market, 356–364, 375–395
imports in, 356–358
interest rates in, 369–371
policy in, 404–406
Open market, 73–75. *See also* open economy
Openness, 355. *See also* open economy
Operation Desert Storm, 455
Optimal control, 447, 513
Optimal currency area, 431
Optimal inflation rate, 491–496
Option value of waiting, 348
"Opt-out" clause, 433
Ordinary least squares (OLS), A–13
Organization for Economic Cooperation and Development (OECD), 18, 112, 155, 206, 208–209, 211, 230, 236, 256, 277, 445, 450, 491

Organization of Petroleum Exporting Countries (OPEC), 189
Output, 201–218. *See also* aggregate output; expectations; growth; specific topics
alternative saving rates and, 223–230
behavior of, 229
capital and, 220–223, 237–238
challenges of, 265–284
climate change, impact on, 278–280
current, 327
equilibrium, 51–58
expansion of, 345–346
fluctuations in, 193, 489
future, 327, 343, 344
gap in, 179
global warming, impact on, 278–280
in goods market, determining, 89–91
growth in, 3, 7, 215, 227, 234, 446–447
inequality, impact on, 269–278
investment and, 221–222
level of, 150, 379–380, 386, 403
magnitudes of, sense of, 230–236
per hour worked, 205
per person, 17, 202, 204–205 (*See also* standard of living)
per worker, 205, 213–214, 229–230, 245–250
potential, 179
robots, impact on, 267–269
saving rate and, 225–228
stabilization of, 471–472
steady-state, 232–233
technological progress and, 243–263
unemployment, impact on, 267–269
Overevaluation, 423–424
Overshooting exchange rates, 514

P

Panel datasets, 315
Panel Study of Income Dynamics (PSID), 314–315
Paradox of saving (thrift), 61
Parameters, 49
Parities in power, 454
Participation rate, 26, 136
Patents, 253
Pay-as-you-go (PAYGO) rule, 455
Pay-as-you-go social security system, 231
PAYGO. *See* pay-as-you-go (PAYGO) rule
Payment
balance of, 365
coupon, 294
expected present discounted values and, 290–291
of factor income, A–2
nominal, 312
real, 312, 463
Social Security, 479
Payroll taxes, 526
Peg/pegging, 409–410, 434–435
Penn World Tables, 204
Pension Protection Act of 2006, 236
Permanent income, 314
Persian Gulf War, 368, 455
Personal consumption expenditures, A–3
Personal income, 407, A–3
Personal taxes, 526
Peso, 434
Phelps, Edmund, 163, 511
Phillips, A. W., 32, 155, 158
Phillips curve, 32–33, 155, 341, 422, 449, 489, 510–512, 514
accelerationist, 161–162
deflation and, 169–170
expectations using, 158–162, 512–513
inflation and, 166–169
macroeconomics and, 510–511
medium run, related to, 158–164
natural rate of employment and, 162–164
original, 158
Physical capital, 236, 237–238, 241
Piketty, Thomas, 274
Pissarides, Christopher, 516
Players, 447
Policy. *See also* fiscal policy; monetary policy
coordination of, 385
credibility of, 449–451
expectations and, 447–451
financial, 126
hostage takings and, 448
housing problems, responses to, 125–128
on inflation, 448–449
in macroeconomics, role of, 511
negotiations for, 448
politics and, 451–457
rate of, 120–122, 338–339
rational expectations on theory of, 514
restraints on, 451
time consistency on, 451
uncertainty and, 444–447
on unemployment, 448–449
voters and, 452–453
Policymakers. *See* policy
Political business cycle, 452
Politics, 451–457
Potential output, 179
Poterba, James, 318
Power, 142, 203, 454
PPP. *See* purchasing power parity (PPP)
Premium, 294, 300
Prescott, Edward, 513, 515
Present discounted value, 288–289
Present value, 289–290, 294–295
Pre-tariff price, 408
Price/pricing
asset, 336, 499
bond, 74–75, 293–295, 299
common, 22
computer, 25
decisions, 514
ex-dividend, 300
fire sale, 117–119
hedonic, 25
housing, 308–309, 499
international dollar, 204
in labor market, determination of, 145–146
level of, 27, 492
of oil, 188–191, 192
pre-tariff, 408
relative, 23
risk of, 293
setter/setting of, 147–148, 154, 189, 514, 516
on stock, 299–305
Primary balance, A–5
Primary deficit, 465
Primary surplus, 465–467, 469, 474
Princeton University, 451, 493
Private domestic fixed investment, A–3
Private investment, 473
Private saving, 59
Private sector involvement, 476
Private spending, 337
Prodi, Romano, 456
Product/production
function of, 145–146, 212–213, 236–237, 244–246
GDP, side of, 20–22, A–3–A–4
growth of, 267–268
labor, 145–146
net national, A–2
Profit
corporate, A–2
current, 324–326
expectations of, 320, 324–326
future, 324
of income, 21
from investment, 320–322, 324–328
rate of, 324
sales and, 326–327
Profitability, 326
Propagation mechanism, 193
Propensity to consume, 49
Property rights, 255–257
Proprietors' income, A–2
Protection during employment, 145
PSID. *See* Panel Study of Income Dynamics (PSID)
Public debt, 519
Public saving, 59
Purchases by government, A–3–A–4
Purchasing power, 203
Purchasing power parity (PPP), 14, 203–209, 255–256
Pure inflation, 30

