macroeconomics

fifth edition

N. Gregory Mankiw

part I

Introduction

CHAPTER ONE

The Science of Macroeconomics

> The whole of science is nothing more than the refinement of everyday thinking.
>
> — Albert Einstein

1-1 What Macroeconomists Study

Why have some countries experienced rapid growth in incomes over the past century while others stay mired in poverty? Why do some countries have high rates of inflation while others maintain stable prices? Why do all countries experience recessions and depressions—times of falling incomes and rising unemployment—and how can government policy reduce the frequency and severity of these episodes? Macroeconomics, the study of the economy as a whole, attempts to answer these and many related questions.

To appreciate the importance of macroeconomics, you need only read the newspaper or listen to the news. Every day you can see headlines such as INCOME GROWTH SLOWS, FED MOVES TO COMBAT INFLATION, or STOCKS FALL AMID RECESSION FEARS. Although these macroeconomic events may seem abstract, they touch all of our lives. Business executives forecasting the demand for their products must guess how fast consumers' incomes will grow. Senior citizens living on fixed incomes wonder how fast prices will rise. Recent college graduates looking for jobs hope that the economy will boom and that firms will be hiring.

Because the state of the economy affects everyone, macroeconomic issues play a central role in political debate. Voters are aware of how the economy is doing, and they know that government policy can affect the economy in powerful ways. As a result, the popularity of the incumbent president rises when the economy is doing well and falls when it is doing poorly.

Macroeconomic issues are also at the center of world politics. In recent years, Europe has moved toward a common currency, many Asian countries have experienced financial turmoil and capital flight, and the United States has financed large trade deficits by borrowing from abroad. When world leaders meet, these topics are often high on their agendas.

Although the job of making economic policy falls to world leaders, the job of explaining how the economy as a whole works falls to macroeconomists. Toward this end, macroeconomists collect data on incomes, prices, unemployment, and many other variables from different time periods and different countries. They then attempt to formulate general theories that help to explain these data. Like astronomers studying the evolution of stars or biologists studying the evolution of species, macroeconomists cannot conduct controlled experiments. Instead, they must make use of the data that history gives them. Macroeconomists observe that economies differ from one another and that they change over time. These observations provide both the motivation for developing macroeconomic theories and the data for testing them.

To be sure, macroeconomics is a young and imperfect science. The macroeconomist's ability to predict the future course of economic events is no better than the meteorologist's ability to predict next month's weather. But, as you will see, macroeconomists do know quite a lot about how the economy works. This knowledge is useful both for explaining economic events and for formulating economic policy.

Every era has its own economic problems. In the 1970s, Presidents Richard Nixon, Gerald Ford, and Jimmy Carter all wrestled in vain with a rising rate of inflation. In the 1980s, inflation subsided, but Presidents Ronald Reagan and George Bush presided over large federal budget deficits. Deficits returned in the 1990s, with President Bill Clinton in the Oval Office, the budget deficit shrank and even turned into a budget surplus, but federal taxes as a share of national income reached a historic high. So it was no surprise that when President George W. Bush moved into the White House in 2001, he put a tax cut high on his agenda. The basic principles of macroeconomics do not change from decade to decade, but the macroeconomist must apply these principles with flexibility and creativity to meet changing circumstances.

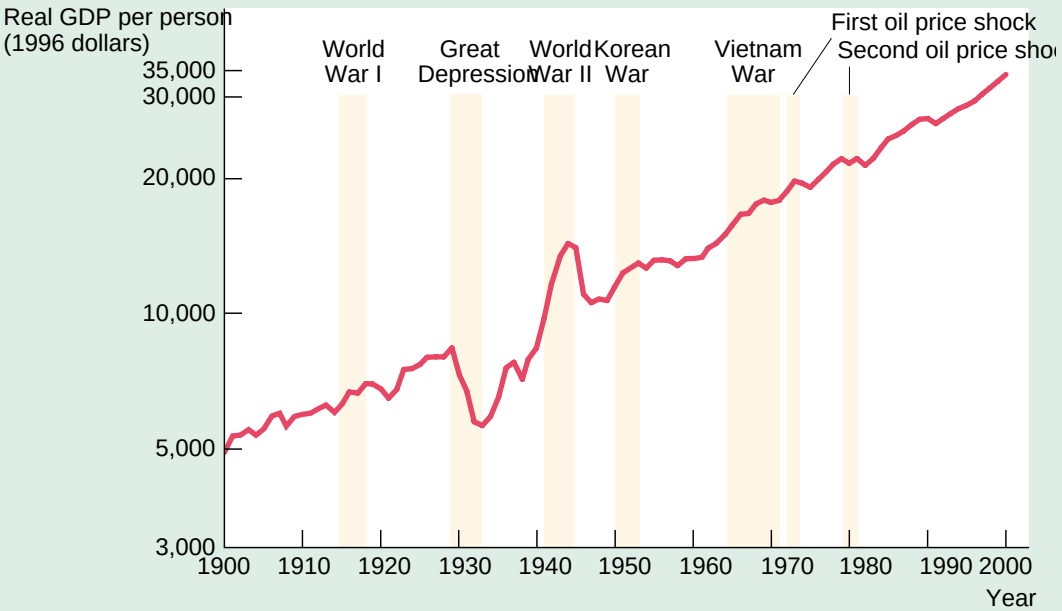
CASE STUDY

The Historical Performance of the U.S. Economy

Economists use many types of data to measure the performance of an economy. Three macroeconomic variables are especially important: real gross domestic product (GDP), the inflation rate, and the unemployment rate. Real GDP measures the total income of everyone in the economy (adjusted for the level of prices). The inflation rate measures how fast prices are rising. The unemployment rate measures the fraction of the labor force that is out of work. Macroeconomists study how these variables are determined, why they change over time, and how they interact with one another.

Figure 1-1 shows real GDP per person in the United States. Two aspects of this figure are noteworthy. First, real GDP grows over time. Real GDP per person is today about five times its level in 1900. This growth in average

figure 1-1



Real GDP per Person in the U.S. Economy

Real GDP measures the total income of everyone in the economy, and real GDP per person measures the income of the average person in the economy. This figure shows that real GDP per person tends to grow over time and that this normal growth is sometimes interrupted by periods of declining income, called recessions or depressions.

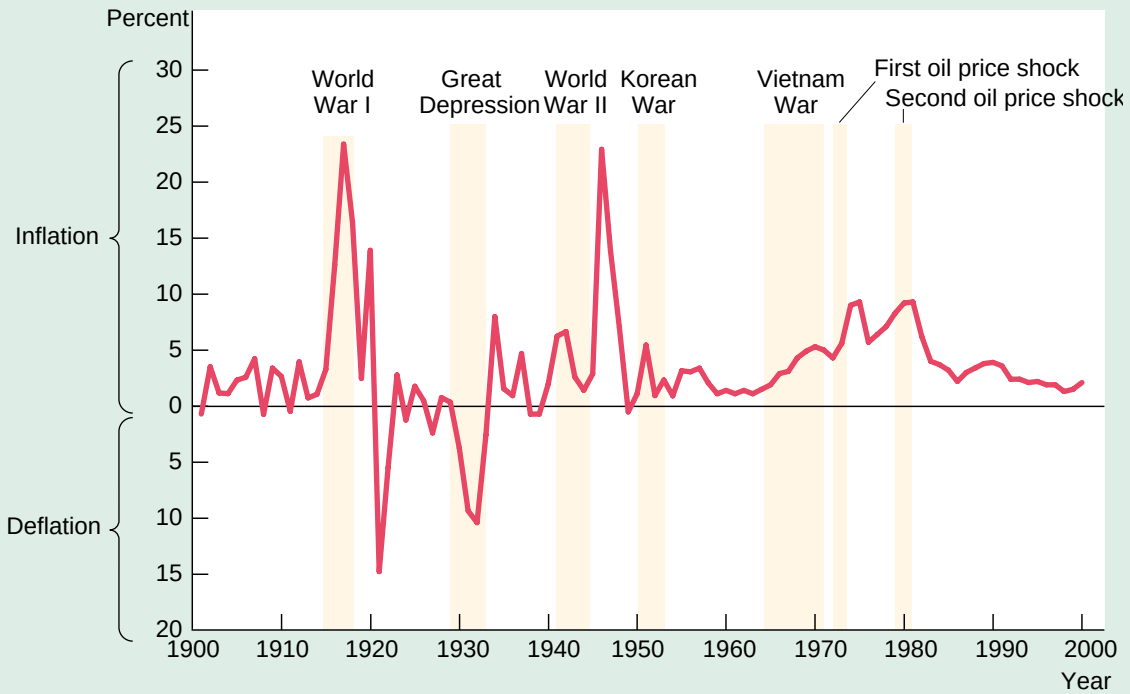
Note: Real GDP is plotted here on a logarithmic scale. On such a scale, equal distances on the vertical axis represent equal percentage changes. Thus, the distance between $5,000 and $10,000 (a 100 percent change) is the same as the distance between $10,000 and $20,000 (a 100 percent change).

Source: U.S. Bureau of the Census, Historical Statistics of the United States: Colonial Times to 1970, and U.S. Department of Commerce.

income allows us to enjoy a higher standard of living than our great-grandparents did. Second, although real GDP rises in most years, this growth is not steady. There are repeated periods during which real GDP falls, the most dramatic instance being the early 1930s. Such periods are called recessions if they are mild and depressions if they are more severe. Not surprisingly, periods of declining income are associated with substantial economic hardship.

Figure 1-2 shows the U.S. inflation rate. You can see that inflation varies substantially. In the first half of the twentieth century, the inflation rate averaged only slightly above zero. Periods of falling prices, called deflation, were almost as common as periods of rising prices. In the past half century, inflation has been the norm. The inflation problem became most severe during the late 1970s, when prices rose at a rate of almost 10 percent per year. In recent years,

figure 1-2



The Inflation Rate in the U.S. Economy
The inflation rate measures the percentage change in the average level of prices from the year before. When the inflation rate is above zero, prices are rising. When it is below zero, prices are falling. If the inflation rate declines but remains positive, prices are rising but at a slower rate.

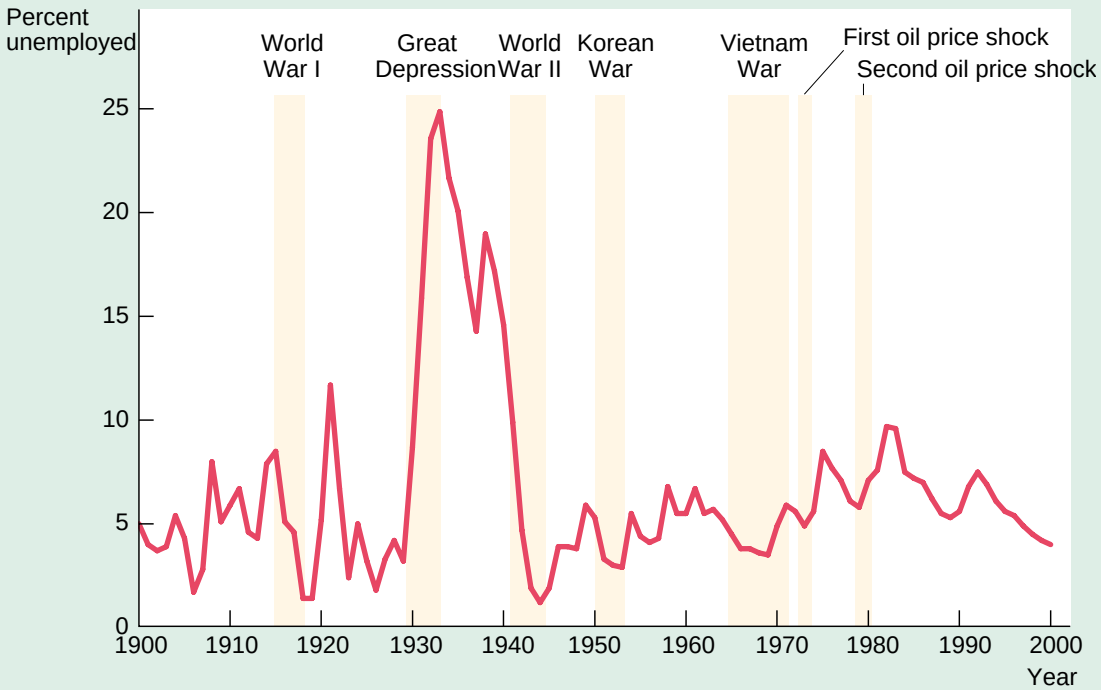
Note: The inflation rate is measured here using the GDP deflator.
Source: U.S. Bureau of the Census, Historical Statistics of the United States: Colonial Times to 1970, and U.S. Department of Commerce.

the inflation rate has been about 2 or 3 percent per year, indicating that prices have been fairly stable.

Figure 1-3 shows the U.S. unemployment rate. Notice that there is always some unemployment in our economy. In addition, although there is no long-term trend, the amount of unemployment varies from year to year. Recessions and depressions are associated with unusually high unemployment. The highest rates of unemployment were reached during the Great Depression of the 1930s.

These three figures offer a glimpse at the history of the U.S. economy. In the chapters that follow, we first discuss how these variables are measured and then develop theories to explain how they behave.

figure 1-3



The Unemployment Rate in the U.S. Economy
The unemployment rate measures the percentage of people in the labor force who do not have jobs. This figure shows that the economy always has some unemployment and that the amount fluctuates from year to year.

Source: U.S. Bureau of the Census, Historical Statistics of the United States: Colonial Times to 1970, and U.S. Department of Commerce.

1-2 How Economists Think

Although economists often study politically charged issues, they try to address these issues with a scientific objectivity. Like any science, economics has its own set of tools—terminology, data, and a way of thinking—that can seem foreign and arcane to the layman. The best way to become familiar with these tools is to practice using them, and this book will afford you ample opportunity to do so. To make these tools less forbidding, however, let's discuss a few of them here.

Theory as Model Building

Young children learn much about the world around them by playing with toy versions of real objects. For instance, they often put together models of cars, trains, or planes. These models are far from realistic, but the model-builder learns

a lot from them nonetheless. The model illustrates the essence of the real object it is designed to resemble.

Economists also use models to understand the world, but an economist's model is more likely to be made of symbols and equations than plastic and glue. Economists build their toy economies to help explain economic variables, such as GDP, inflation, and unemployment. Economic models illustrate, often in mathematical terms, the relationships among the variables. They are useful because they help us to dispense with irrelevant details and to focus on important connections.

Models have two kinds of variables: endogenous variables and exogenous variables. **Endogenous variables** are those variables that a model tries to explain. **Exogenous variables** are those variables that a model takes as given. The purpose of a model is to show how the exogenous variables affect the endogenous variables. In other words, as Figure 1-4 illustrates, exogenous variables come from outside the model and serve as the model's input, whereas endogenous variables are determined inside the model and are the model's output.

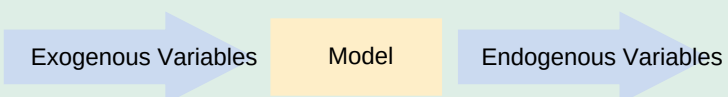
To make these ideas more concrete, let's review the most celebrated of all economic models: the model of supply and demand. Imagine that an economist were interested in figuring out what factors influence the price of pizza and the quantity of pizza sold. He or she would develop a model that described the behavior of pizza buyers, the behavior of pizza sellers, and their interaction in the market for pizza. For example, the economist supposes that the quantity of pizza demanded by consumers (Q^d) depends on the price of pizza (P) and on aggregate income (Y). This relationship is expressed in the equation

$$Q^d = D(P, Y),$$

where $D( )$ represents the demand function. Similarly, the economist supposes that the quantity of pizza supplied by pizzerias (Q^s) depends on the price of pizza (P) and on the price of materials (P_m), such as cheese, tomatoes, flour, and anchovies. This relationship is expressed as

$$Q^s = S(P, P_m),$$

figure 1-4



How Models Work
Models are simplified theories that show the key relationships among economic variables. The exogenous variables are those that come from outside the model. The endogenous variables are those that the model explains. The model shows how changes in the exogenous variables affect the endogenous variables.

where $S(\ )$ represents the supply function. Finally, the economist assumes that the price of pizza adjusts to bring the quantity supplied and quantity demanded into balance:

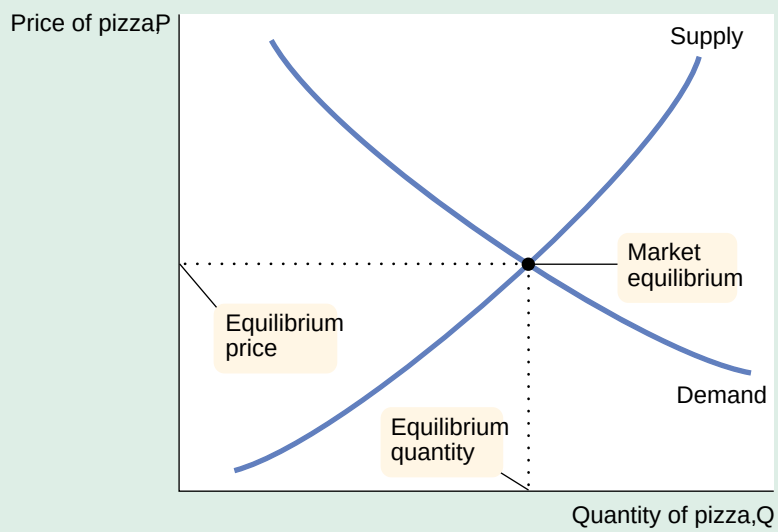
$$Q^s = Q^d.$$

These three equations compose a model of the market for pizza.

The economist illustrates the model with a supply-and-demand diagram, as in Figure 1-5. The demand curve shows the relationship between the quantity of pizza demanded and the price of pizza, while holding aggregate income constant. The demand curve slopes downward because a higher price of pizza encourages consumers to switch to other foods and buy less pizza. The supply curve shows the relationship between the quantity of pizza supplied and the price of pizza, while holding the price of materials constant. The supply curve slopes upward because a higher price of pizza makes selling pizza more profitable, which encourages pizzerias to produce more of it. The equilibrium for the market is the price and quantity at which the supply and demand curves intersect. At the equilibrium price, consumers choose to buy the amount of pizza that pizzerias choose to produce.

This model of the pizza market has two exogenous variables and two endogenous variables. The exogenous variables are aggregate income and the price of materials. The model does not attempt to explain them but takes them as given (perhaps to be explained by another model). The endogenous variables are the

figure 1-5

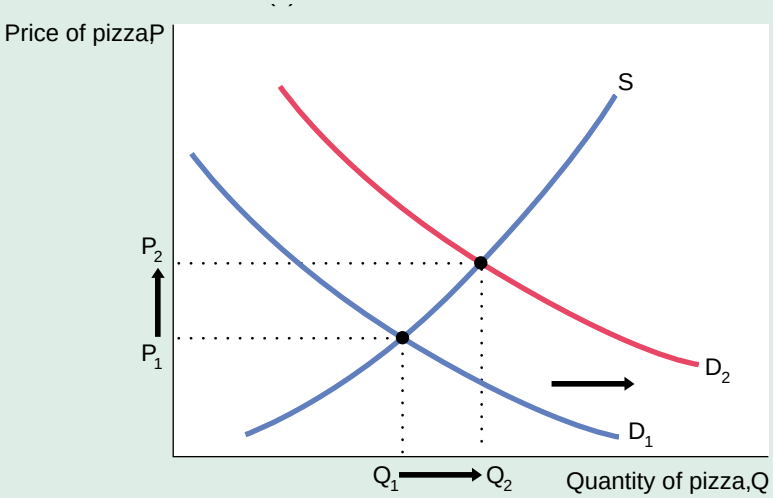


The Model of Supply and Demand
The most famous economic model is that of supply and demand for a good or service—in this case, pizza. The demand curve is a downward-sloping curve relating the price of pizza to the quantity of pizza that consumers demand. The supply curve is an upward-sloping curve relating the price of pizza to the quantity of pizza that pizzerias supply. The price of pizza adjusts until the quantity supplied equals the quantity demanded. The point where the two curves cross is the market equilibrium, which shows the equilibrium price of pizza and the equilibrium quantity of pizza.