Q

QE3, 496
QE2, 496
Quantitative easing, 496–497
Quits, 137

R

Random walk of consumption, 513
Rate/rating. *See also* specific types of
agencies for, 124
bond, 294
borrowing, 120
discount, 288
employment, 138
federal funds, 8, 79–80, 497–498
of growth, 262, 511
of interest, 183
participation, 26, 136
of policy, 120–121
of profit, 324
of return, 368, 399
teaser, 123
of TFP growth, 262
Rational expectations, 341, 342, 513, 514. *See also* expectations
game theory and, 513
implications of, 513–514
Lucas critique on, 512
in macroeconomics, 511–514
optimal control and, 513
Phillips curve and, 512–513
on policy, theory of, 514
of price setting, 514
of wage setting, 514
Rational speculative bubbles, 307
RBC. *See* real business cycle (RBC) models
R&D. *See* research and development (R&D)
Reagan, Ronald, 407
Reagan administration, 407, 454–455
Real appreciation, 361, 399, 423
Real business cycle (RBC) models, 515, 517
Real debt, 464
Real depreciation, 361, 385–386, 421, 433
Real exchange rate, 358–363, 376–377, 385–386, 399, 408, 420–424, 439–440
Real GDP, 22–25, 29, 40–41
Real income, 92–93
Real interest payments, 463
Real interest rate, 110–111, 292, 312, 320, 421, 463, 474, 480
current, 339
defined, 110
deflation, relation to, 113–114
domestic, 439–440
ex-ante, 113
expected present discounted values and, 291–292, 312
ex-post, 113
foreign, 428, 439–440
negative, 494–495
one-year, 111
in the United States, 112–113
zero lower bound, relation to, 113–114
Real money, 92–93
Real payments, 312
Real policy rate, 338–339

Real US exchange rate, 363
Real wage, 30, 148–149
Re-anchored expectations, 161–162
Recessions, 24, 98–100, 446, 455, 487, 515
Redistribution, 277
Reduction in deficit, 101
Reenen, John Van, 254
Regime, 409–410, 422. *See also* exchange rate regime
Regression, A–13
Relative price, 23
Relative wage, 271–272
Rental cost, 324
Rental income, A–2
Republicans, 443, 454
Reputation, 514
Rescheduling debt, 476
Research
  fertility of, 251
  innovation *vs.* imitation in, 254
  results of, appropriability of, 252–253
  technological progress and process for, 251–252
Research and development (R&D), 250–251, 517
Reservation wage, 142
Reserve, 77, 416
Reserve ratio, 78
Residential investment, 46, A–3
Restrepo, Pascual, 268
Restructuring debt, 476
Retirement, 318
Returns to capital, 213–214
Returns to labor, 213
Revaluation, 359, 410
Ricardian equivalence, 470–471
Ricardo, David, 470
Ricardo-Barro proposition, 470
Right to property, 255–257
Risk
  asset prices and, 306
  aversion to, 114
  of bond, 293, 297–298
  default, 293
  in financial market, 114–115
  premium of, 114–115, 120, 294
  of price, 293
"Robinson Crusoe" economy, 59
Robots, 267–269
Romer, Christina, 348
Romer, Paul, 238, 516
Roubini, Nouriel, 518
Russia, 306
Russian Federation, 306