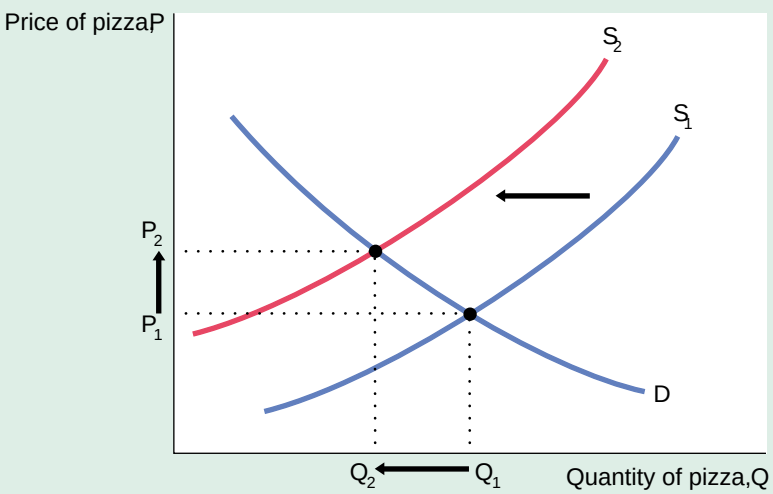
price of pizza and the quantity of pizza exchanged. These are the variables that the model attempts to explain.

The model can be used to show how a change in one of the exogenous variables affects both endogenous variables. For example, if aggregate income increases, then the demand for pizza increases, as in panel (a) of Figure 1-6. The model shows that both the equilibrium price and the equilibrium quantity of pizza rise. Similarly, if the price of materials increases, then the supply of pizza decreases, as in panel (b) of Figure 1-6. The model shows that in this case the equilibrium price of pizza rises and the equilibrium quantity of pizza falls. Thus, the model shows how changes in aggregate income or in the price of materials affect price and quantity in the market for pizza.

figure 1-6



(b) A Shift in Supply



Changes in Equilibrium
In panel (a), a rise in aggregate income causes the demand for pizza to increase: at any given price, consumers now want to buy more pizza. This is represented by a rightward shift in the demand curve from D_1 to D_2. The market moves to the new intersection of supply and demand. The equilibrium price rises from P_1 to P_2, and the equilibrium quantity of pizza rises from Q_1 to Q_2. In panel (b), a rise in the price of materials decreases the supply of pizza: at any given price, pizzerias find that the sale of pizza is less profitable and therefore choose to produce less pizza. This is represented by a leftward shift in the supply curve from S_1 to S_2. The market moves to the new intersection of supply and demand. The equilibrium price rises from P_1 to P_2, and the equilibrium quantity falls from Q_1 to Q_2.

Like all models, this model of the pizza market makes simplifying assumptions. The model does not take into account, for example, that every pizzeria is in a different location. For each customer, one pizzeria is more convenient than the others, and thus pizzerias have some ability to set their own prices. Although the model assumes that there is a single price for pizza, in fact there could be a different price at every pizzeria.

How should we react to the model's lack of realism? Should we discard the simple model of pizza supply and pizza demand? Should we attempt to build a more complex model that allows for diverse pizza prices? The answers to these questions depend on our purpose. If our goal is to explain how the price of cheese affects the average price of pizza and the amount of pizza sold, then the diversity of pizza prices is probably not important. The simple model of the pizza market does a good job of addressing that issue. Yet if our goal is to explain why towns with three pizzerias have lower pizza prices than towns with one pizzeria, the simple model is less useful.

FYI

Using Functions to Express Relationships Among Variables

All economic models express relationships among economic variables. Often, these relationships are expressed as functions. A function is a mathematical concept that shows how one variable depends on a set of other variables. For example, in the model of the pizza market, we said that the quantity of pizza demanded depends on the price of pizza and on aggregate income. To express this, we use functional notation to write

$$Q^d = D(P, Y).$$

This equation says that the quantity of pizza demanded Q^d is a function of the price of pizza P and aggregate income Y. In functional notation, the variable preceding the parentheses denotes the function. In this case, $D(\ )$ is the function expressing how the variables in parentheses determine the quantity of pizza demanded.

If we knew more about the pizza market, we could give a numerical formula for the quantity of pizza demanded. We might be able to write

$$Q^d = 60 - 10P + 2Y.$$

In this case, the demand function is

$$D(P, Y) = 60 - 10P + 2Y.$$

For any price of pizza and aggregate income, this function gives the corresponding quantity of pizza demanded. For example, if aggregate income is $10 and the price of pizza is $2, then the quantity of pizza demanded is 60 pies; if the price of pizza rises to $3, the quantity of pizza demanded falls to 50 pies.

Functional notation allows us to express a relationship among variables even when the precise numerical relationship is unknown. For example, we might know that the quantity of pizza demanded falls when the price rises from $2 to $3, but we might not know by how much it falls. In this case, functional notation is useful: as long as we know that a relationship among the variables exists, we can remind ourselves of that relationship using functional notation.

The art in economics is in judging when an assumption is clarifying and when it is misleading. Any model constructed to be completely realistic would be too complicated for anyone to understand. Simplification is a necessary part of building a useful model. Yet models lead to incorrect conclusions if they assume away features of the economy that are crucial to the issue at hand. Economic modeling therefore requires care and common sense.

A Multitude of Models

Macroeconomists study many facets of the economy. For example, they examine the role of saving in economic growth, the impact of labor unions on unemployment, the effect of inflation on interest rates, and the influence of trade policy on the trade balance and exchange rates. Macroeconomics is as diverse as the economy.

Although economists use models to address all these issues, no single model can answer all questions. Just as carpenters use different tools for different tasks, economists uses different models to explain different economic phenomena. Students of macroeconomics, therefore, must keep in mind that there is no single "correct" model useful for all purposes. Instead, there are many models, each of which is useful for shedding light on a different facet of the economy. The field of macroeconomics is like a Swiss army knife—a set of complementary but distinct tools that can be applied in different ways in different circumstances.

This book therefore presents many different models that address different questions and that make different assumptions. Remember that a model is only as good as its assumptions and that an assumption that is useful for some purposes may be misleading for others. When using a model to address a question, the economist must keep in mind the underlying assumptions and judge whether these are reasonable for the matter at hand.

Prices: Flexible Versus Sticky

Throughout this book, one group of assumptions will prove especially important—those concerning the speed with which wages and prices adjust. Economists normally presume that the price of a good or a service moves quickly to bring quantity supplied and quantity demanded into balance. In other words, they assume that a market goes to the equilibrium of supply and demand. This assumption is called *market clearing* and is central to the model of the pizza market discussed earlier. For answering most questions, economists use market-clearing models.

Yet the assumption of continuous market clearing is not entirely realistic. For markets to clear continuously, prices must adjust instantly to changes in supply and demand. In fact, however, many wages and prices adjust slowly. Labor contracts often set wages for up to three years. Many firms leave their product prices the same for long periods of time—for example, magazine publishers typically

change their newsstand prices only every three or four years. Although market-clearing models assume that all wages and prices are flexible, in the real world some wages and prices are sticky.

The apparent stickiness of prices does not make market-clearing models useless. After all, prices are not stuck forever; eventually, they do adjust to changes in supply and demand. Market-clearing models might not describe the economy at every instant, but they do describe the equilibrium toward which the economy gravitates. Therefore, most macroeconomists believe that price flexibility is a good assumption for studying long-run issues, such as the growth in real GDP that we observe from decade to decade.

For studying short-run issues, such as year-to-year fluctuations in real GDP and unemployment, the assumption of price flexibility is less plausible. Over short periods, many prices are fixed at predetermined levels. Therefore, most macroeconomists believe that price stickiness is a better assumption for studying the behavior of the economy in the short run.

Microeconomic Thinking and Macroeconomic Models

Microeconomics is the study of how households and firms make decisions and how these decisionmakers interact in the marketplace. A central principle of microeconomics is that households and firms optimize—they do the best they can for themselves given their objectives and the constraints they face. In microeconomic models, households choose their purchases to maximize their level of satisfaction, which economists call utility, and firms make production decisions to maximize their profits.

Because economy-wide events arise from the interaction of many households and many firms, macroeconomics and microeconomics are inextricably linked. When we study the economy as a whole, we must consider the decisions of individual economic actors. For example, to understand what determines total consumer spending, we must think about a family deciding how much to spend today and how much to save for the future. To understand what determines total investment spending, we must think about a firm deciding whether to build a new factory. Because aggregate variables are the sum of the variables describing many individual decisions, macroeconomic theory rests on a microeconomic foundation.

Although microeconomic decisions always underlie economic models, in many models the optimizing behavior of households and firms is implicit rather than explicit. The model of the pizza market we discussed earlier is an example. Household decisions about how much pizza to buy underlie the demand for pizza, and pizzeria decisions about how much pizza to produce underlie the supply of pizza. Presumably, households make their decisions to maximize utility, and pizzerias make their decisions to maximize profit. Yet the model did not focus on these microeconomic decisions; it left them in the background. Similarly, in much of macroeconomics, the optimizing behavior of households and firms is left implicit.

1-3 How This Book Proceeds

This book has six parts. This chapter and the next make up Part One, the Introduction. Chapter 2 discusses how economists measure economic variables, such as aggregate income, the inflation rate, and the unemployment rate.

Part Two, "Classical Theory: The Economy in the Long Run," presents the classical model of how the economy works. The key assumption of the classical model is that prices are flexible. That is, with rare exceptions, the classical model assumes market clearing. Because the assumption of flexible prices typically describes the economy only in the long run, classical theory is best suited for analyzing a time horizon of at least several years.

Part Three, "Growth Theory: The Economy in the Very Long Run," builds on the classical model. It maintains the assumption of market clearing but adds a new emphasis on growth in the capital stock, the labor force, and technological knowledge. Growth theory is designed to explain how the economy evolves over a period of several decades.

Part Four, "Business Cycle Theory: The Economy in the Short Run," examines the behavior of the economy when prices are sticky. The non-market-clearing model developed here is designed to analyze short-run issues, such as the reasons for economic fluctuations and the influence of government policy on those fluctuations. It is best suited to analyzing the changes in the economy we observe from month to month or from year to year.

Part Five, "Macroeconomic Policy Debates," builds on the previous analysis to consider what role the government should take in the economy. It considers how, if at all, the government should respond to short-run fluctuations in real GDP and unemployment. It also examines the various views on the effects of government debt.

Part Six, "More on the Microeconomics Behind Macroeconomics," presents some of the microeconomic models that are useful for analyzing macroeconomic issues. For example, it examines the household decisions regarding how much to consume and how much money to hold and the firm's decision regarding how much to invest. These individual decisions together form the larger macroeconomic picture. The goal of studying these microeconomic decisions in detail is to refine our understanding of the aggregate economy.

Summary

1. Macroeconomics is the study of the economy as a whole, including growth in incomes, changes in prices, and the rate of unemployment. Macroeconomists attempt both to explain economic events and to devise policies to improve economic performance.
2. To understand the economy, economists use models—theories that simplify reality in order to reveal how exogenous variables influence endogenous variables. The art in the science of economics is in judging whether a model

captures the important economic relationships for the matter at hand. Because no single model can answer all questions, macroeconomists use different models to look at different issues.

3. A key feature of a macroeconomic model is whether it assumes that prices are flexible or sticky. According to most macroeconomists, models with flexible prices describe the economy in the long run, whereas models with sticky prices offer a better description of the economy in the short run.
4. Microeconomics is the study of how firms and individuals make decisions and how these decisionmakers interact. Because macroeconomic events arise from many microeconomic interactions, macroeconomists use many of the tools of microeconomics.

KEY CONCEPTS

Macroeconomics
Real GDP
Inflation rate
Unemployment rate
Recession
Depression
Deflation
Models
Endogenous variables
Exogenous variables
Market clearing
Flexible and sticky prices
Microeconomics

QUESTIONS FOR REVIEW

1. Explain the difference between macroeconomics and microeconomics. How are these two fields related?
2. Why do economists build models?
3. What is a market-clearing model? When is the assumption of market clearing appropriate?

PROBLEMS AND APPLICATIONS

1. What macroeconomic issues have been in the news lately?
2. What do you think are the defining characteristics of a science? Does the study of the economy have these characteristics? Do you think macroeconomics should be called a science? Why or why not?
3. Use the model of supply and demand to explain how a fall in the price of frozen yogurt would affect the price of ice cream and the quantity of ice cream sold. In your explanation, identify the exogenous and endogenous variables.
4. How often does the price you pay for a haircut change? What does your answer imply about the usefulness of market-clearing models for analyzing the market for haircuts?

CHAPTER 2 TWO

The Data of Macroeconomics

> It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to Þt facts.
>
> Ñ Sherlock Holmes

Scientists, economists, and detectives have much in common: they all want to Þgure out whatÕs going on in the world around them. To do this, they rely on both theory and observation. They build theories in an attempt to make sense of what they see happening. They then turn to more systematic observation to evaluate the theoriesÕ validity. Only when theory and evidence come into line do they feel they understand the situation. This chapter discusses the types of observation that economists use to develop and test their theories.

Casual observation is one source of information about whatÕs happening in the economy. When you go shopping, you see how fast prices are rising. When you look for a job, you learn whether Þrms are hiring. Because we are all participants in the economy, we get some sense of economic conditions as we go about our lives.

A century ago, economists monitoring the economy had little more to go on than these casual observations. Such fragmentary information made economic policymaking all the more difÞcult. One personÕs anecdote would suggest the economy was moving in one direction, while a different personÕs anecdote would suggest it was moving in another. Economists needed some way to combine many individual experiences into a coherent whole. There was an obvious solution: as the old quip goes, the plural of ÒanecdoteÓ is Ódata.Ó

Today, economic data offer a systematic and objective source of information, and almost every day the newspaper has a story about some newly released statistic. Most of these statistics are produced by the government. Various government agencies survey households and Þrms to learn about their economic activityÑ how much they are earning, what they are buying, what prices they are charging, whether they have a job or are looking for work, and so on. From these surveys, various statistics are computed that summarize the state of the economy. Economists use these statistics to study the economy; policymakers use them to monitor developments and formulate policies.

This chapter focuses on the three statistics that economists and policymakers use most oftenGross domestic product , or GDP , tells us the nationÕs total income

and the total expenditure on its output of goods and services. The consumer price index, or CPI, measures the level of prices. The unemployment rate tells us the fraction of workers who are unemployed. In the following pages, we see how these statistics are computed and what they tell us about the economy.

2-1 Measuring the Value of Economic Activity: Gross Domestic Product

Gross domestic product is often considered the best measure of how well the economy is performing. This statistic is computed every three months by the Bureau of Economic Analysis (a part of the U.S. Department of Commerce) from a large number of primary data sources. The goal of GDP is to summarize in a single number the dollar value of economic activity in a given period of time.

There are two ways to view this statistic. One way to view GDP is as the total income of everyone in the economy. Another way to view GDP is as the total expenditure on the economy's output of goods and services. From either viewpoint, it is clear why GDP is a gauge of economic performance. GDP measures something people care about—their incomes. Similarly, an economy with a large output of goods and services can better satisfy the demands of households, businesses, and the government.

How can GDP measure both the economy's income and the expenditure on its output? The reason is that these two quantities are really the same: for the economy as a whole, income must equal expenditure. That fact, in turn, follows from an even more fundamental one: because every transaction has both a buyer and a seller, every dollar of expenditure by a buyer must become a dollar of income to a seller. When Joe paints Jane's house for $1,000, that $1,000 is income to Joe and expenditure by Jane. The transaction contributes $1,000 to GDP, regardless of whether we are adding up all income or adding up all expenditure.

To understand the meaning of GDP more fully, we turn to national income accounting, the accounting system used to measure GDP and many related statistics.

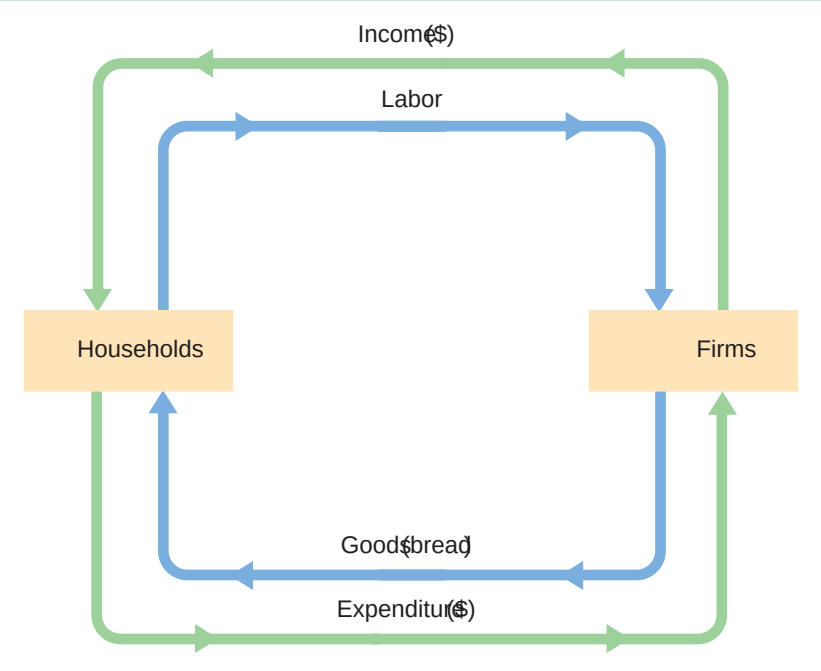
Income, Expenditure, and the Circular Flow

Imagine an economy that produces a single good, bread, from a single input, labor. Figure 2-1 illustrates all the economic transactions that occur between households and firms in this economy.

The inner loop in Figure 2-1 represents the flows of bread and labor. The households sell their labor to the firms. The firms use the labor of their workers to produce bread, which the firms in turn sell to the households. Hence, labor flows from households to firms, and bread flows from firms to households.

The outer loop in Figure 2-1 represents the corresponding flow of dollars. The households buy bread from the firms. The firms use some of the revenue

figure 2-1



The Circular Flow This figure illustrates the flows between firms and households in an economy that produces one good, bread, from one input, labor. The inner loop represents the flows of labor and bread: households sell their labor to firms, and the firms sell the bread they produce to households. The outer loop represents the corresponding flows of dollars: households pay the firms for the bread, and the firms pay wages and profit to the households. In this economy, GDP is both the total expenditure on bread and the total income from the production of bread.

from these sales to pay the wages of their workers, and the remainder is the profit belonging to the owners of the firms (who themselves are part of the household sector). Hence, expenditure on bread flows from households to firms, and income in the form of wages and profit flows from firms to households.

GDP measures the flow of dollars in this economy. We can compute it in two ways. GDP is the total income from the production of bread, which equals the sum of wages and profit—the top half of the circular flow of dollars. GDP is also the total expenditure on purchases of bread—the bottom half of the circular flow of dollars. To compute GDP, we can look at either the flow of dollars from firms to households or the flow of dollars from households to firms.

These two ways of computing GDP must be equal because the expenditure of buyers on products is, by the rules of accounting, income to the sellers of those products. Every transaction that affects expenditure must affect income, and every transaction that affects income must affect expenditure. For example, suppose that a farm produces and sells one more loaf of bread to a household. Clearly this transaction raises total expenditure on bread, but it also has an equal effect on total income. If the farm produces the extra loaf without hiring any more labor (such as by making the production process more efficient), then profit increases. If the farm produces the extra loaf by hiring more labor, then wages increase. In both cases, expenditure and income increase equally.

FYI

Stocks and Flows

Many economic variables measure a quantity of something—a quantity of money, a quantity of goods, and so on. Economists distinguish between two types of quantity variables: stocks and flows. A stock is a quantity measured at a given point in time, whereas a flow is a quantity measured per unit of time.

The bathtub, shown in Figure 2-2, is the classic example used to illustrate stocks and flows. The amount of water in the tub is a stock: it is the quantity of water in the tub at a given point in time. The amount of water coming out of the faucet is a flow: it is the quantity of water being added to the tub per unit of time. Note that we measure stocks and flows in different units. We say that the bathtub contains 50 gallons of water, but that water is coming out of the faucet at 5 gallons per minute.

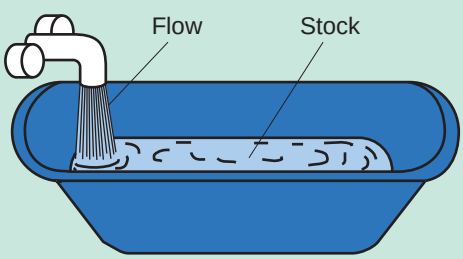
GDP is probably the most important flow variable in economics: it tells us how many dollars are flowing around the economy's circular flow per unit of time. When you hear someone say that the U.S. GDP is $10 trillion, you should understand that this means that it is $10 trillion per year. (Equivalently, we could say that U.S. GDP is $317,000 per second.)

Stocks and flows are often related. In the bathtub example, these relationships are clear. The stock of water in the tub represents the accumulation of the flow out of the faucet, and the flow of water represents the change in the stock. When building theories to explain economic variables, it is often useful to determine whether the variables are stocks or flows and whether any relationships link them.

Here are some examples of related stocks and flows that we study in future chapters:

- A person's wealth is a stock; income and expenditure are flows.
- The number of unemployed people is a stock; the number of people losing their jobs is a flow.
- The amount of capital in the economy is a stock; the amount of investment is a flow.
- The government debt is a stock; the government budget deficit is a flow.

figure 2-2



Stocks and Flows The amount of water in a bathtub is a stock: it is a quantity measured at a given moment in time. The amount of water coming out of the faucet is a flow: it is a quantity measured per unit of time.

Rules for Computing GDP

In an economy that produces only bread, we can compute GDP by adding up the total expenditure on bread. Real economies, however, include the production and sale of a vast number of goods and services. To compute GDP for such a complex economy, it will be helpful to have a more precise definition: gross

domestic product (GDP) is the market value of all final goods and services produced within an economy in a given period of time. To see how this definition is applied, let's discuss some of the rules that economists follow in constructing this statistic.

Adding Apples and Oranges The U.S. economy produces many different goods and services—hamburgers, haircuts, cars, computers, and so on. GDP combines the value of these goods and services into a single measure. The diversity of products in the economy complicates the calculation of GDP because different products have different values.

Suppose, for example, that the economy produces four apples and three oranges. How do we compute GDP? We could simply add apples and oranges and conclude that GDP equals seven pieces of fruit. But this makes sense only if we thought apples and oranges had equal value, which is generally not true. (This would be even clearer if the economy had produced four watermelons and three grapes.)

To compute the total value of different goods and services, the national income accounts use market prices because these prices reflect how much people are willing to pay for a good or service. Thus, if apples cost $0.50 each and oranges cost $1.00 each, GDP would be

$$\begin{aligned}\text{GDP} &= (\text{Price of Apples} \times \text{Quantity of Apples}) \\ &\quad + (\text{Price of Oranges} \times \text{Quantity of Oranges}) \\ &= (\$0.50 \times 4) + (\$1.00 \times 3) \\ &= \$5.00.\end{aligned}$$

GDP equals $5.00—the value of all the apples, $2.00, plus the value of all the oranges, $3.00.

Used Goods When the Topps Company makes a package of baseball cards and sells it for 50 cents, that 50 cents is added to the nation's GDP. But what about when a collector sells a rare Mickey Mantle card to another collector for $500? That $500 is not part of GDP. GDP measures the value of currently produced goods and services. The sale of the Mickey Mantle card is the transfer of an asset, not an addition to the economy's income. Thus, the sale of used goods is not included as part of GDP.

The Treatment of Inventories Imagine that a bakery hires workers to produce more bread, pays their wages, and then fails to sell the additional bread. How does this transaction affect GDP?

The answer depends on what happens to the unsold bread. First, suppose that the bread spoils. In this case, the firm has paid more in wages but has not received any additional revenue, so the firm's profit is reduced by the amount that wages are increased. Total expenditure in the economy hasn't changed because no one buys the bread. Total income has changed either—although more is distributed as wages and less as profit. Because the transaction affects neither expenditure nor income, it does not alter GDP.

Now suppose, instead, that the bread is put into inventory to be sold later. In this case, the transaction is treated differently. The owners of the firm are assumed to have "purchased" the bread for the firm's inventory, and the firm's profit is not

reduced by the additional wages it has paid. Because the higher wages raise total income, and greater spending on inventory raises total expenditure, the economy's GDP rises.

What happens later when the firm sells the bread out of inventory? This case is much like the sale of a used good. There is spending by bread consumers, but there is inventory disinvestment by the firm. This negative spending by the firm offsets the positive spending by consumers, so the sale out of inventory does not affect GDP.

The general rule is that when a firm increases its inventory of goods, this investment in inventory is counted as expenditure by the firm owners. Thus, production for inventory increases GDP just as much as production for sale. A sale out of inventory, however, is a combination of positive spending (the purchase) and negative spending (inventory disinvestment), so it does not change GDP. This treatment of inventories ensures that GDP reflects the economy's current production of goods and services.

Intermediate Goods and Value Added Many goods are produced in stages: raw materials are processed into intermediate goods by one firm and then sold to another firm for final processing. How should we treat such products when computing GDP? For example, suppose a cattle rancher sells one-quarter pound of meat to McDonald's for $0.50, and then McDonald's sells you a hamburger for $1.50. Should GDP include both the meat and the hamburger (a total of $2.00), or just the hamburger ($1.50)?

The answer is that GDP includes only the value of final goods. Thus, the hamburger is included in GDP but the meat is not: GDP increases by $1.50, not by $2.00. The reason is that the value of intermediate goods is already included as part of the market price of the final goods in which they are used. To add the intermediate goods to the final goods would be double counting—that is, the meat would be counted twice. Hence, GDP is the total value of final goods and services produced.

One way to compute the value of all final goods and services is to sum the value added at each stage of production. The value added of a firm equals the value of the firm's output less the value of the intermediate goods that the firm purchases. In the case of the hamburger, the value added of the rancher is $0.50 (assuming that the rancher bought no intermediate goods), and the value added of McDonald's is $1.50 − $0.50, or $1.00. Total value added is $0.50 + $1.00, which equals $1.50. For the economy as a whole, the sum of all value added must equal the value of all final goods and services. Hence, GDP is also the total value added of all firms in the economy.

Housing Services and Other Imputations Although most goods and services are valued at their market prices when computing GDP, some are not sold in the marketplace and therefore do not have market prices. If GDP is to include the value of these goods and services, we must use an estimate of their value. Such an estimate is called an *imputed value*.

Imputations are especially important for determining the value of housing. A person who rents a house is buying housing services and providing income for the landlord; the rent is part of GDP, both as expenditure by the renter and as income for the landlord. Many people, however, live in their own homes. Although they do not pay rent to a landlord, they are enjoying housing services similar to those that

renters purchase. To take account of the housing services enjoyed by homeowners, GDP includes the "rent" that these homeowners "pay" to themselves. Of course, homeowners do not in fact pay themselves this rent. The Department of Commerce estimates what the market rent for a house would be if it were rented and includes that imputed rent as part of GDP. This imputed rent is included both in the homeowner's expenditure and in the homeowner's income.

Imputations also arise in valuing government services. For example, police officers, firefighters, and senators provide services to the public. Giving a value to these services is difficult because they are not sold in a marketplace and therefore do not have a market price. The national income accounts include these services in GDP by valuing them at their cost. That is, the wages of these public servants are used as a measure of the value of their output.

In many cases, an imputation is called for in principle but, to keep things simple, is not made in practice. Because GDP includes the imputed rent on owner-occupied houses, one might expect it also to include the imputed rent on cars, lawn mowers, jewelry, and other durable goods owned by households. Yet the value of these rental services is left out of GDP. In addition, some of the output of the economy is produced and consumed at home and never enters the marketplace. For example, meals cooked at home are similar to meals cooked at a restaurant, yet the value added in meals at home is left out of GDP.

Finally, no imputation is made for the value of goods and services sold in the underground economy. The underground economy is the part of the economy that people hide from the government either because they wish to evade taxation or because the activity is illegal. Domestic workers "paid off the books" is one example. The illegal drug trade is another.

Because the imputations necessary for computing GDP are only approximate, and because the value of many goods and services is left out altogether, GDP is an imperfect measure of economic activity. These imperfections are most problematic when comparing standards of living across countries. The size of the underground economy, for instance, varies from country to country. Yet as long as the magnitude of these imperfections remains fairly constant over time, GDP is useful for comparing economic activity from year to year.

Real GDP Versus Nominal GDP

Economists use the rules just described to compute GDP, which values the economy's total output of goods and services. But is GDP a good measure of economic well-being? Consider once again the economy that produces only apples and oranges. In this economy GDP is the sum of the value of all the apples produced and the value of all the oranges produced. That is,

$$\begin{aligned}\text{GDP} = &\ (\text{Price of Apples} \times \text{Quantity of Apples}) \\ &+ (\text{Price of Oranges} \times \text{Quantity of Oranges}).\end{aligned}$$

Notice that GDP can increase either because prices rise or because quantities rise.

It is easy to see that GDP computed this way is not a good gauge of economic well-being. That is, this measure does not accurately reflect how well the

economy can satisfy the demands of households, firms, and the government. If all prices doubled without any change in quantities, GDP would double. Yet it would be misleading to say that the economy's ability to satisfy demands has doubled, because the quantity of every good produced remains the same. Economists call the value of goods and services measured at current prices **nominal GDP**.

A better measure of economic well-being would tally the economy's output of goods and services and would not be influenced by changes in prices. For this purpose, economists use **real GDP**, which is the value of goods and services measured using a constant set of prices. That is, real GDP shows what would have happened to expenditure on output if quantities had changed but prices had not.

To see how real GDP is computed, imagine we wanted to compare output in 2002 and output in 2003 in our apple-and-orange economy. We could begin by choosing a set of prices, called **base-year prices**, such as the prices that prevailed in 2002. Goods and services are then added up using these base-year prices to value the different goods in both years. Real GDP for 2002 would be

$$\begin{aligned}\text{Real GDP} = {} & (\text{2002 Price of Apples} \times \text{2002 Quantity of Apples}) \\ & + (\text{2002 Price of Oranges} \times \text{2002 Quantity of Oranges}).\end{aligned}$$

Similarly, real GDP in 2003 would be

$$\begin{aligned}\text{Real GDP} = {} & (\text{2002 Price of Apples} \times \text{2003 Quantity of Apples}) \\ & + (\text{2002 Price of Oranges} \times \text{2003 Quantity of Oranges}).\end{aligned}$$

And real GDP in 2004 would be

$$\begin{aligned}\text{Real GDP} = {} & (\text{2002 Price of Apples} \times \text{2004 Quantity of Apples}) \\ & + (\text{2002 Price of Oranges} \times \text{2004 Quantity of Oranges}).\end{aligned}$$

Notice that 2002 prices are used to compute real GDP for all three years. Because the prices are held constant, real GDP varies from year to year only if the quantities produced vary. Because a society's ability to provide economic satisfaction for its members ultimately depends on the quantities of goods and services produced, real GDP provides a better measure of economic well-being than nominal GDP.

The GDP Deflator

From nominal GDP and real GDP we can compute a third statistic: the GDP deflator. The **GDP deflator**, also called the **implicit price deflator** for GDP, is defined as the ratio of nominal GDP to real GDP:

$$\text{GDP Deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}}.$$

The GDP deflator reflects what is happening to the overall level of prices in the economy.

To better understand this, consider again an economy with only one good, bread. If P is the price of bread and Q is the quantity sold, then nominal GDP is

the total number of dollars spent on bread in that year, $P \times Q$. Real GDP is the number of loaves of bread produced in that year times the price of bread in some base year, $P_{base} \times Q$. The GDP deflator is the price of bread in that year relative to the price of bread in the base year, P / P_{base}.

The definition of the GDP deflator allows us to separate nominal GDP into two parts: one part measures quantities (real GDP) and the other measures prices (the GDP deflator). That is,

$$\text{Nominal GDP} = \text{Real GDP} \times \text{GDP Deflator}.$$

Nominal GDP measures the current dollar value of the output of the economy. Real GDP measures output valued at constant prices. The GDP deflator measures the price of output relative to its price in the base year. We can also write this equation as

$$\text{Real GDP} = \frac{\text{Nominal GDP}}{\text{GDP Deflator}}.$$

In this form, you can see how the deflator earns its name: it is used to deflate (that is, take inflation out of) nominal GDP to yield real GDP.

Chain-Weighted Measures of Real GDP

We have been discussing real GDP as if the prices used to compute this measure never change from their base-year values. If this were truly the case, over time the prices would become more and more dated. For instance, the price of computers has fallen substantially in recent years, while the price of a year at college has risen. When valuing the production of computers and education, it would be misleading to use the prices that prevailed ten or twenty years ago.

To solve this problem, the Bureau of Economic Analysis used to update periodically the prices used to compute real GDP. About every five years, a new base year was chosen. The prices were then held fixed and used to measure year-to-year changes in the production of goods and services until the base year was updated once again.

In 1995, the bureau announced a new policy for dealing with changes in the base year. In particular, it now emphasizes *chain-weighted* measures of real GDP. With these new measures, the base year changes continuously over time. In essence, average prices in 2001 and 2002 are used to measure real growth from 2001 to 2002; average prices in 2002 and 2003 are used to measure real growth from 2002 to 2003; and so on. These various year-to-year growth rates are then put together to form a "chain" that can be used to compare the output of goods and services between any two dates.

This new chain-weighted measure of real GDP is better than the more traditional measure because it ensures that the prices used to compute real GDP are never far out of date. For most purposes, however, the differences are not important. It turns out that the two measures of real GDP are highly correlated with each other. The reason for this close association is that most relative prices change slowly over time. Thus, both measures of real GDP track the same thing: economy-wide changes in the production of goods and services.

FYI

Two Arithmetic Tricks for Working With Percentage Changes

For manipulating many relationships in economics, there is an arithmetic trick that is useful to know: the percentage change of a product of two variables is approximately the sum of the percentage changes in each of the variables.

To see how this trick works, consider an example. Let P denote the GDP deflator and Y denote real GDP. Nominal GDP is $P \times Y$. The trick states that

$$\begin{aligned}\text{Percentage Change in } (P \times Y) \\ \approx (\text{Percentage Change in } P) \\ + (\text{Percentage Change in } Y).\end{aligned}$$

For instance, suppose that in one year, real GDP is 100 and the GDP deflator is 2; the next year, real GDP is 103 and the GDP deflator is 2.1. We can calculate that real GDP rose by 3 percent and that the GDP deflator rose by 5 percent. Nominal GDP rose from 200 the first year to 216.3 the second year, an increase of 8.15 percent. Notice that the growth in nominal GDP (8.15 percent) is approximately the sum of the growth in the GDP deflator (5 percent) and the growth in real GDP (3 percent).[1]

A second arithmetic trick follows as a corollary to the first: the percentage change of a ratio is approximately the percentage change in the numerator minus the percentage change in the denominator. Again, consider an example. Let Y denote GDP and L denote the population, so that Y/L is GDP per person. The second trick states

$$\begin{aligned}\text{Percentage Change in } (Y/L) \\ \approx (\text{Percentage Change in } Y) \\ - (\text{Percentage Change in } L).\end{aligned}$$

For instance, suppose that in the first year, Y is 100,000 and L is 100, so Y/L is 1,000; in the second year, Y is 110,000 and L is 103, so Y/L is 1,068. Notice that the growth in GDP per person (6.8 percent) is approximately the growth in income (10 percent) minus the growth in population (3 percent).

The Components of Expenditure

Economists and policymakers care not only about the economy's total output of goods and services but also about the allocation of this output among alternative uses. The national income accounts divide GDP into four broad categories of spending:

- Consumption (C)
- Investment (I)
- Government purchases (G)
- Net exports (NX).

Thus, letting Y stand for GDP,

$$Y = C + I + G + NX.$$

---

[1] Mathematical note: The proof that this trick works begins with the chain rule from calculus:

$$d(PY) = Y\,dP + P\,dY.$$

Now divide both sides of this equation by PY to obtain

$$d(PY)/(PY) = dP/P + dY/Y.$$

Notice that all three terms in this equation are percentage changes.

FYI

What Is Investment?

Newcomers to macroeconomics are sometimes confused by how macroeconomists use familiar words in new and specific ways. One example is the term "investment." The confusion arises because what looks like investment for an individual may not be investment for the economy as a whole. The general rule is that the economy's investment does not include purchases that merely reallocate existing assets among different individuals. Investment, as macroeconomists use the term, creates new capital.

Let's consider some examples. Suppose we observe these two events:

- Smith buys for himself a 100-year-old Victorian house.
- Jones builds for herself a brand-new contemporary house.

What is total investment here? Two houses, one house, or zero?

A macroeconomist seeing these two transactions counts only the Jones house as investment. Smith's transaction has not created new housing for the economy; it has merely reallocated existing housing. Smith's purchase is investment for Smith, but it is disinvestment for the person selling the house. By contrast, Jones has added new housing to the economy; her new house is counted as investment.

Similarly, consider these two events:

- Gates buys $5 million in IBM stock from Buffett on the New York Stock Exchange.
- General Motors sells $10 million in stock to the public and uses the proceeds to build a new car factory.

Here, investment is $10 million. In the first transaction, Gates is investing in IBM stock, and Buffett is disinvesting; there is no investment for the economy. By contrast, General Motors is using some of the economy's output of goods and services to add to its stock of capital; hence, its new factory is counted as investment.

GDP is the sum of consumption, investment, government purchases, and net exports. Each dollar of GDP falls into one of these categories. This equation is an identity—an equation that must hold because of the way the variables are defined. It is called the **national income accounts identity**.

Consumption consists of the goods and services bought by households. It is divided into three subcategories: nondurable goods, durable goods, and services. Nondurable goods are goods that last only a short time, such as food and clothing. Durable goods are goods that last a long time, such as cars and TVs. Services include the work done for consumers by individuals and firms, such as haircuts and doctor visits.

Investment consists of goods bought for future use. Investment is also divided into three subcategories: **business fixed investment**, **residential investment**, and **inventory investment**. Business fixed investment is the purchase of new plant and equipment by firms. Residential investment is the purchase of new housing by households and landlords. Inventory investment is the increase in firms' inventories of goods (if inventories are falling, inventory investment is negative).