S
Saboteur, 267
Safe haven, 400
Sale
  in goods market, 88–89
  profit from, 326–327
  tax on, A–2
Samuelson, Paul, 155, 158, 508
Sargent, Thomas, 341, 511–513
Satisfaction in life, 207
"Save More Tomorrow" approach, 236
Save/saving, 59, 69. *See also* saving rate
  equilibrium, impact on, 58–59
  in goods market, 390–391
  paradox of, 61
  private, 59, 101
  propensity to, 60
  public, 59
  for retirement, 318
Saving rate, 215, 219, 250
  alternative, 223–230
  consumption and, 228–230
  golden rule on, 235–236
  household, 236
  increase in, dynamic effects of, 233–235
  output and, 225–228
  steady-state output, effects on, 232–233
  technological progress and, 249–250
  US, 235–236
Scholz, John, 318
Schumpeter, Joseph, 517
Schwartz, Anna, 510
Security/securitization, 124
Seignorage, 477, 494
Semantic traps, 69
Separations, 136–137
September 11, 2001, 98–100
Seshadri, Ananth, 318
SGP. *See* Stability and Growth Pact (SGP)
Shadow cost, 324
Shafir, Eldar, 493
Shares, 299. *See also* stock
Shiller, Robert, 308
Shleifer, Andrei, 517, 518
Shock, 193, 431–432, 515
  financial, 121–122
  propagation mechanisms and, 193
Shoe-leather costs, 491
Short run, 34. *See also* financial market; goods market
  aggregate output, 34
  equilibrium in, 181–182, 186–187, 190, 421–422
  medium run *vs.*, 193
Short-term interest rates, 293
SIVs. *See* structured investment vehicles (SIVs)
Skill-biased technological progress, 272
Slope, A–10
Smooth landing, 298
Smooth taxes, 474
Social Science Research Council, 510
Social Security Administration, 232, 270
Social Security system, 231–232, 318
  fully funded, 231
  pay-as-you-go, 231
  payments, 479
  trust fund, 231, 479
Soft peg, 434
Software, A–3
Solow, Robert, 155, 158, 263, 509
Solow residual, 262
South Korea, 256–257
S&P. *See* Standard and Poor's Corporation (S&P)
Speculative bubbles, 307
Spending
  autonomous, 52
  caps in, 455
  deficit, 472
  government, 47, 50, 405, 473
  private, 337
Spread, 475
Stability and Growth Pact (SGP), 443, 456–457
Stabilization of output, 471–472
Stagflation, 190, 511
Staggering of wage, 514
Standard and Poor's 500 Composite Index, 299
Standard and Poor's Corporation (S&P), 294, 299
Standardized employment deficit, 471
Standard of living, 17, 202, 204, 387
Stanford University, 254, 493, 514
"Starve the beast" strategy, 407, 453
State of technology, 213, 216, 244–245
Static expectations, 322, 333
Statistical discrepancy, 366, A–2
Steady rate of growth, 511
Steady state, 225, 232–233, 239, 241, 248–250
Stevenson, Betsey, 206
Stewart, James, 118
Stock, 247, 299–302, 329. *See also* stock market
  prices on, 299–312
Stock market, 305
  consumer spending in, increase in, 304
  economic activity in, 302–303
  investment and, 323–324
  monetary expansion and, 303–304
  movements in, 299–312
  stock prices in, 299–312
  well-functioning, 302
Strategic interactions, 447
Strauss-Kahn, Dominique, 384
Structural deficit, 471
Structured investment vehicles (SIVs), 123–125
Subprime mortgages, 122–123
Subprimes, 122–123
Sudden stops, 400
Summers, Robert, 204
Supply, 71–73, 93, 154, 488
Supply side/siders, 407, 453
Surplus
  account, 365, 366, 428
  budget, 59, 228, 457
  primary, 465–467, 469, 474
  in trade, 47