Government purchases are the goods and services bought by federal, state, and local governments. This category includes such items as military equipment, highways, and the services that government workers provide. It does not include

transfer payments to individuals, such as Social Security and welfare. Because transfer payments reallocate existing income and are not made in exchange for goods and services, they are not part of GDP.

The last category, *net exports*, takes into account trade with other countries. Net exports are the value of goods and services exported to other countries minus the value of goods and services that foreigners provide us. Net exports represent the net expenditure from abroad on our goods and services, which provides income for domestic producers.

CASE STUDY

GDP and Its Components

In 2000 the GDP of the United States totaled about $10 trillion. This number is so large that it is almost impossible to comprehend. We can make it easier to understand by dividing it by the 2000 U.S. population of 275 million. In this way, we obtain GDP per person—the amount of expenditure for the average American—which equaled $36,174 in 2000.

table 2-1

GDP and the Components of Expenditure: 2000

	Total (billions of dollars)	Per Person (dollars)
Gross Domestic Product	9,963.1	36,174
Consumption	6,757.3	24,534
Nondurable goods	2,010.0	7,298
Durable goods	820.3	2,978
Services	3,927.0	14,258
Investment	1,832.7	6,654
Nonresidential fixed investment	1,362.2	4,946
Residential fixed investment	416.0	1,510
Inventory investment	54.5	198
Government Purchases	1,743.7	6,331
Federal	595.2	2,161
Defense	377.0	1,369
Nondefense	218.2	792
State and local	1,148.6	4,170
Net Exports	–370.7	–1,346
Exports	1,097.3	3,984
Imports	1,468.0	5,330

Source: U.S. Department of Commerce.

How did this GDP get used? Table 2-1 shows that about two-thirds of it, or $24,534 per person, was spent on consumption. Investment was $6,654 per person. Government purchases were $6,331 per person, $1,369 of which was spent by the federal government on national defense.

The average American bought $5,330 of goods imported from abroad and produced $3,984 of goods that were exported to other countries. Because the average American imported more than he exported, net exports were negative. Furthermore, because the average American earned less from selling to foreigners than he spent on foreign goods, the difference must have been financed by taking out loans from foreigners (or, equivalently, by selling them some assets). Thus, the average American borrowed $1,346 from abroad in 2000.

Other Measures of Income

The national income accounts include other measures of income that differ slightly in definition from GDP. It is important to be aware of the various measures, because economists and the press often refer to them.

To see how the alternative measures of income relate to one another, we start with GDP and add or subtract various quantities. To obtain gross national product (GNP), we add receipts of factor income (wages, profit, and rent) from the rest of the world and subtract payments of factor income to the rest of the world:

$$\text{GNP} = \text{GDP} + \text{Factor Payments From Abroad} - \text{Factor Payments to Abroad}.$$

Whereas GDP measures the total income produced domestically, GNP measures the total income earned by nationals (residents of a nation). For instance, if a Japanese resident owns an apartment building in New York, the rental income he earns is part of U.S. GDP because it is earned in the United States. But because this rental income is a factor payment to abroad, it is not part of U.S. GNP. In the United States, factor payments from abroad and factor payments to abroad are similar in size—each representing about 3 percent of GDP—so GDP and GNP are quite close.

To obtain net national product (NNP), we subtract the depreciation of capital—the amount of the economy's stock of plants, equipment, and residential structures that wears out during the year:

$$\text{NNP} = \text{GNP} - \text{Depreciation}.$$

In the national income accounts, depreciation is called the consumption of fixed capital. It equals about 10 percent of GNP. Because the depreciation of capital is a cost of producing the output of the economy, subtracting depreciation shows the net result of economic activity.

The next adjustment in the national income accounts is for indirect business taxes, such as sales taxes. These taxes, which make up about 10 percent of NNP, place a wedge between the price that consumers pay for a good and the price

that firms receive. Because firms never receive this tax wedge, it is not part of their income. Once we subtract indirect business taxes from NNP, we obtain a measure called national income

National Income = NNP − Indirect Business Taxes.

National income measures how much everyone in the economy has earned.

The national income accounts divide national income into five components, depending on the way the income is earned. The categories, and the percentage of national income paid in each category, are

- Compensation of employees (70%). The wages and fringe benefits earned by workers.
- Proprietors' income (9%). The income of noncorporate businesses, such as small farms, mom-and-pop stores, and law partnerships.
- Rental income (2%). The income that landlords receive, including the imputed rent that homeowners pay themselves, less expenses, such as depreciation.
- Corporate profits (12%). The income of corporations after payments to their workers and creditors.
- Net interest (7%). The interest domestic businesses pay minus the interest they receive, plus interest earned from foreigners.

A series of adjustments takes us from national income to personal income, the amount of income that households and noncorporate businesses receive. Three of these adjustments are most important. First, we reduce national income by the amount that corporations earn but do not pay out, either because the corporations are retaining earnings or because they are paying taxes to the government. This adjustment is made by subtracting corporate profits (which equals the sum of corporate taxes, dividends, and retained earnings) and adding back dividends. Second, we increase national income by the net amount the government pays out in transfer payments. This adjustment equals government transfers to individuals minus social insurance contributions paid to the government. Third, we adjust national income to include the interest that households earn rather than the interest that businesses pay. This adjustment is made by adding personal interest income and subtracting net interest. (The difference between personal interest and net interest arises in part from the interest on the government debt.) Thus, personal income is

Personal Income = National Income
− Corporate Profits
− Social Insurance Contributions
− Net Interest
+ Dividends
+ Government Transfers to Individuals
+ Personal Interest Income.

Next, if we subtract personal tax payments and certain nontax payments to the government (such as parking tickets), we obtain disposable personal income:

$$\text{Disposable Personal Income} = \text{Personal Income} - \text{Personal Tax and Nontax Payments}.$$

We are interested in disposable personal income because it is the amount households and noncorporate businesses have available to spend after satisfying their tax obligations to the government.

CASE STUDY

The Seasonal Cycle and Seasonal Adjustment

Because real GDP and the other measures of income tell us how well the economy is performing, economists are interested in studying the quarter-to-quarter fluctuations in these variables. Yet when we start to do so, one fact leaps out: all these measures of income exhibit a regular seasonal pattern. The output of the economy rises during the year, reaching a peak in the fourth quarter (October, November, and December), and then falling in the first quarter (January, February, and March) of the next year. These regular seasonal changes are substantial. From the fourth quarter to the first quarter, real GDP falls on average about 8 percent.[2]

It is not surprising that real GDP follows a seasonal cycle. Some of these changes are attributable to changes in our ability to produce: for example, building homes is more difficult during the cold weather of winter than during other seasons. In addition, people have seasonal tastes: they have preferred times for such activities as vacations and holiday shopping.

When economists study fluctuations in real GDP and other economic variables, they often want to eliminate the portion of fluctuations caused by predictable seasonal changes. You will find that most of the economic statistics reported in the newspaper are seasonally adjusted. This means that the data have been adjusted to remove the regular seasonal fluctuations. (The precise statistical procedures used are too elaborate to bother with here, but in essence they involve subtracting those changes in income that are predictable simply from the change in season.) Therefore, when you observe a rise or fall in real GDP or any other data series, you must look beyond the seasonal cycle for the explanation.

[2] Robert B. Barsky and Jeffrey A. Miron, "The Seasonal Cycle and the Business Cycle," *Journal of Political Economy* 97 (June 1989): 503-534.

2-2 Measuring the Cost of Living: The Consumer Price Index

A dollar today doesn't buy as much as it did 20 years ago. The cost of almost everything has gone up. This increase in the overall level of prices is called inflation, and it is one of the primary concerns of economists and policymakers. In later chapters we examine in detail the causes and effects of inflation. Here we discuss how economists measure changes in the cost of living.

The Price of a Basket of Goods

The most commonly used measure of the level of prices is the consumer price index (CPI). The Bureau of Labor Statistics, which is part of the U.S. Department of Labor, has the job of computing the CPI. It begins by collecting the prices of thousands of goods and services. Just as GDP turns the quantities of many goods and services into a single number measuring the value of production, the CPI turns the prices of many goods and services into a single index measuring the overall level of prices.

How should economists aggregate the many prices in the economy into a single index that reliably measures the price level? They could simply compute an average of all prices. Yet this approach would treat all goods and services equally. Because people buy more chicken than caviar, the price of chicken should have a greater weight in the CPI than the price of caviar. The Bureau of Labor Statistics weights different items by computing the price of a basket of goods and services purchased by a typical consumer. The CPI is the price of this basket of goods and services relative to the price of the same basket in some base year.

For example, suppose that the typical consumer buys 5 apples and 2 oranges every month. Then the basket of goods consists of 5 apples and 2 oranges, and the CPI is

$$\text{CPI} = \frac{(5 \times \text{Current Price of Apples}) + (2 \times \text{Current Price of Oranges})}{(5 \times \text{2002 Price of Apples}) + (2 \times \text{2002 Price of Oranges})}.$$

In this CPI, 2002 is the base year. The index tells us how much it costs now to buy 5 apples and 2 oranges relative to how much it cost to buy the same basket of fruit in 2002.

The consumer price index is the most closely watched index of prices, but it is not the only such index. Another is the producer price index, which measures the price of a typical basket of goods bought by firms rather than consumers. In addition to these overall price indices, the Bureau of Labor Statistics computes price indices for specific types of goods, such as food, housing, and energy.

The CPI Versus the GDP Deflator

Earlier in this chapter we saw another measure of prices: the implicit price deflator for GDP, which is the ratio of nominal GDP to real GDP. The GDP deflator and the CPI give somewhat different information about what is happening to the overall level of prices in the economy. There are three key differences between the two measures.

The first difference is that the GDP deflator measures the prices of all goods and services produced, whereas the CPI measures the prices of only the goods and services bought by consumers. Thus, an increase in the price of goods bought by firms or the government will show up in the GDP deflator but not in the CPI.

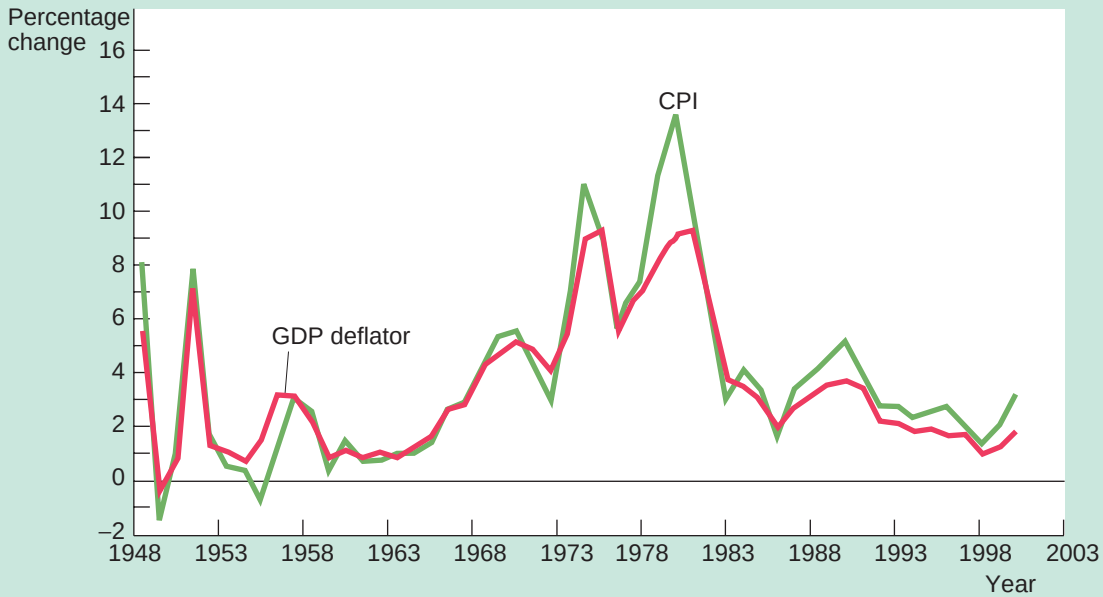
The second difference is that the GDP deflator includes only those goods produced domestically. Imported goods are not part of GDP and do not show up in the GDP deflator. Hence, an increase in the price of a Toyota made in Japan and sold in this country affects the CPI, because the Toyota is bought by consumers, but it does not affect the GDP deflator.

The third and most subtle difference results from the way the two measures aggregate the many prices in the economy. The CPI assigns fixed weights to the prices of different goods, whereas the GDP deflator assigns changing weights. In other words, the CPI is computed using a fixed basket of goods, whereas the GDP deflator allows the basket of goods to change over time as the composition of GDP changes. The following example shows how these approaches differ. Suppose that major frosts destroy the entire orange crop. The quantity of oranges produced falls to zero, and the price of the few oranges that remain on grocers' shelves is driven sky-high. Because oranges are no longer part of GDP, the increase in the price of oranges does not show up in the GDP deflator. But because the CPI is computed with a fixed basket of goods that includes oranges, the increase in the price of oranges causes a substantial rise in the CPI.

Economists call a price index with a fixed basket of goods a *Laspeyres index* and a price index with a changing basket a *Paasche index*. Economic theorists have studied the properties of these different types of price indices to determine which is a better measure of the cost of living. The answer, it turns out, is that neither is clearly superior. When prices of different goods are changing by different amounts, a Laspeyres (fixed basket) index tends to overstate the increase in the cost of living because it does not take into account that consumers have the opportunity to substitute less expensive goods for more expensive ones. By contrast, a Paasche (changing basket) index tends to understate the increase in the cost of living. Although it accounts for the substitution of alternative goods, it does not reflect the reduction in consumers' welfare that may result from such substitutions.

The example of the destroyed orange crop shows the problems with Laspeyres and Paasche price indices. Because the CPI is a Laspeyres index, it overstates the impact of the increase in orange prices on consumers: by using a fixed basket of goods, it ignores consumers' ability to substitute apples for oranges. By contrast,

figure 2-3



The GDP Deflator and the CPI This figure shows the percentage change in the GDP deflator and in the CPI for every year since 1948. Although these two measures of prices diverge at times, they usually tell the same story about how quickly prices are rising. Both the CPI and the GDP deflator show that prices rose slowly in most of the 1950s and 1960s, that they rose much more quickly in the 1970s, and that they rose slowly again in the 1980s and 1990s.

Source: U.S. Department of Commerce, U.S. Department of Labor.

because the GDP deflator is a Paasche index, it understates the impact on consumers: the GDP deflator shows no rise in prices, yet surely the higher price of oranges makes consumers worse off.

Luckily, the difference between the GDP deflator and the CPI is usually not large in practice. Figure 2-3 shows the percentage change in the GDP deflator and the percentage change in the CPI for each year since 1948. Both measures usually tell the same story about how quickly prices are rising.

CASE STUDY

Does the CPI Overstate Inflation?

The consumer price index is a closely watched measure of inflation. Policymakers in the Federal Reserve monitor the CPI when choosing monetary policy. In addition, many laws and private contracts have cost-of-living allowances, called COLAs, which use the CPI to adjust for changes in the price level. For instance,

Social Security benefits are adjusted automatically every year so that inflation will not erode the living standard of the elderly.

Because so much depends on the CPI, it is important to ensure that this measure of the price level is accurate. Many economists believe that, for a number of reasons, the CPI tends to overstate inflation.

One problem is the substitution bias we have already discussed. Because the CPI measures the price of a fixed basket of goods, it does not reflect the ability of consumers to substitute toward goods whose relative prices have fallen. Thus, when relative prices change, the true cost of living rises less rapidly than the CPI.

A second problem is the introduction of new goods. When a new good is introduced into the marketplace, consumers are better off, because they have more products from which to choose. In effect, the introduction of new goods increases the real value of the dollar. Yet this increase in the purchasing power of the dollar is not reflected in a lower CPI.

A third problem is unmeasured changes in quality. When a firm changes the quality of a good it sells, not all of the good's price change reflects a change in the cost of living. The Bureau of Economic Analysis does its best to account for changes in the quality of goods over time. For example, if Ford increases the horsepower of a particular car model from one year to the next, the CPI will reflect the change: the quality-adjusted price of the car will not rise as fast as the unadjusted price. Yet many changes in quality, such as comfort or safety, are hard to measure. If unmeasured quality improvement (rather than unmeasured quality deterioration) is typical, then the measured CPI rises faster than it should.

Because of these measurement problems, some economists have suggested revising laws to reduce the degree of indexation. For example, Social Security benefits could be indexed to CPI inflation minus 1 percent. Such a change would provide a rough way of offsetting these measurement problems. At the same time, it would automatically slow the growth in government spending.

In 1995, the Senate Finance Committee appointed a panel of five noted economists—Michael Boskin, Ellen Dulberger, Robert Gordon, Zvi Griliches, and Dale Jorgenson—to study the magnitude of the measurement error in the CPI. The panel concluded that the CPI was biased upward by 0.8 to 1.6 percentage points per year, with the best estimate being 1.1 percentage points. This report led to some changes in the way the CPI is calculated, so the bias is now thought to be slightly under 1 percentage point. The CPI still overstates inflation, but not by as much as it once did.[3]

[3] For further discussion of these issues, see Matthew Shapiro and David Wilcox, "Mismeasurement in the Consumer Price Index: An Evaluation," *NBER Macroeconomics Annual 1996*, and the symposium on "Measuring the CPI" in the Winter 1998 issue of *The Journal of Economic Perspectives*.

2-3 Measuring Joblessness: The Unemployment Rate

One aspect of economic performance is how well an economy uses its resources. Because an economy's workers are its chief resource, keeping workers employed is a paramount concern of economic policymakers. The unemployment rate is the statistic that measures the percentage of those people wanting to work who do not have jobs.

Drawing M. Stevens; © 1980 The New Yorker Magazine, Inc.

"Well, so long Eddie, the recession's over."

Every month the U.S. Bureau of Labor Statistics computes the unemployment rate and many other statistics that economists and policymakers use to monitor developments in the labor market. These statistics come from a survey of about 60,000 households. Based on the responses to survey questions, each adult (16 years and older) in each household is placed into one of three categories: employed, unemployed, or not in the labor force. A person is employed if he or she spent some of the previous week working at a paid job. A person is unemployed if he or she is not employed and has been looking for a job or is on temporary layoff. A person who fits into neither of the first two categories, such as a full-time student or retiree, is not in the labor force. A person who wants a job but has given up looking is a discouraged worker and is counted as not being in the labor force.

The labor force is defined as the sum of the employed and unemployed, and the unemployment rate is defined as the percentage of the labor force that is unemployed. That is,

$$\text{Labor Force} = \text{Number of Employed} + \text{Number of Unemployed},$$

and

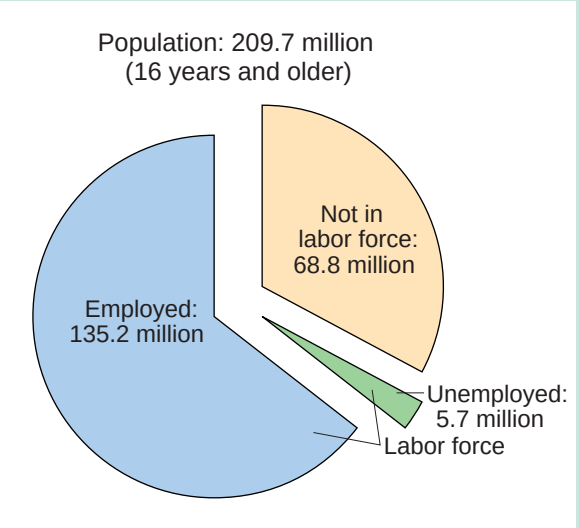
$$\text{Unemployment Rate} = \frac{\text{Number of Unemployed}}{\text{Labor Force}} \times 100.$$

A related statistic is the **labor-force participation rate**, the percentage of the adult population that is in the labor force:

$$\text{Labor-Force Participation Rate} = \frac{\text{Labor Force}}{\text{Adult Population}} \times 100.$$

The Bureau of Labor Statistics computes these statistics for the overall population and for groups within the population: men and women, whites and blacks, teenagers and prime-age workers.

figure 2-4



The Three Groups of the Population When the Bureau of Labor Statistics surveys the population, it places all adults into one of three categories: employed, unemployed, or not in the labor force. This figure shows the number of people in each category in 2000.

Source: U.S. Department of Labor.

Figure 2-4 shows the breakdown of the population into the three categories for 2000. The statistics broke down as follows:

$$\text{Labor Force} = 135.2 + 5.7 = 140.9 \text{ million.}$$
$$\text{Unemployment Rate} = (5.7/140.9) \times 100 = 4.0\%.$$
$$\text{Labor-Force Participation Rate} = (140.9/209.7) \times 100 = 67.2\%.$$

Hence, about two-thirds of the adult population was in the labor force, and about 4 percent of those in the labor force did not have a job.

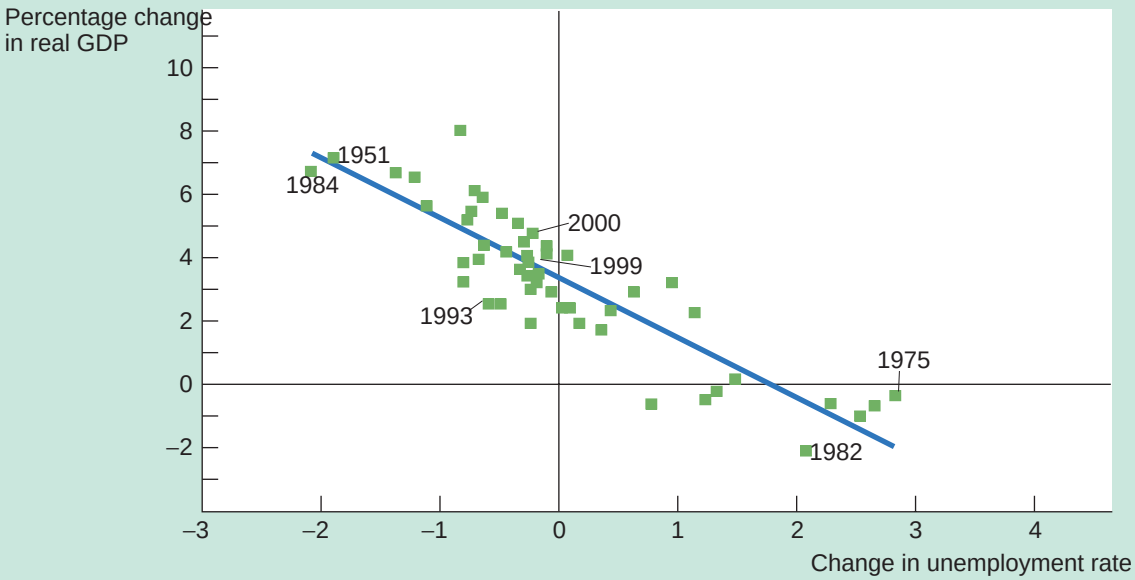
CASE STUDY

Unemployment, GDP, and Okun's Law

What relationship should we expect to find between unemployment and real GDP? Because employed workers help to produce goods and services and unemployed workers do not, increases in the unemployment rate should be associated with decreases in real GDP. This negative relationship between unemployment and GDP is called Okun's law, after Arthur Okun, the economist who first studied it.[4]

[4] Arthur M. Okun, "Potential GNP: Its Measurement and Significance," Proceedings of the Business and Economics Statistics Section, American Statistical Association (Washington, DC: American Statistical Association, 1962), 98-103; reprinted in Arthur M. Okun, *Economics for Policymaking* (Cambridge, MA: MIT Press, 1983), 145-158.

figure 2-5



Okun's Law This figure is a scatterplot of the change in the unemployment rate on the horizontal axis and the percentage change in real GDP on the vertical axis, using data on the U.S economy. Each point represents one year. The negative correlation between these variables shows that increases in unemployment tend to be associated with lower-than-normal growth in real GDP.

Source: U.S. Department of Commerce, U.S. Department of Labor.

Figure 2-5 uses annual data for the United States to illustrate Okun's Law. This figure is a scatterplot—a scatter of points where each point represents one observation (in this case, the data for a particular year). The horizontal axis represents the change in the unemployment rate from the previous year, and the vertical axis represents the percentage change in GDP. This figure shows clearly that year-to-year changes in the unemployment rate are closely associated with year-to-year changes in real GDP.

We can be more precise about the magnitude of the Okun's Law relationship. The line drawn through the scatter of points (estimated with a statistical procedure called ordinary least squares) tells us that

$$\text{Percentage Change in Real GDP} = 3\% - 2 \times \text{Change in the Unemployment Rate}.$$

If the unemployment rate remains the same, real GDP grows by about 3 percent; this normal growth in the production of goods and services is a result of growth in the labor force, capital accumulation, and technological progress. In addition, for every percentage point the unemployment rate rises, real GDP growth typi-

cally falls by 2 percent. Hence, if the unemployment rate rises from 6 to 8 percent, then real GDP growth would be

$$\begin{aligned}\text{Percentage Change in Real GDP} &= 3\% - 2 \times (8\% - 6\%) \\ &= -1\%.\end{aligned}$$

In this case, Okun's law says that GDP would fall by 1 percent, indicating that the economy is in a recession.

2-4 Conclusion: From Economic Statistics to Economic Models

The three statistics discussed in this chapter—gross domestic product, the consumer price index, and the unemployment rate—quantify the performance of the economy. Public and private decisionmakers use these statistics to monitor changes in the economy and to formulate appropriate policies. Economists use these statistics to develop and test theories about how the economy works.

In the chapters that follow, we examine some of these theories. That is, we build models that explain how these variables are determined and how economic policy affects them. Having learned how to measure economic performance, we are now ready to learn how to explain it.

Summary

1. Gross domestic product (GDP) measures both the income of everyone in the economy and the total expenditure on the economy's output of goods and services.
2. Nominal GDP values goods and services at current prices. Real GDP values goods and services at constant prices. Real GDP rises only when the amount of goods and services has increased, whereas nominal GDP can rise either because output has increased or because prices have increased.
3. GDP is the sum of four categories of expenditure: consumption, investment, government purchases, and net exports.
4. The consumer price index (CPI) measures the price of a fixed basket of goods and services purchased by a typical consumer. Like the GDP deflator, which is the ratio of nominal GDP to real GDP, the CPI measures the overall level of prices.
5. The unemployment rate shows what fraction of those who would like to work do not have a job. When the unemployment rate rises, real GDP typically grows slower than its normal rate and may even fall.

KEY CONCEPTS

Gross domestic product (GDP)
Consumer price index (CPI)
Unemployment rate
National income accounting
Stocks and flows
Value added
Imputed value
Nominal versus real GDP
GDP deflator
National income accounts identity
Consumption
Investment
Government purchases
Net exports
Labor force
Labor-force participation rate
Okun's law

QUESTIONS FOR REVIEW

1. List the two things that GDP measures. How can GDP measure two things at once?
2. What does the consumer price index measure?
3. List the three categories used by the Bureau of Labor Statistics to classify everyone in the economy. How does the bureau compute the unemployment rate?
4. Explain Okun's law.

PROBLEMS AND APPLICATIONS

1. Look at the newspapers for the past few days. What new economic statistics have been released? How do you interpret these statistics?
2. A farmer grows a bushel of wheat and sells it to a miller for $1.00. The miller turns the wheat into flour and then sells the flour to a baker for $3.00. The baker uses the flour to make bread and sells the bread to an engineer for $6.00. The engineer eats the bread. What is the value added by each person? What is GDP?
3. Suppose a woman marries her butler. After they are married, her husband continues to wait on her as before, and she continues to support him as before (but as a husband rather than as an employee). How does the marriage affect GDP? How should it affect GDP?
4. Place each of the following transactions in one of the four components of expenditure: consumption, investment, government purchases, and net exports.
   a. Boeing sells an airplane to the Air Force.
   b. Boeing sells an airplane to American Airlines.
   c. Boeing sells an airplane to Air France.
   d. Boeing sells an airplane to Amelia Earhart.
   e. Boeing builds an airplane to be sold next year.
5. Find data on GDP and its components, and compute the percentage of GDP for the following components for 1950, 1975, and 2000.
   a. Personal consumption expenditures
   b. Gross private domestic investment
   c. Government purchases
   d. Net exports
   e. National defense purchases
   f. State and local purchases
   g. Imports

   Do you see any stable relationships in the data? Do you see any trends? (Hint: A good place to look for data is the statistical appendices of the Economic Report of the President, which is written each year by the Council of Economic Advisers. Alternatively, you can go to www.bea.doc.gov, which is the Web site of the Bureau of Economic Analysis.)
6. Consider an economy that produces and consumes bread and automobiles. In the following table are data for two different years.

	Year 2000	Year 2010
Price of an automobile	$50,000	$60,000
Price of a loaf of bread	$10	$20
Number of automobiles produced	100	120
Number of loaves of bread produced	500,000	400,000

a. Using the year 2000 as the base year, compute the following statistics for each year: nominal GDP, real GDP, the implicit price deflator for GDP, and a fixed-weight price index such as the CPI.

b. How much have prices risen between year 2000 and year 2010? Compare the answers given by the Laspeyres and Paasche price indices. Explain the difference.

c. Suppose you are a senator writing a bill to index Social Security and federal pensions. That is, your bill will adjust these benefits to offset changes in the cost of living. Will you use the GDP deflator or the CPI? Why?

7. Abby consumes only apples. In year 1, red apples cost $1 each, green apples cost $2 each, and Abby buys 10 red apples. In year 2, red apples cost $2, green apples cost $1, and Abby buys 10 green apples.

a. Compute a consumer price index for apples for each year. Assume that year 1 is the base year in which the consumer basket is fixed. How does your index change from year 1 to year 2?

b. Compute Abby's nominal spending on apples in each year. How does it change from year 1 to year 2?

c. Using year 1 as the base year, compute Abby's real spending on apples in each year. How does it change from year 1 to year 2?

d. Defining the implicit price deflator as nominal spending divided by real spending, compute the deflator for each year. How does the deflator change from year 1 to year 2?

e. Suppose that Abby is equally happy eating red or green apples. How much has the true cost of living increased for Abby? Compare this answer to your answers to parts (a) and (d). What does this example tell you about Laspeyres and Paasche price indexes?

8. Consider how each of the following events is likely to affect real GDP. Do you think the change in real GDP reflects a similar change in economic well-being?

a. A hurricane in Florida forces Disney World to shut down for a month.

b. The discovery of a new, easy-to-grow strain of wheat increases farm harvests.

c. Increased hostility between unions and management sparks a rash of strikes.

d. Firms throughout the economy experience falling demand, causing them to lay off workers.

e. Congress passes new environmental laws that prohibit firms from using production methods that emit large quantities of pollution.

f. More high-school students drop out of school to take jobs mowing lawns.

g. Fathers around the country reduce their workweeks to spend more time with their children.

9. In a speech that Senator Robert Kennedy gave when he was running for president in 1968, he said the following about GDP:

> [It] does not allow for the health of our children, the quality of their education, or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages, the intelligence of our public debate or the integrity of our public officials. It measures neither our courage, nor our wisdom, nor our devotion to our country. It measures everything, in short, except that which makes life worthwhile, and it can tell us everything about America except why we are proud that we are Americans.

Was Robert Kennedy right? If so, why do we care about GDP?

part II

Classical Theory: The Economy in the Long Run

CHAPTER 3 THREE

National Income: Where It Comes From and Where It Goes

> A large income is the best recipe for happiness I ever heard of.
>
> — Jane Austen

The most important macroeconomic variable is gross domestic product (GDP). As we have seen, GDP measures both a nation's total output of goods and services and its total income. To appreciate the significance of GDP, one need only take a quick look at international data: compared with their poorer counterparts, nations with a high level of GDP per person have everything from better childhood nutrition to more televisions per household. A large GDP does not ensure that all of a nation's citizens are happy, but it may be the best recipe for happiness that macroeconomists have to offer.

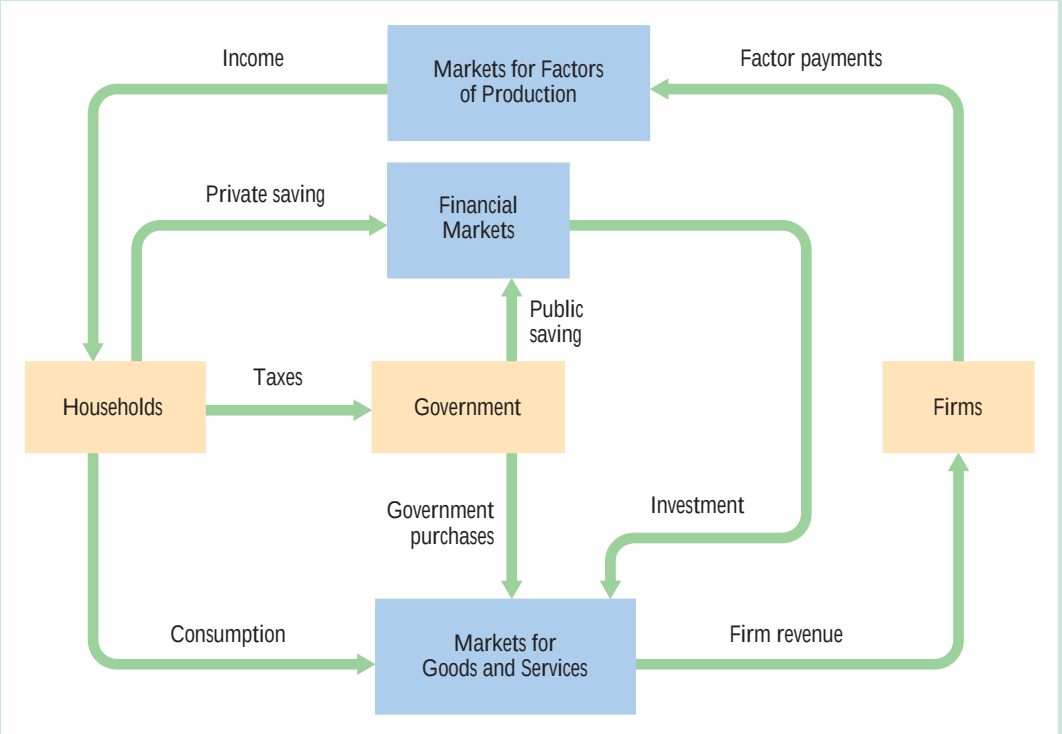
This chapter addresses four groups of questions about the sources and uses of a nation's GDP:

- How much do the firms in the economy produce? What determines a nation's total income?
- Who gets the income from production? How much goes to compensate workers, and how much goes to compensate owners of capital?
- Who buys the output of the economy? How much do households purchase for consumption, how much do households and firms purchase for investment, and how much does the government buy for public purposes?
- What equilibrates the demand for and supply of goods and services? What ensures that desired spending on consumption, investment, and government purchases equals the level of production?

To answer these questions, we must examine how the various parts of the economy interact.

A good place to start is the circular flow diagram. In Chapter 2 we traced the circular flow of dollars in a hypothetical economy that produced one product, bread, from labor services. Figure 3-1 more accurately depicts how real economies function. It shows the linkages among the economic actors—households, firms,

figure 3-1



The Circular Flow of Dollars Through the Economy This figure is a more realistic version of the circular flow diagram found in Chapter 2. Each yellow box represents an economic actor—households, firms, and the government. Each blue box represents a type of market—the markets for goods and services, the markets for the factors of production, and financial markets. The green arrows show the flow of dollars among the economic actors through the three types of markets.

and the government and how dollars flow among them through the various markets in the economy.

Let's look at the flow of dollars from the viewpoints of these economic actors. Households receive income and use it to pay taxes to the government, to consume goods and services, and to save through the financial markets. Firms receive revenue from the sale of goods and services and use it to pay for the factors of production. Both households and firms borrow in financial markets to buy investment goods, such as houses and factories. The government receives revenue from taxes and uses it to pay for government purchases. Any excess of tax revenue over government spending is called public saving, which can be either positive (a budget surplus) or negative (a budget deficit).

In this chapter we develop a basic classical model to explain the economic interactions depicted in Figure 3-1. We begin with firms and look at what

determines their level of production (and, thus, the level of national income). Then we examine how the markets for the factors of production distribute this income to households. Next, we consider how much of this income households consume and how much they save. In addition to discussing the demand for goods and services arising from the consumption of households, we discuss the demand arising from investment and government purchases. Finally, we come full circle and examine how the demand for goods and services (the sum of consumption, investment, and government purchases) and the supply of goods and services (the level of production) are brought into balance.

3-1 What Determines the Total Production of Goods and Services?

An economy's output of goods and services—its GDP—depends on (1) its quantity of inputs, called the factors of production, and (2) its ability to turn inputs into output, as represented by the production function. We discuss each of these in turn.

The Factors of Production

Factors of production are the inputs used to produce goods and services. The two most important factors of production are capital and labor. Capital is the set of tools that workers use: the construction worker's crane, the accountant's calculator, and this author's personal computer. Labor is the time people spend working. We use the symbol K to denote the amount of capital and the symbol L to denote the amount of labor.

In this chapter we take the economy's factors of production as given. In other words, we assume that the economy has a fixed amount of capital and a fixed amount of labor. We write

$$K = \bar{K}.$$
$$L = \bar{L}.$$

The overbar means that each variable is fixed at some level. In Chapter 7 we examine what happens when the factors of production change over time, as they do in the real world. For now, to keep our analysis simple, we assume fixed amounts of capital and labor.

We also assume here that the factors of production are fully utilized—that is, that no resources are wasted. Again, in the real world, part of the labor force is unemployed, and some capital lies idle. In Chapter 6 we examine the reasons for unemployment, but for now we assume that capital and labor are fully employed.

The Production Function

The available production technology determines how much output is produced from given amounts of capital and labor. Economists express the available technology using a **production function**. Letting Y denote the amount of output, we write the production function as

$$Y = F(K, L).$$

This equation states that output is a function of the amount of capital and the amount of labor.

The production function reflects the available technology for turning capital and labor into output. If someone invents a better way to produce a good, the result is more output from the same amounts of capital and labor. Thus, technological change alters the production function.

Many production functions have a property called **constant returns to scale**. A production function has constant returns to scale if an increase of an equal percentage in all factors of production causes an increase in output of the same percentage. If the production function has constant returns to scale, then we get 10 percent more output when we increase both capital and labor by 10 percent. Mathematically, a production function has constant returns to scale if

$$zY = F(zK, zL)$$

for any positive number z. This equation says that if we multiply both the amount of capital and the amount of labor by some number z, then output is also multiplied by z. In the next section we see that the assumption of constant returns to scale has an important implication for how the income from production is distributed.

As an example of a production function, consider production at a bakery. The kitchen and its equipment are the bakery's capital, the workers hired to make the bread are its labor, and the loaves of bread are its output. The bakery's production function shows that the number of loaves produced depends on the amount of equipment and the number of workers. If the production function has constant returns to scale, then doubling the amount of equipment and the number of workers doubles the amount of bread produced.