T
Targeting, 485, 489–490
Tariff, 355, 408
TARP. *See* Troubled Asset Relief Program (TARP)
Tax
  corporate, 407
  cuts in, 452–453, 470–471
  distortions in, 473–474, 492
  income, 526
  indirect, A–2
  payroll, 526
  personal, 407, 526
  sales, A–2
  smooth, 474
Taylor, John, 490, 514
Taylor rule, 490
T-bills. *See* Treasury bills (T-bills)
Teaser rates, 123
Technological progress, 215, 243–263, 269–270. *See also* research
  capital and, 246–249
  in China, 262–263
  determinants of, 250–264
  future of, 266–267
  GDP, associated with, 25
  growth, rate of, 244–250, 255–258
  institutions and, 255–258
  management of, 254
  output and, 243–263
  production function and, 244–246
  research, and process for, 251–252
  saving rate and, 249–250
  skill-biased, 272
Technological unemployment, 267
Technology. *See also* technological progress
  frontier of, 254
  general-purpose, 266
  movement in, 267
  state of, 213, 216, 244–245
Term premium, 294
Term structure of interest rates, 293
Thaler, Richard, 236, 518
Time consistency, 448–449, 451
Time inconsistency, 513
TIPS. *See* Treasury Inflation Protected Securities (TIPS)
Tirole, Jean, 518
Tobin, James, 322–323, 509
Tobin's q, 323–324
Total wealth, 314
"Township and Village Enterprises," 258
Toxic assets, 124
Trade
  balance in, 47, 379–380, 381, 383, 386, 395, 405, 408, 421, A–4
  deficit in, 47, 365, 381–382, 384–385, 388, 407–409, 422, 434
  goods for, 357
  secrets to, 253
  surplus in, 47
Transactions, 68, 365
Transfers, 47, 366, A–5
Treasury bills (T-bills), 74, 110, 112, 118, 186, 294, 480
Treasury bonds, 294
Treasury Inflation Protected Securities (TIPS), 294
Treasury notes, 294

Treaty on Stability, Coordination and Governance in the Economic and Monetary Union, 457
Trichet, Jean Claude, 347
Troubled Asset Relief Program (TARP), 126
Truman, Harry, 308
Trump, President, 408
Trump administration, 349, 406, 408–409
T-statistic, A–14
"Tulip bubble," 306
"Tulipmania," 306
Tversky, Amos, 493
Twin deficits, 407
Two-year bond, 294–298

U

Uncertainty, 348–349, 444–447
Unconventional monetary policy, 127, 496, 496–498
Uncovered interest parity relation, 368
Underwater, 123
Unemployment, 25–26, 139–141. *See also* unemployment rate
  duration of, 137
  economy, impact on, 26–27
  in euro area, 11–12
  European, 167–168
  fine tuning, 447
  happiness associated with, 28
  inflation, impact on, 156–158, 175
  insurance for, 145, 167
  interest rate and, decreasing of, 490
  in labor market, 139–141
  in medium run, 156–158
  natural rate of, 146, 148–149, 162–164
  output, impact on, 267–269
  policy on, 448–449
  structural rate of, 149
  technological, 267
Unemployment rate, 7, 25–27, 136, 138–140, 446. *See also* unemployment
  civilian, 17
  equilibrium, 148
  harmonized, 17
  level of productivity and, 267
  natural rate of, 148
  robots and, 268–269
  zero inflation and, 448–449
United States (US), 6–9
  capital accumulation in, 231–232
  central bank in, 4 (*See also* Federal Reserve Bank)
  consumption in, 473
  deficits in, 473
  exchange rate in, 363
  exports in, 356–357, 364
  growth in, 7–9
  imports in, 356–357, 364
  inflation rate in, 7
  interest rate in, 112–113
  investment in, 473
  payments, balance of, 365
  policy in, 407
  recessions in, 7–8, 98–100
  saving in, 231–232, 235–236
  Social Security in, 231–232
  trade in, 408–409
  unemployment rate in, 7
University of California, Berkeley, 348
University of Chicago, 314, 341, 447
University of Michigan, 318
University of Minnesota, 341
University of Pennsylvania, 509
University of Wisconsin, 318
Unpredictable consumption, 513
US. *See* United States (US)
Usable observations, A–14
User cost, 324
US Treasury, 71, 126