The Supply of Goods and Services

We can now see that the factors of production and the production function together determine the quantity of goods and services supplied, which in turn equals the economy's output. To express this mathematically, we write

$$\begin{aligned} Y &= F(\bar{K}, \bar{L}) \\ &= \bar{Y}. \end{aligned}$$

In this chapter, because we assume that the supplies of capital and labor and the technology are fixed, output is also fixed (at a level denoted here as $\bar{Y}$). When we

discuss economic growth in Chapters 7 and 8, we will examine how increases in capital and labor and improvements in the production technology lead to growth in the economy's output.

3-2 How Is National Income Distributed to the Factors of Production?

As we discussed in Chapter 2, the total output of an economy equals its total income. Because the factors of production and the production function together determine the total output of goods and services, they also determine national income. The circular-flow diagram in Figure 3-1 shows that this national income flows from firms to households through the markets for the factors of production.

In this section we continue developing our model of the economy by discussing how these factor markets work. Economists have long studied factor markets to understand the distribution of income. (For example, Karl Marx, the noted nineteenth-century economist, spent much time trying to explain the incomes of capital and labor. The political philosophy of communism was in part based on Marx's now-discredited theory.) Here we examine the modern theory of how national income is divided among the factors of production. This theory, called the *neoclassical theory of distribution*, is accepted by most economists today.

Factor Prices

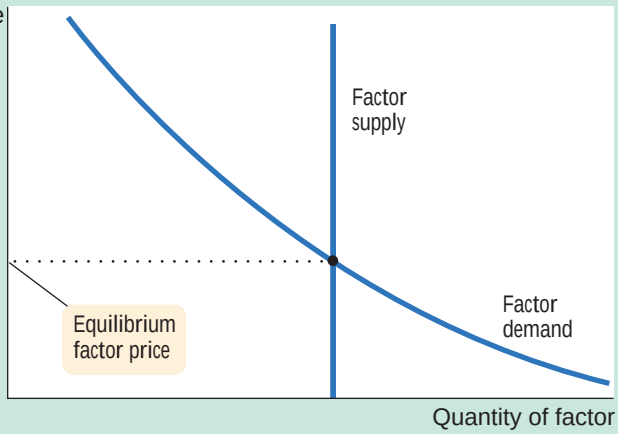
The distribution of national income is determined by factor prices. *Factor prices* are the amounts paid to the factors of production—the wage workers earn and the rent the owners of capital collect. As Figure 3-2 illustrates, the price each factor of production receives for its services is in turn determined by the supply and demand for that factor. Because we have assumed that the economy's factors of production are fixed, the factor supply curve in Figure 3-2 is vertical. The intersection of the downward-sloping factor demand curve and the vertical supply curve determines the equilibrium factor price.

To understand factor prices and the distribution of income, we must examine the demand for the factors of production. Because factor demand arises from the thousands of firms that use capital and labor, we now look at the decisions faced by a typical firm about how much of these factors to employ.

The Decisions Facing the Competitive Firm

The simplest assumption to make about a typical firm is that it is competitive. A *competitive firm* is small relative to the markets in which it trades, so it has little influence on market prices. For example, our firm produces a good and sells it at the market price. Because many firms produce this good, our firm can sell as

figure 3-2



How a Factor of Production Is Compensated The price paid to any factor of production depends on the supply and demand for that factor's services. Because we have assumed that supply is fixed, the supply curve is vertical. The demand curve is downward sloping. The intersection of supply and demand determines the equilibrium factor price.

much as it wants without causing the price of the good to fall, or it can stop selling altogether without causing the price of the good to rise. Similarly, your firm cannot influence the wages of the workers it employs because many other local firms also employ workers. The firm has no reason to pay more than the market wage, and if it tried to pay less, its workers would take jobs elsewhere. Therefore, the competitive firm takes the prices of its output and its inputs as given.

To make its product, the firm needs two factors of production, capital and labor. As we did for the aggregate economy, we represent the firm's production technology by the production function

$$Y = F(K, L),$$

where Y is the number of units produced (the firm's output), K the number of machines used (the amount of capital), and L the number of hours worked by the firm's employees (the amount of labor). The firm produces more output if it has more machines or if its employees work more hours.

The firm sells its output at a price P, hires workers at a wage W, and rents capital at a rate R. Notice that when we speak of firms renting capital, we are assuming that households own the economy's stock of capital. In this analysis, households rent out their capital, just as they sell their labor. The firm obtains both factors of production from the households that own them.[1]

The goal of the firm is to maximize profit. Profit is revenue minus costs. It is what the owners of the firm keep after paying for the costs of production. Revenue equals $P \times Y$, the selling price of the good, P, multiplied by the amount of

[1] This is a simplification. In the real world, the ownership of capital is indirect because firms own capital and households own the firms. That is, real firms have two functions: owning capital and producing output. To help us understand how the factors of production are compensated, however, we assume that firms only produce output and that households own capital directly.

the good the firm produces (Y). Costs include both labor costs and capital costs. Labor costs equal $W \times L$, the wage (W) times the amount of labor (L). Capital costs equal $R \times K$, the rental price of capital (R) times the amount of capital (K). We can write

$$\begin{aligned}\text{Profit} &= \text{Revenue} - \text{Labor Costs} - \text{Capital Costs} \\ &= PY - WL - RK.\end{aligned}$$

To see how profit depends on the factors of production, we use the production function $Y = F(K, L)$ to substitute for Y to obtain

$$\text{Profit} = PF(K, L) - WL - RK.$$

This equation shows that profit depends on the product price (P), the factor prices W and R, and the factor quantities (L and K). The competitive firm takes the product price and the factor prices as given and chooses the amounts of labor and capital that maximize profit.

The Firm's Demand for Factors

We now know that our firm will hire labor and rent capital in the quantities that maximize profit. But what are those profit-maximizing quantities? To answer this question, we first consider the quantity of labor and then the quantity of capital.

The Marginal Product of Labor The more labor the firm employs, the more output it produces. The **marginal product of labor** (*MPL*) is the extra amount of output the firm gets from one extra unit of labor, holding the amount of capital fixed. We can express this using the production function:

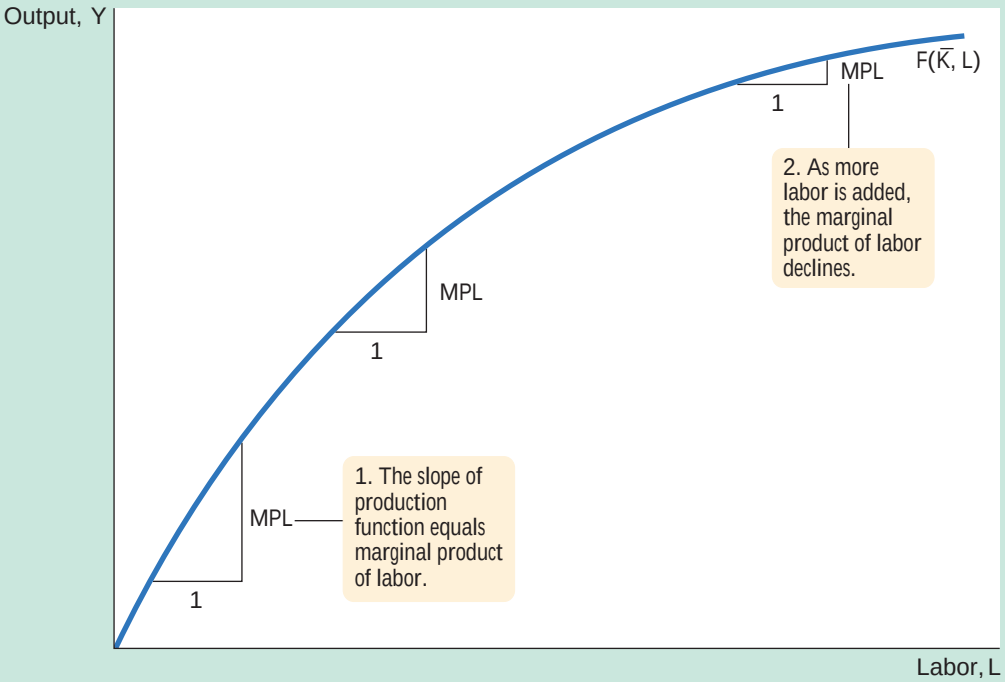
$$MPL = F(K, L + 1) - F(K, L).$$

The first term on the right-hand side is the amount of output produced with K units of capital and $L + 1$ units of labor; the second term is the amount of output produced with K units of capital and L units of labor. This equation states that the marginal product of labor is the difference between the amount of output produced with $L + 1$ units of labor and the amount produced with only L units of labor.

Most production functions have the property of **diminishing marginal product**: holding the amount of capital fixed, the marginal product of labor decreases as the amount of labor increases. Consider again the production of bread at a bakery. As a bakery hires more labor, it produces more bread. That is, *MPL* is the amount of extra bread produced when an extra unit of labor is hired. As more labor is added to a fixed amount of capital, however, the *MPL* falls. Fewer additional loaves are produced because workers are less productive when the kitchen is more crowded. In other words, holding the size of the kitchen fixed, each additional worker adds fewer loaves of bread to the bakery's output.

Figure 3-3 graphs the production function. It illustrates what happens to the amount of output when we hold the amount of capital constant and vary the

figure 3-3



The Production Function This curve shows how output depends on labor input, holding the amount of capital constant. The marginal product of labor, MPL, is the change in output when the labor input is increased by 1 unit. As the amount of labor increases, the production function becomes flatter, indicating diminishing marginal product.

amount of labor. This figure shows that the marginal product of labor is the slope of the production function. As the amount of labor increases, the production function becomes flatter, indicating diminishing marginal product.

From the Marginal Product of Labor to Labor Demand When the competitive, profit-maximizing firm is deciding whether to hire an additional unit of labor, it considers how that decision would affect profit. It therefore compares the extra revenue from the increased production that results from the added labor to the extra cost of higher spending on wages. The increase in revenue from an additional unit of labor depends on two variables: the marginal product of labor and the price of the output. Because an extra unit of labor produces MPL units of output and each unit of output sells for P dollars, the extra revenue is P × MPL. The extra cost of hiring one more unit of labor is the wage, W. Thus, the change in profit from hiring an additional unit of labor is

$$\begin{aligned} \Delta \text{Profit} &= \Delta\text{Revenue} - \Delta\text{Cost} \\ &= (P \times \text{MPL}) - W. \end{aligned}$$

The symbol Δ (called delta) denotes the change in a variable.

We can now answer the question we asked at the beginning of this section: How much labor does the firm hire? The firm's manager knows that if the extra revenue $P \times MPL$ exceeds the wage W, an extra unit of labor increases profit. Therefore, the manager continues to hire labor until the next unit would no longer be profitable—that is, until the MPL falls to the point where the extra revenue equals the wage. The firm's demand for labor is determined by

$$P \times MPL = W.$$

We can also write this as

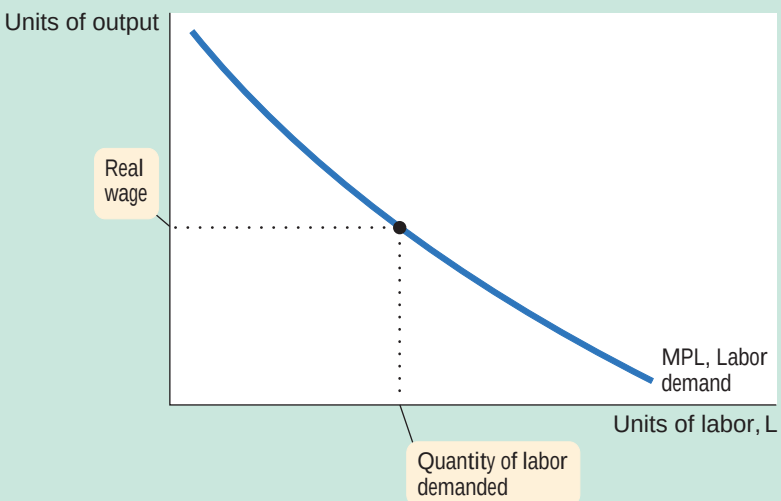
$$MPL = W/P.$$

W/P is the **real wage**—the payment to labor measured in units of output rather than in dollars. To maximize profit, the firm hires up to the point at which the marginal product of labor equals the real wage.

For example, again consider a bakery. Suppose the price of bread P is $2 per loaf, and a worker earns a wage W of $20 per hour. The real wage W/P is 10 loaves per hour. In this example, the firm keeps hiring workers as long as each additional worker would produce at least 10 loaves per hour. When the MPL falls to 10 loaves per hour or less, hiring additional workers is no longer profitable.

Figure 3-4 shows how the marginal product of labor depends on the amount of labor employed (holding the firm's capital stock constant). That is, this figure graphs the MPL schedule. Because the MPL diminishes as the amount of labor increases, this curve slopes downward. For any given real wage, the firm hires up to the point at which the MPL equals the real wage. Hence, the MPL schedule is also the firm's labor demand curve.

figure 3-4



The Marginal Product of Labor Schedule The marginal product of labor MPL depends on the amount of labor. The MPL curve slopes downward because the MPL declines as L increases. The firm hires labor up to the point where the real wage W/P equals the MPL. Hence, this schedule is also the firm's labor demand curve.

The Marginal Product of Capital and Capital Demand The firm decides how much capital to rent in the same way it decides how much labor to hire. The marginal product of capital (*MPK*) is the amount of extra output the firm gets from an extra unit of capital, holding the amount of labor constant:

$$MPK = F(K + 1, L) - F(K, L).$$

Thus, the marginal product of capital is the difference between the amount of output produced with $K + 1$ units of capital and that produced with only K units of capital. Like labor, capital is subject to diminishing marginal product.

The increase in profit from renting an additional machine is the extra revenue from selling the output of that machine minus the machine's rental price:

$$\begin{aligned}\Delta\text{Profit} &= \Delta\text{Revenue} - \Delta\text{Cost} \\ &= (P \times MPK) - R.\end{aligned}$$

To maximize profit, the firm continues to rent more capital until the MPK falls to equal the real rental price:

$$MPK = R/P.$$

The **real rental price** of capital is the rental price measured in units of goods rather than in dollars.

To sum up, the competitive, profit-maximizing firm follows a simple rule about how much labor to hire and how much capital to rent. The firm demands each factor of production until that factor's marginal product falls to equal its real factor price.

The Division of National Income

Having analyzed how a firm decides how much of each factor to employ, we can now explain how the markets for the factors of production distribute the economy's total income. If all firms in the economy are competitive and profit-maximizing, then each factor of production is paid its marginal contribution to the production process. The real wage paid to each worker equals MPL, and the real rental price paid to each owner of capital equals MPK. The total real wages paid to labor are therefore $MPL \times L$, and the total real return paid to capital owners is $MPK \times K$.

The income that remains after the firms have paid the factors of production is the **economic profit** of the owners of the firms. Real economic profit is

$$\text{Economic Profit} = Y - (MPL \times L) - (MPK \times K).$$

Because we want to examine the distribution of national income, we rearrange the terms as follows:

$$Y = (MPL \times L) + (MPK \times K) + \text{Economic Profit}.$$

Total income is divided among the return to labor, the return to capital, and economic profit.

How large is economic profit? The answer is surprising: if the production function has the property of constant returns to scale, as is often thought to be the case, then economic profit must be zero. That is, nothing is left after the factors of production are paid. This conclusion follows from a famous mathematical result called Euler's theorem,[2] which states that if the production function has constant returns to scale, then

$$F(K, L) = (MPK \times K) + (MPL \times L).$$

If each factor of production is paid its marginal product, then the sum of these factor payments equals total output. In other words, constant returns to scale, profit maximization, and competition together imply that economic profit is zero.

If economic profit is zero, how can we explain the existence of "profit" in the economy? The answer is that the term "profit" as normally used is different from economic profit. We have been assuming that there are three types of agents: workers, owners of capital, and owners of firms. Total income is divided among wages, return to capital, and economic profit. In the real world, however, most firms own rather than rent the capital they use. Because firm owners and capital owners are the same people, economic profit and the return to capital are often lumped together. If we call this alternative definition accounting profit, we can say that

$$\text{Accounting Profit} = \text{Economic Profit} + (MPK \times K).$$

Under our assumptions—constant returns to scale, profit maximization, and competition—economic profit is zero. If these assumptions approximately describe the world, then the "profit" in the national income accounts must be mostly the return to capital.

We can now answer the question posed at the beginning of this chapter about how the income of the economy is distributed from firms to households. Each factor of production is paid its marginal product, and these factor payments exhaust total output. Total output is divided between the payments to capital and the payments to labor, depending on their marginal productivities.

CASE STUDY

The Black Death and Factor Prices

As we have just learned, in the neoclassical theory of distribution, factor prices equal the marginal products of the factors of production. Because the marginal products depend on the quantities of the factors, a change in the quantity of any one factor alters the marginal products of all the factors. Therefore, a change in the supply of a factor alters equilibrium factor prices.

Fourteenth-century Europe provides a vivid example of how factor quantities affect factor prices. The outbreak of the bubonic plague—the Black Death—in

[2] Mathematical note: To prove Euler's theorem, begin with the definition of constant returns to scale: $zY = F(zK, zL)$. Now differentiate with respect to z and then evaluate at $z = 1$.

1348 reduced the population of Europe by about one-third within a few years. Because the marginal product of labor increases as the amount of labor falls, this massive reduction in the labor force raised the marginal product of labor. (The economy moved to the left along the curves in Figures 3-3 and 3-4.) Real wages did increase substantially during the plague years, by some estimates. The peasants who were fortunate enough to survive the plague enjoyed economic prosperity.

The reduction in the labor force caused by the plague also affected the return to land, the other major factor of production in medieval Europe. With fewer workers available to farm the land, an additional unit of land produced less additional output. This fall in the marginal product of land led to a decline in real rents of 50 percent or more. Thus, while the peasant classes prospered, the landed classes suffered reduced incomes.[3]

3-3 What Determines the Demand for Goods and Services?

We have seen what determines the level of production and how the income from production is distributed to workers and owners of capital. We now continue our tour of the circular flow diagram, Figure 3-1, and examine how the output from production is used.

In Chapter 2 we identified the four components of GDP:

- Consumption (C)
- Investment (I)
- Government purchases (G)
- Net exports (NX).

The circular flow diagram contains only the first three components. For now, to simplify the analysis, we assume a closed economy—a country that does not trade with other countries. Thus, net exports are always zero. (We examine the macroeconomics of open economies in Chapter 5.)

A closed economy has three uses for the goods and services it produces. These three components of GDP are expressed in the national income accounts identity:

$$Y = C + I + G.$$

Households consume some of the economy's output; firms and households use some of the output for investment; and the government buys some of the output for public purposes. We want to see how GDP is allocated among these three uses.

[3] Carlo M. Cipolla, *Before the Industrial Revolution: European Society and Economy, 1000–1700*, 2nd ed. (New York: Norton, 1980), 200–202.

Consumption

When we eat food, wear clothing, or go to a movie, we are consuming some of the output of the economy. All forms of consumption together make up two-thirds of GDP. Because consumption is so large, macroeconomists have devoted much energy to studying how households decide how much to consume. Chapter 16 examines this work in detail. Here we consider the simplest story of consumer behavior.

Households receive income from their labor and their ownership of capital, pay taxes to the government, and then decide how much of their after-tax income to consume and how much to save. As we discussed in Section 3-2, the income that households receive equals the output of the economy, Y. The government then taxes households an amount T. (Although the government imposes many kinds of taxes, such as personal and corporate income taxes and sales taxes, for our purposes we can lump all these taxes together.) We refer to income after the payment of all taxes, $Y - T$, as *disposable income*. Households divide their disposable income between consumption and saving.