V

Value
  added, 21, 366
  of bond, 294–295
  expected present discounted, 288–292
  fundamental, 306–307
  present, 289–290, 294–295
Variable/variability
  dependent, A–14
  endogenous, 50
  exogenous, 50
  independent, A–14
  inflation, 493–494
  instrumental, A–16
Venti, Steven, 318
Very foresighted consumer, 314
Vicious cycles, 474–476
Vishny, Robert, 518
VIX, 349
Volcker, Paul, 407
Von Wachter, Till, 270
Voters, 452–453

W

Wage. *See also* wage inequality
  catch-all variables impacting, 145
  determination of, 141–145
  efficiency, 142–144, 516
  indexation of, 168–169
  minimum, 167
  prices, impact on, 144
  real, 30, 148–149
  relative, 271–272
  reservation, 142
  setter/setting, 146–147, 154, 514, 516
  staggering of, 514
  unemployment, impact on, 144–145
Wage inequality, 271–272
  education and, 273
  Gini coefficient and, 276
  growth and, 275–278
  increase in, 271–272
  technological progress and, 273
  and top 1%, 274–275
Waiting, option value of, 348
The Wall Street Journal, 305
War, 454, 472–474
Watson, Thomas, 267
Wealth
  financial, 69, 314
  housing, 314
  human, 314, 470
  nonhuman, 314, 319
  semantic traps associated with, 69
  total, 314
Well-being, emotional, 207
WEO. *See* World Economic Outlook Database (WEO)
Wholesale funding, 125
Wicksell, Knut, 183
Wicksellian rate of interest, 183
Wise, David, 318
Wolfers, Justin, 206
Woodford, Michael, 517
Worker
  capital per, 213–214, 221, 230–232, 245–246, 249
  consumption per, 229, 235
  discouraged, 26, 138
  effective, 245
  investment per effective, 246–248
  output per, 205, 213–214, 227, 229–230, 234, 245–250
World Bank, 18
World Economic Outlook Database (WEO), 18
World output growth, 3
World War I, 423, 452, 478
World War II, 226, 452, 469, 472, 478, 509

X

Xi JinPing, President, 258

Y

Yale University, 280, 322, 509
Yellen, Janet, 516
Yeltsin, Boris, 306
Yield, 293
  bond, 74–75, 293, 296–297
  current, 294
  to maturity, 293–294, 296–299
Yield curve, 293, 298–299

Z

Zero consumption, 229
Zero inflation, 448–449
Zero interest rates, 291
Zero lower bound, 8, 81, 97, 127, 184, 481, 485, 488, 495, 497
  interest rate, relation to, 113–114
  IS-LM model for, 184–185
  monetary policy and, 481, 485

Credits

Material from the International Monetary Fund appears on the following text pages. Courtesy of International Monetary Fund: 004, 007, 011, 011, 013, 014, 168, 499, 354, 382, 382, 382, 387, 387, 407, 410, 421 and 424.

p. iii: Open Your Eyes: Noelle Blanchard
p. xviii: Photo of Olivier Blanchard: Noelle Blanchard
p. 45: Philip J., Juliano/Pearson Education
p. 119: National Archives and Records Administration
p. 203: Dana Fradon/The New Yorker Collection/The Cartoon Bank
p. 253: © Chappatte in "L'Hebdo," Lausanne, www.globecartoon.com
p. 257: NASA Earth Observatory
p. 268: enson/Shutterstock
p. 271: © Chappatte in "Die Weltwoche," Zurich, www.globecartoon.com
p. 305: Dan Wasserman Editorial cartoon/Boston Globe/TNS
p. 309: Jason Zweig.com
p. 317: Roz Chast/The New Yorker Collection/The Cartoon Bank
p. 419: ®1971 by Ed Fisher/The New Yorker collection/The Cartoon Bank
p. 478: BestofColombia, etsy store: www.etsy.com/shop/bestofcolombia
p. 508: J. M. Keynes: Bettmann/Getty Images
P. Samuelson: Rick Friedman/Corbis Historical/Getty Images
p. 509: F. Modigliani: Plinio Lepri/AP Images
J. Tobin: AP Images
R. Solow: Ira Wyman/Sygma/Getty Images
p. 510: L. Klein: B Thumma/AP Images
M. Friedman: Chuck Nacke/Alamy Stock Photo
A. Schwartz: National Women's Hall of Fame
p. 511: E. Phelps: James Leynse/Corbis Historical/Getty Images
R. Lucas: Ralf-Finn Hestoft/Corbis Historical/Getty Images
T. Sargent: Stan Honda/AFP/Getty Images
p. 512: R. Barro: ICHPL Imaginechina/dycj/AP Images
p. 513: R. Hall: Robert Hall
p. 514: R. Dornbusch: Luca Bruno/AP Images
S. Fischer: Achmad Ibrahim/AP Images
J. Taylor: Susana Gonzalez/Getty Images
p. 515: E. Prescott: Giuseppe Aresu/AP Images
p. 516: G. Akerlof: Cabalar/EPA/Shutterstock
J. Yellen: Federal Reserve/Alamy Stock Photo
B. Bernanke: Jay L. Clendenin/Cavan /Aurora Photos/Alamy Stock Photo
P. Romer: Larry Busacca/Getty Images Entertainment/Getty Images
p. 517: P. Aghion: Dr. Philipe Aghion
P. Howitt: Dr. Peter Howitt
A. Shleifer: Dr. Andrei Schleifer
D. Acemoglu: Pearson Education
p. 518: M. Woodford: Dr. Michael Woodford
J. Galí: Arne Dedert/picture-alliance/dpa/AP Images
B. Holmström: Dr. Bengt Holmström
J. Tirole: Pascal Le Segretain/Getty Images
p. 519: Andrews McMeel Universal/GoComics.com