We assume that the level of consumption depends directly on the level of disposable income. The higher the disposable income, the greater the consumption. Thus,

$$C = C(Y - T).$$

This equation states that consumption is a function of disposable income. The relationship between consumption and disposable income is called the *consumption function*.

The *marginal propensity to consume* (*MPC*) is the amount by which consumption changes when disposable income increases by one dollar. The *MPC* is between zero and one: an extra dollar of income increases consumption, but by less than one dollar. Thus, if households obtain an extra dollar of income, they save a portion of it. For example, if the *MPC* is 0.7, then households spend 70 cents of each additional dollar of disposable income on consumer goods and services and save 30 cents.

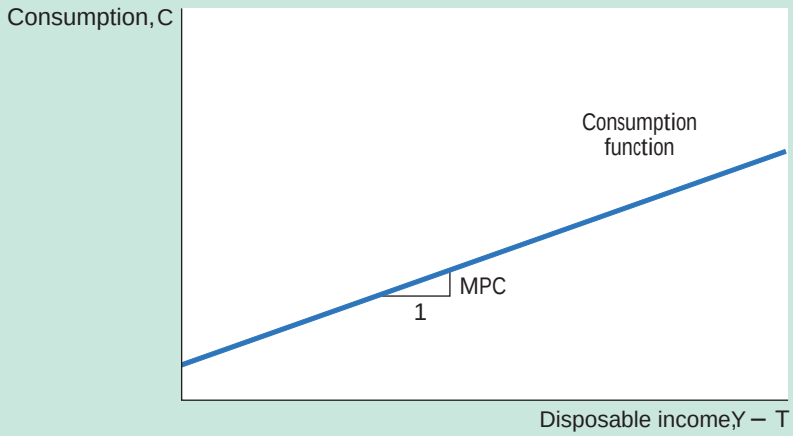
Figure 3-5 illustrates the consumption function. The slope of the consumption function tells us how much consumption increases when disposable income increases by one dollar. That is, the slope of the consumption function is the *MPC*.

Investment

Both firms and households purchase investment goods. Firms buy investment goods to add to their stock of capital and to replace existing capital as it wears out. Households buy new houses, which are also part of investment. Total investment in the United States averages about 15 percent of GDP.

The quantity of investment goods demanded depends on the interest rate, which measures the cost of the funds used to finance investment. For an investment project to be profitable, its return (the revenue from increased future

figure 3-5



The Consumption Function
The consumption function relates consumption, C, to disposable income, $Y - T$. The marginal propensity to consume, MPC, is the amount by which consumption increases when disposable income increases by one dollar.

production of goods and services) must exceed its cost (the payments for borrowed funds). If the interest rate rises, fewer investment projects are profitable, and the quantity of investment goods demanded falls.

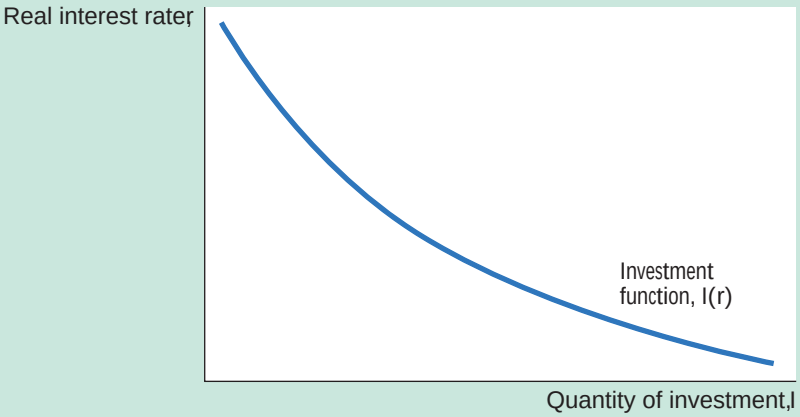
For example, suppose a firm is considering whether it should build a $1 million factory that would yield a return of $100,000 per year, or 10 percent. The firm compares this return to the cost of borrowing the $1 million. If the interest rate is below 10 percent, the firm borrows the money in financial markets and makes the investment. If the interest rate is above 10 percent, the firm forgoes the investment opportunity and does not build the factory.

The firm makes the same investment decision even if it does not have to borrow the $1 million but rather uses its own funds. The firm can always deposit this money in a bank or a money market fund and earn interest on it. Building the factory is more profitable than the deposit if and only if the interest rate is less than the 10 percent return on the factory.

A person wanting to buy a new house faces a similar decision. The higher the interest rate, the greater the cost of carrying a mortgage. A $100,000 mortgage costs $8,000 per year if the interest rate is 8 percent and $10,000 per year if the interest rate is 10 percent. As the interest rate rises, the cost of owning a home rises, and the demand for new homes falls.

When studying the role of interest rates in the economy, economists distinguish between the nominal interest rate and the real interest rate. This distinction is relevant when the overall level of prices is changing. The nominal interest rate is the interest rate as usually reported: it is the rate of interest that investors pay to borrow money. The real interest rate is the nominal interest rate corrected for the effects of inflation. If the nominal interest rate is 8 percent and the inflation rate is 3 percent, then the real interest rate is 5 percent. In Chapter 4 we discuss the relation between nominal and real interest rates in detail. Here it is sufficient to note that the real interest rate measures the true cost of borrowing and, thus, determines the quantity of investment.

figure 3-6



The Investment Function
The investment function relates the quantity of investment, I, to the real interest rate, r. Investment depends on the real interest rate because the interest rate is the cost of borrowing. The investment function slopes downward: when the interest rate rises, fewer investment projects are profitable.

We can summarize this discussion with an equation relating investment, I, to the real interest rate, r:

$$I = I(r).$$

Figure 3-6 shows this investment function. It slopes downward, because as the interest rate rises, the quantity of investment demanded falls.

Government Purchases

Government purchases are the third component of the demand for goods and services. The federal government buys guns, missiles, and the services of government employees. Local governments buy library books, build schools, and hire teachers. Governments at all levels build roads and other public works. All these transactions make up government purchases of goods and services, which account for about 20 percent of GDP in the United States.

These purchases are only one type of government spending. The other type is transfer payments to households, such as welfare for the poor and Social Security payments for the elderly. Unlike government purchases, transfer payments are not made in exchange for some of the economy's output of goods and services. Therefore, they are not included in the variable G.

Transfer payments do affect the demand for goods and services indirectly. Transfer payments are the opposite of taxes: they increase household disposable income, just as taxes reduce disposable income. Thus, an increase in transfer payments financed by an increase in taxes leaves disposable income unchanged. We can now revise our definition of T to equal taxes minus transfer payments. Disposable income, Y – T, includes both the negative impact of taxes and the positive impact of transfer payments.

If government purchases equal taxes minus transfers, G = T, and the government has a balanced budget. If G exceeds T, the government runs a budget

FYI The Many Different Interest Rates

If you look in the business section of a newspaper, you will find many different interest rates reported. By contrast, throughout this book, we will talk about "the" interest rate, as if there were only one interest rate in the economy. The only distinction we will make is between the nominal interest rate (which is not corrected for inflation) and the real interest rate (which is corrected for inflation). Almost all of the interest rates reported in the newspaper are nominal.

Why does the newspaper report so many interest rates? The various interest rates differ in three ways:

- *Term.* Some loans in the economy are for short periods of time, even as short as overnight. Other loans are for 30 years or even longer. The interest rate on a loan depends on its term. Long-term interest rates are usually, but not always, higher than short-term interest rates.
- *Credit risk.* In deciding whether to make a loan, a lender must take into account the probability that the borrower will repay. The law allows borrowers to default on their loans by declaring bankruptcy. The higher the perceived probability of default, the higher the interest rate. The safest credit risk is the government, and so government bonds tend to pay a low interest rate. At the other extreme, financially shaky corporations can raise funds only by issuing *junk bonds*, which pay a high interest rate to compensate for the high risk of default.
- *Tax treatment.* The interest on different types of bonds is taxed differently. Most important, when state and local governments issue bonds, called *municipal bonds*, the holders of the bonds do not pay federal income tax on the interest income. Because of this tax advantage, municipal bonds pay a lower interest rate.

When you see two different interest rates in the newspaper, you can almost always explain the difference by considering the term, the credit risk, and the tax treatment of the loan.

Although there are many different interest rates in the economy, macroeconomists can usually ignore these distinctions. The various interest rates tend to move up and down together. The assumption that there is only one interest rate is, for our purposes, a useful simplification.

deficit, which it funds by issuing government debt—that is, by borrowing in the financial markets. If G is less than T, the government runs a budget surplus, which it can use to repay some of its outstanding debt.

Here we do not try to explain the political process that leads to a particular fiscal policy—that is, to the level of government purchases and taxes. Instead, we take government purchases and taxes as exogenous variables. To denote that these variables are fixed outside of our model of national income, we write

$$G = \bar{G},$$
$$T = \bar{T}.$$

We do, however, want to examine the impact of fiscal policy on the variables determined within the model, the endogenous variables. The endogenous variables here are consumption, investment, and the interest rate.

To see how the exogenous variables affect the endogenous variables, we must complete the model. This is the subject of the next section.

3-4 What Brings the Supply and Demand for Goods and Services Into Equilibrium?

We have now come full circle in the circular flow diagram, Figure 3-1. We began by examining the supply of goods and services, and we have just discussed the demand for them. How can we be certain that all these flows balance? In other words, what ensures that the sum of consumption, investment, and government purchases equals the amount of output produced? We will see that in this classical model, the interest rate has the crucial role of equilibrating supply and demand.

There are two ways to think about the role of the interest rate in the economy. We can consider how the interest rate affects the supply and demand for goods or services. Or we can consider how the interest rate affects the supply and demand for loanable funds. As we will see, these two approaches are two sides of the same coin.

Equilibrium in the Market for Goods and Services: The Supply and Demand for the Economy's Output

The following equations summarize the discussion of the demand for goods and services in Section 3-3:

$$Y = C + I + G.$$
$$C = C(Y - T).$$
$$I = I(r).$$
$$G = \bar{G}.$$
$$T = \bar{T}.$$

The demand for the economy's output comes from consumption, investment, and government purchases. Consumption depends on disposable income; investment depends on the real interest rate; and government purchases and taxes are the exogenous variables set by fiscal policymakers.

To this analysis, let's add what we learned about the supply of goods and services in Section 3-1. There we saw that the factors of production and the production function determine the quantity of output supplied to the economy:

$$Y = F(\bar{K}, \bar{L})$$
$$= \bar{Y}.$$

Now let's combine these equations describing the supply and demand for output. If we substitute the consumption function and the investment function into the national income accounts identity, we obtain

$$Y = C(Y - T) + I(r) + G.$$

Because the variables G and T are fixed by policy, and the level of output Y is fixed by the factors of production and the production function, we can write

$$\bar{Y} = C(\bar{Y} - \bar{T}) + I(r) + \bar{G}.$$

This equation states that the supply of output equals its demand, which is the sum of consumption, investment, and government purchases.

Notice that the interest rate r is the only variable not already determined in the last equation. This is because the interest rate still has a key role to play: it must adjust to ensure that the demand for goods equals the supply. The greater the interest rate, the lower the level of investment, and thus the lower the demand for goods and services $C + I + G$. If the interest rate is too high, investment is too low, and the demand for output falls short of the supply. If the interest rate is too low, investment is too high, and the demand exceeds the supply. At the equilibrium interest rate, the demand for goods and services equals the supply.

This conclusion may seem somewhat mysterious. One might wonder how the interest rate gets to the level that balances the supply and demand for goods and services. The best way to answer this question is to consider how financial markets fit into the story.

Equilibrium in the Financial Markets: The Supply and Demand for Loanable Funds

Because the interest rate is the cost of borrowing and the return to lending in financial markets, we can better understand the role of the interest rate in the economy by thinking about the financial markets. To do this, rewrite the national income accounts identity as

$$Y - C - G = I.$$

The term $Y - C - G$ is the output that remains after the demands of consumers and the government have been satisfied; it is called national saving or simply saving (S). In this form, the national income accounts identity shows that saving equals investment.

To understand this identity more fully, we can split national saving into two parts—one part representing the saving of the private sector and the other representing the saving of the government:

$$(Y - T - C) + (T - G) = I.$$

The term $(Y - T - C)$ is disposable income minus consumption, which is private saving. The term $(T - G)$ is government revenue minus government spending, which is public saving. (If government spending exceeds government revenue, the government runs a budget deficit, and public saving is negative.) National saving is the sum of private and public saving. The circular flow

diagram in Figure 3-1 reveals an interpretation of this equation: this equation states that the flows into the financial markets (private and public saving) must balance the flows out of the financial markets (investment).

To see how the interest rate brings financial markets into equilibrium, substitute the consumption function and the investment function into the national income accounts identity:

$$Y - C(Y - T) - G = I(r).$$

Next, note that G and T are fixed by policy and Y is fixed by the factors of production and the production function:

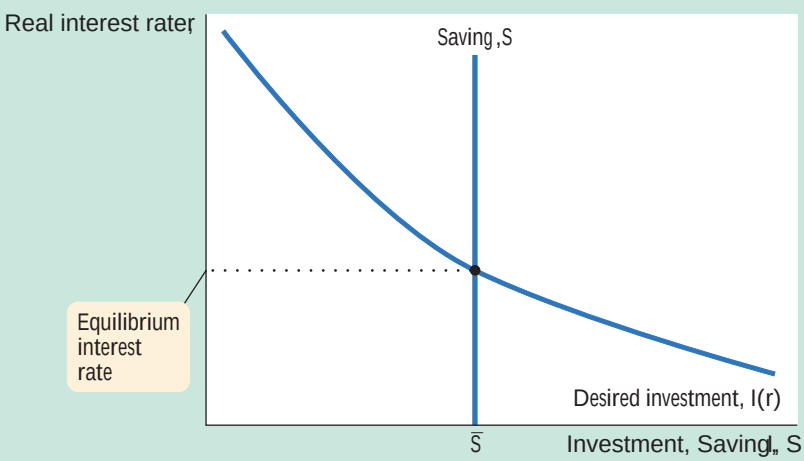
$$\bar{Y} - C(\bar{Y} - \bar{T}) - \bar{G} = I(r)$$
$$\bar{S} = I(r).$$

The left-hand side of this equation shows that national saving depends on income Y and the fiscal-policy variables G and T. For fixed values of Y, G, and T, national saving S is also fixed. The right-hand side of the equation shows that investment depends on the interest rate.

Figure 3-7 graphs saving and investment as a function of the interest rate. The saving function is a vertical line because in this model saving does not depend on the interest rate (although we relax this assumption later). The investment function slopes downward: the higher the interest rate, the fewer profitable investment projects.

From a quick glance at Figure 3-7, one might think it was a supply-and-demand diagram for a particular good. In fact, saving and investment can be interpreted in terms of supply and demand. In this case, the "good" is loanable funds, and its "price" is the interest rate. Saving is the supply of loanable funds

figure 3-7



Saving, Investment, and the Interest Rate The interest rate adjusts to bring saving and investment into balance. The vertical line represents saving—the supply of loanable funds. The downward-sloping line represents investment—the demand for loanable funds. The intersection of these two curves determines the equilibrium interest rate.

households lend their saving to investors or deposit their saving in a bank that then loans the funds out. Investment is the demand for loanable funds—firms borrow from the public directly by selling bonds or indirectly by borrowing from banks. Because investment depends on the interest rate, the quantity of loanable funds demanded also depends on the interest rate.

The interest rate adjusts until the amount that firms want to invest equals the amount that households want to save. If the interest rate is too low, investors want more of the economy's output than households want to save. Equivalently, the quantity of loanable funds demanded exceeds the quantity supplied. When this happens, the interest rate rises. Conversely, if the interest rate is too high, households want to save more than firms want to invest; because the quantity of loanable funds supplied is greater than the quantity demanded, the interest rate falls. The equilibrium interest rate is found where the two curves cross. At the equilibrium interest rate, household's desired saving balances firms' desire to invest, and the quantity of loanable funds supplied equals the quantity demanded.

Changes in Saving: The Effects of Fiscal Policy

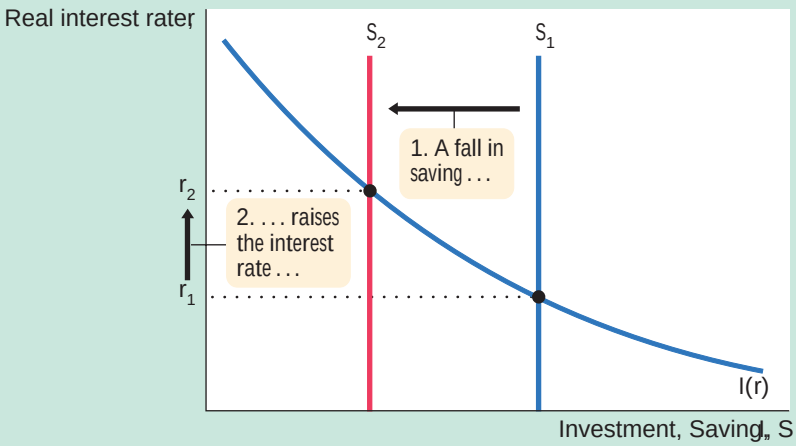
We can use our model to show how fiscal policy affects the economy. When the government changes its spending or the level of taxes, it affects the demand for the economy's output of goods and services and alters national saving, investment, and the equilibrium interest rate.

An Increase in Government Purchases Consider first the effects of an increase in government purchases of an amount ΔG. The immediate impact is to increase the demand for goods and services by ΔG. But since total output is fixed by the factors of production, the increase in government purchases must be met by a decrease in some other category of demand. Because disposable income $Y - T$ is unchanged, consumption C is unchanged. The increase in government purchases must be met by an equal decrease in investment.

To induce investment to fall, the interest rate must rise. Hence, the increase in government purchases causes the interest rate to increase and investment to decrease. Government purchases are said to *crowd out* investment.

To grasp the effects of an increase in government purchases, consider the impact on the market for loanable funds. Because the increase in government purchases is not accompanied by an increase in taxes, the government finances the additional spending by borrowing—that is, by reducing public saving. With private saving unchanged, this government borrowing reduces national saving. As Figure 3-8 shows, a reduction in national saving is represented by a leftward shift in the supply of loanable funds available for investment. At the initial interest rate, the demand for loanable funds exceeds the supply. The equilibrium interest rate rises to the point where the investment schedule crosses the new saving schedule. Thus, an increase in government purchases causes the interest rate to rise from r_1 to r_2.

figure 3-8



A Reduction in Saving A reduction in saving, possibly the result of a change in fiscal policy, shifts the saving schedule to the left. The new equilibrium is the point at which the new saving schedule crosses the investment schedule. A reduction in saving lowers the amount of investment and raises the interest rate. Fiscal-policy actions that reduce saving are said to crowd out investment.