Symbols Used in This Book

Symbol	Term	Introduced in Chapter
$(\ )^d$	Superscript ^{d} means demanded	
$(\ )^e$	Superscript ^{e} means expected	
A	Aggregate private spending	16
	Also: Labor productivity/states of technology	7, 12
α	Effect on the inflation rate of the unemployment rate, given expected inflation	8
B	Goverment debt	22
C	Consumption	3
CU	Currency	4
c	Proportion of money held as currency	4
c_0	Consumption when disposable income equals zero	3
c_1	Propensity to consume	3
D	Checkable deposits	4
	Also: Real dividend on a stock	14
$\$D$	Nominal dividend on a stock	14
δ	Depreciation rate	11
E	Nominal exchange rate (price of domestic currency in terms of foreign currency)	17
$\bar{E}$	Fixed nominal exchange rate	19
E^e	Expected future exchange rate	17
ε	Real exchange rate	17
G	Government spending	3
g_A	Growth rate of technological progress	12
g_K	Growth rate of capital	12
g_N	Growth rate of population	12
g, g_y	Growth rate of output	8
H	High powered money/monetary base/central bank money	4
	Also: Human capital	11
I	Fixed investment	3
IM	Imports	3
i	Nominal interest rate	4
i_1	One-year nominal interest rate	14
i_2	Two-year nominal interest rate	14
i^*	Foreign nominal interest rate	17
	Also: Target interest rate for central bank	
K	Capital stock	10
L	Labor force	2

Symbol	Term	Introduced in Chapter
M	Money stock (nominal)	4
M^d	Money demand (nominal)	4
M^s	Money supply (nominal)	4
m	Markup of prices over wages	7
N	Employment	2
N_n	Natural level of employment	9
NI	Net income payments from the rest of the world	17
NX	Net exports	18
P	GDP deflator/CPI/price level	2
P^*	Foreign price level	17
π	Inflation	2
Π	Profit per unit of capital	15
Q	Real stock price	14
$\$Q$	Nominal stock price	14
R	Bank reserves	4
r	Real interest rate	6
S	Private saving	3
s	Private saving rate	11
T	Net taxes (taxes paid by consumers minus transfers)	3
Tr	Government transfers	22
θ	Reserve ratio of banks	4
U	Unemployment	2
u	Unemployment rate	2
u_n	Natural rate of unemployment	7
V	Present value of a sequence of real payments z	14
$\$V$	Present value of a sequence of nominal payments $\$z$	14
W	Nominal wage	7
Y	Real GDP/Output/Production	2
$\$Y$	Nominal GDP	2
Y_D	Disposable income	3
Y_L	Labor income	15
Y_n	Natural level of output	9
Y^*	Foreign output	18
X	Exports	3
Z	Demand for goods	3
z	Factors that affect the wage, given unemployment	7
	Also: A real payment	14
$\$z$	Nominal payment	14