CASE STUDY

Wars and Interest Rates in the United Kingdom, 1730–1920

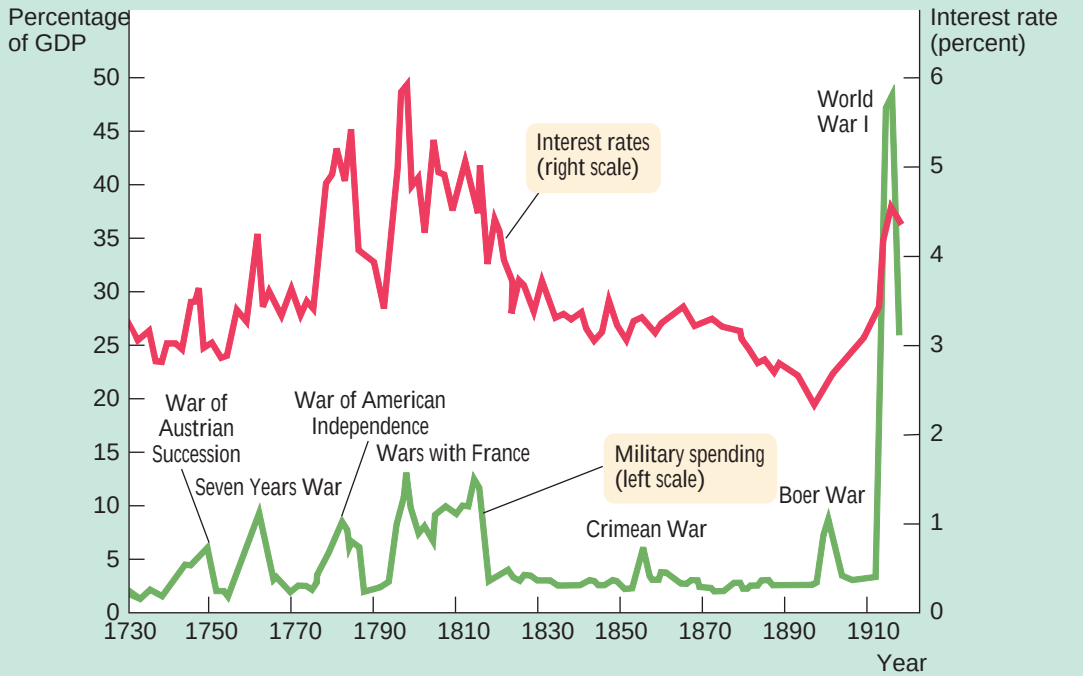
Wars are traumatic both for those who fight them and for a nation's economy. Because the economic changes accompanying them are often large, wars provide a natural experiment with which economists can test their theories. We can learn about the economy by seeing how in wartime the endogenous variables respond to the major changes in the exogenous variables.

One exogenous variable that changes substantially in wartime is the level of government purchases. Figure 3-9 shows military spending as a percentage of GDP for the United Kingdom from 1730 to 1919. This graph shows, as one would expect, that government purchases rose suddenly and dramatically during the eight wars of this period.

Our model predicts that this wartime increase in government purchases and the increase in government borrowing to finance the wars should have raised the demand for goods and services, reduced the supply of loanable funds, and raised the interest rate. To test this prediction, Figure 3-9 also shows the interest rate on long-term government bonds, called consols, in the United Kingdom. A positive association between military purchases and interest rates is apparent in this figure. These data support the model's prediction: interest rates do tend to rise when government purchases increase.[4]

[4] Daniel K. Benjamin and Levis A. Kochin, "War, Prices, and Interest Rates: A Martial Solution to Gibson's Paradox," in M. D. Bordo and A. J. Schwartz, eds., A Retrospective on the Classical Gold Standard, 1821–1931 (Chicago: University of Chicago Press, 1984), 587; Robert J. Barro, "Government Spending, Interest Rates, Prices, and Budget Deficits in the United Kingdom, 1701–1918," Journal of Monetary Economics 20 (September 1987): 221–248.

figure 3-9



Military Spending and the Interest Rate in the United Kingdom This figure shows military spending as a percentage of GDP in the United Kingdom from 1730 to 1919. Not surprisingly, military spending rose substantially during each of the eight wars of this period. This figure also shows that the interest rate tended to rise when military spending rose.

Source: Series constructed from various sources described in Robert J. Barro, "Government Spending, Interest Rates, Prices, and Budget Deficits in the United Kingdom, 1701–1918," Journal of Monetary Economics 20 (September 1987): 221–248.

One problem with using wars to test theories is that many economic changes may be occurring at the same time. For example, in World War II, while government purchases increased dramatically, rationing also restricted consumption of many goods. In addition, the risk of defeat in the war and default by the government on its debt presumably increases the interest rate the government must pay. Economic models predict what happens when one exogenous variable changes and all the other exogenous variables remain constant. In the real world, however, many exogenous variables may change at once. Unlike controlled laboratory experiments, the natural experiments on which economists must rely are not always easy to interpret.

A Decrease in Taxes Now consider a reduction in taxes ΔT. The immediate impact of the tax cut is to raise disposable income and thus to raise consumption. Disposable income rises by ΔT, and consumption rises by an amount equal to

ΔT times the marginal propensity to consume, MPC. The higher the MPC, the greater the impact of the tax cut on consumption.

Because the economy's output is fixed by the factors of production and the level of government purchases is fixed by the government, the increase in consumption must be met by a decrease in investment. For investment to fall, the interest rate must rise. Hence, a reduction in taxes, like an increase in government purchases, crowds out investment and raises the interest rate.

We can also analyze the effect of a tax cut by looking at saving and investment. Because the tax cut raises disposable income, ΔT, consumption goes up by MPC $\times$ ΔT. National saving, S, which equals $Y - C - G$, falls by the same amount as consumption rises. As in Figure 3-8, the reduction in saving shifts the supply of loanable funds to the left, which increases the equilibrium interest rate and crowds out investment.

Changes in Investment Demand

So far, we have discussed how fiscal policy can change national saving. We can also use our model to examine the other side of the market—the demand for investment. In this section we look at the causes and effects of changes in investment demand.

One reason investment demand might increase is technological innovation. Suppose, for example, that someone invents a new technology, such as the railroad or the computer. Before a firm or household can take advantage of the innovation, it must buy investment goods. The invention of the railroad had no value until railroad cars were produced and tracks were laid. The idea of the computer was not productive until computers were manufactured. Thus, technological innovation leads to an increase in investment demand.

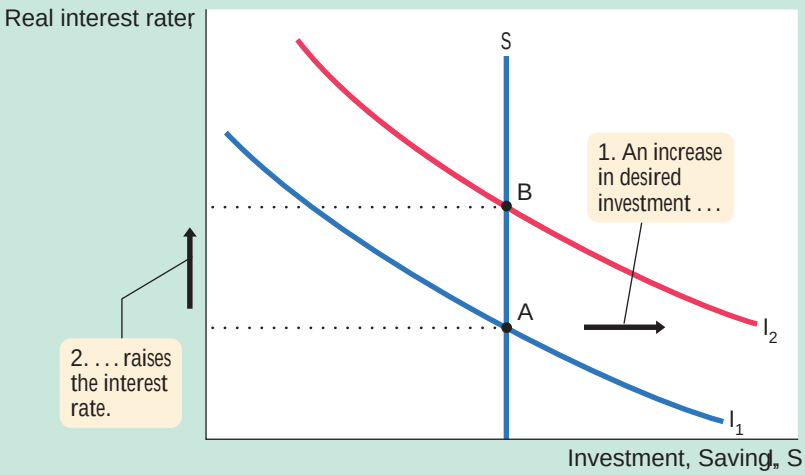
Investment demand may also change because the government encourages or discourages investment through the tax laws. For example, suppose that the government increases personal income taxes and uses the extra revenue to provide tax cuts for those who invest in new capital. Such a change in the tax laws makes more investment projects profitable and, like a technological innovation, increases the demand for investment goods.

Figure 3-10 shows the effects of an increase in investment demand. At any given interest rate, the demand for investment goods (and also for loanable funds) is higher. This increase in demand is represented by a shift in the investment schedule to the right. The economy moves from the old equilibrium, point A, to the new equilibrium, point B.

The surprising implication of Figure 3-10 is that the equilibrium amount of investment is unchanged. Under our assumptions, the fixed level of saving determines the amount of investment; in other words, there is a fixed supply of loanable funds. An increase in investment demand merely raises the equilibrium interest rate.

We would reach a different conclusion, however, if we modified our simple consumption function and allowed consumption (and its flip side, saving) to

figure 3-10

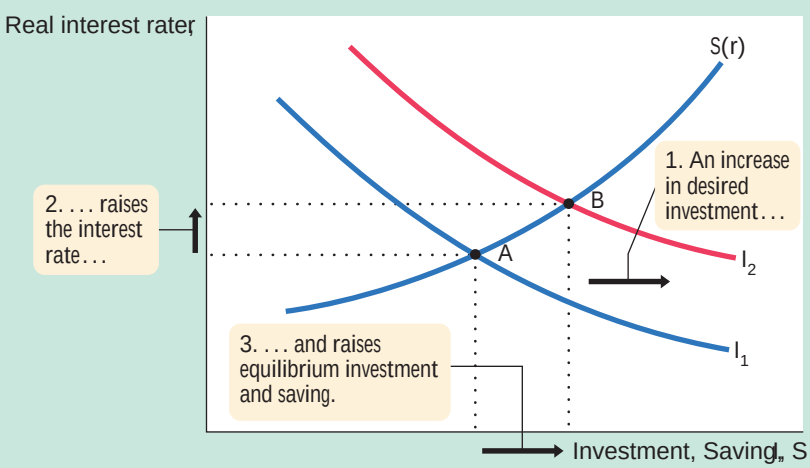


An Increase in the Demand for Investment An increase in the demand for investment goods shifts the investment schedule to the right. At any given interest rate, the amount of investment is greater. The equilibrium moves from point A to point B. Because the amount of saving is fixed, the increase in investment demand raises the interest rate while leaving the equilibrium amount of investment unchanged.

depend on the interest rate. Because the interest rate is the return to saving (as well as the cost of borrowing), a higher interest rate might reduce consumption and increase saving. If so, the saving schedule would be upward sloping, rather than vertical.

With an upward-sloping saving schedule, an increase in investment demand would raise both the equilibrium interest rate and the equilibrium quantity of investment. Figure 3-11 shows such a change. The increase in the interest rate causes households to consume less and save more. The decrease in consumption frees resources for investment.

figure 3-11



An Increase in Investment Demand When Saving Depends on the Interest Rate When saving is positively related to the interest rate, a rightward shift in the investment schedule increases the interest rate and the amount of investment. The higher interest rate induces people to increase saving, which in turn allows investment to increase.

FYI

The Identification Problem

In our model, investment depends on the interest rate. The higher the interest rate, the fewer investment projects there are that are profitable. The investment schedule therefore slopes downward.

Economists who look at macroeconomic data, however, usually fail to find an obvious association between investment and interest rates. In years when interest rates are high, investment is not always low. In years when interest rates are low, investment is not always high.

How do we interpret this finding? Does it mean that investment does not depend on the interest rate? Does it suggest that our model of saving, investment, and the interest rate is inconsistent with how the economy actually functions?

Luckily, we do not have to discard our model. The inability to find an empirical relationship between investment and interest rates is an example of the *identification problem*. The identification problem arises when variables are related in more than one way. When we look at data, we are observing a combination of these different relationships, and it is difficult to "identify" any one of them.

To understand this problem more concretely, consider the relationships among saving, investment, and the interest rate. Suppose, on the one hand, that all changes in the interest rate resulted from changes in saving—that is, from shifts in the saving schedule. Then, as shown in the left-hand side of panel (a) in Figure 3-12, all changes would represent movement along a fixed investment schedule. As the right-hand side of panel (a) shows, the data would trace out this investment schedule. Thus, we would observe a negative relationship between investment and interest rates.

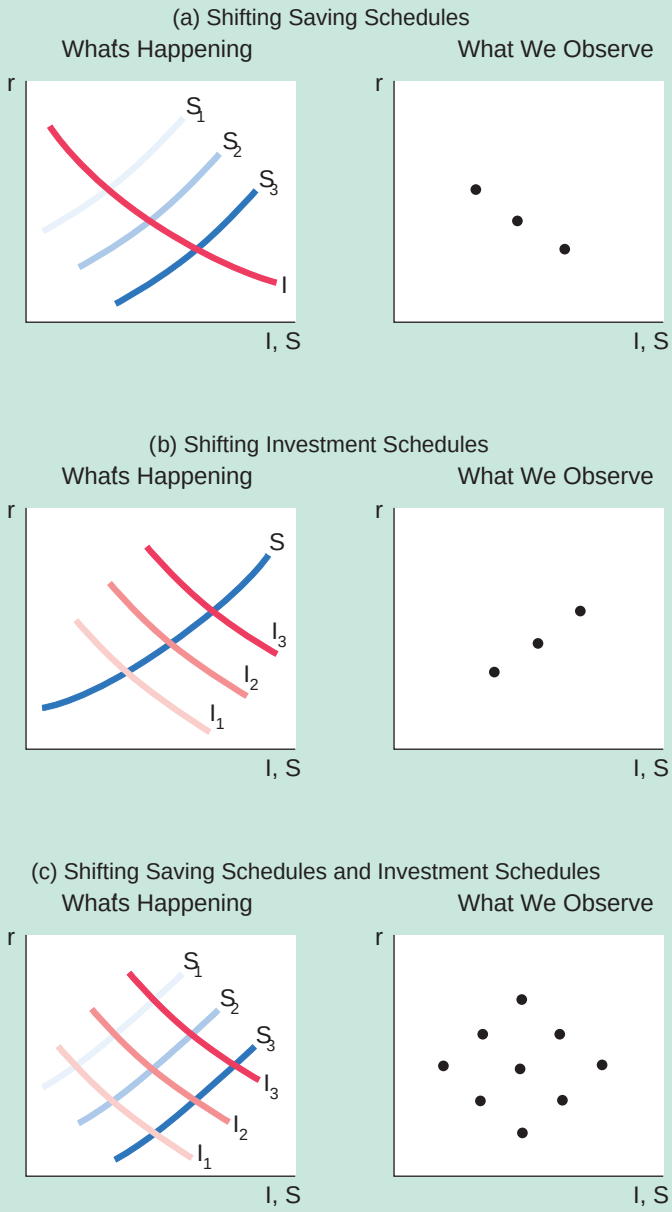
Suppose, on the other hand, that all changes in the interest rate resulted from technological innovations—that is, from shifts in the investment schedule. Then, as shown in panel (b), all changes would represent movements in the investment schedule along a fixed saving schedule. As the right-hand side of panel (b) shows, the data would reflect this saving schedule. Thus, we would observe a positive relationship between investment and interest rates.

In the real world, interest rates change sometimes because of shifts in the saving schedule and sometimes because of shifts in the investment schedule. In this mixed case, as shown in panel (c), a plot of the data would reveal no recognizable relation between interest rates and the quantity of investment, just as economists observe in actual data. The moral of the story is simple and is applicable to many other situations: the empirical relationship we expect to observe depends crucially on which exogenous variables we think are changing.

3-5 Conclusion

In this chapter we have developed a model that explains the production, distribution, and allocation of the economy's output of goods and services. Because the model incorporates all the interactions illustrated in the circular flow diagram in Figure 3-1, it is sometimes called a *general equilibrium model*. The model emphasizes how prices adjust to equilibrate supply and demand. Factor prices equilibrate factor markets. The interest rate equilibrates the supply and demand for goods and services (or, equivalently, the supply and demand for loanable funds).

figure 3-12



Identifying the Investment Function
When we look at data on interest rates r and investment I, what we find depends on which exogenous variables are changing. In panel (a), the saving schedule is shifting, perhaps because of changes in fiscal policy; we would observe a negative correlation between r and I. In panel (b), the investment schedule is shifting, perhaps because of technological innovations; we would observe a positive correlation between r and I. In the more realistic situation shown in panel (c), both schedules are shifting. In the data, we would observe no correlation between r and I, which is in fact what researchers typically find.

Throughout the chapter, we have discussed various applications of the model. The model can explain how income is divided among the factors of production and how factor prices depend on factor supplies. We have also used the model to discuss how fiscal policy alters the allocation of output among its alternative uses—consumption, investment, and government purchases—and how it affects the equilibrium interest rate.

At this point it is useful to review some of the simplifying assumptions we have made in this chapter. In the following chapters we relax some of these assumptions in order to address a greater range of questions.

- We have ignored the role of money, the asset with which goods and services are bought and sold. In Chapter 4 we discuss how money affects the economy and the influence of monetary policy.
- We have assumed that there is no trade with other countries. In Chapter 5 we consider how international interactions affect our conclusions.
- We have assumed that the labor force is fully employed. In Chapter 6 we examine the reasons for unemployment and see how public policy influences the level of unemployment.
- We have assumed that the capital stock, the labor force, and the production technology are fixed. In Chapters 7 and 8 we see how changes over time in each of these lead to growth in the economy's output of goods and services.
- We have ignored the role of short-run sticky prices. In Chapters 9 through 13, we develop a model of short-run fluctuations that includes sticky prices. We then discuss how the model of short-run fluctuations relates to the model of national income developed in this chapter.

Before going on to these chapters, go back to the beginning of this one and make sure you can answer the four groups of questions about national income that begin the chapter.

Summary

1. The factors of production and the production technology determine the economy's output of goods and services. An increase in one of the factors of production or a technological advance raises output.
2. Competitive, profit-maximizing firms hire labor until the marginal product of labor equals the real wage. Similarly, they rent capital until the marginal product of capital equals the real rental price. Therefore, each factor of production is paid its marginal product. If the production function has constant returns to scale, all output is used to compensate the inputs.
3. The economy's output is used for consumption, investment, and government purchases. Consumption depends positively on disposable income. Investment depends negatively on the real interest rate. Government purchases and taxes are the exogenous variables of fiscal policy.
4. The real interest rate adjusts to equilibrate the supply and demand for the economy's output—or, equivalently, to equilibrate the supply of loanable funds (saving) and the demand for loanable funds (investment). A decrease in national saving, perhaps because of an increase in government purchases or a

decrease in taxes, reduces the equilibrium amount of investment and raises the interest rate. An increase in investment demand, perhaps because of a technological innovation or a tax incentive for investment, also raises the interest rate. An increase in investment demand increases the quantity of investment only if higher interest rates stimulate additional saving.

KEY CONCEPTS

Factors of production
Production function
Constant returns to scale
Factor prices
Competition
Marginal product of labor (MPL)
Diminishing marginal product
Real wage
Marginal product of capital (MPK)
Real rental price of capital
Economic profit versus accounting profit
Disposable income
Consumption function
Marginal propensity to consume (MPC)
Nominal interest rate
Real interest rate
National saving (saving)
Private saving
Public saving
Loanable funds
Crowding out

QUESTIONS FOR REVIEW

1. What determines the amount of output an economy produces?
2. Explain how a competitive, profit-maximizing firm decides how much of each factor of production to demand.
3. What is the role of constant returns to scale in the distribution of income?
4. What determines consumption and investment?
5. Explain the difference between government purchases and transfer payments. Give two examples of each.
6. What makes the demand for the economy's output of goods and services equal the supply?
7. Explain what happens to consumption, investment, and the interest rate when the government increases taxes.

PROBLEMS AND APPLICATIONS

1. Use the neoclassical theory of distribution to predict the impact on the real wage and the real rental price of capital of each of the following events:
   a. A wave of immigration increases the labor force.
   b. An earthquake destroys some of the capital stock.
   c. A technological advance improves the production function.
2. If a 10-percent increase in both capital and labor causes output to increase by less than 10 percent, the production function is said to exhibit decreasing returns to scale. If it causes output to increase by more than 10 percent, the production function is said to exhibit increasing returns to scale. What might a production function exhibit decreasing or increasing returns to scale?
3. According to the neoclassical theory of distribution, the real wage earned by any worker equals that worker's marginal productivity. Let us use this insight to examine the incomes of two groups of workers: farmers and barbers.

a. Over the past century, the productivity of farmers has risen substantially because of technological progress. According to the neoclassical theory, what should have happened to their real wage?

b. In what units is the real wage discussed in part (a) measured?

c. Over the same period, the productivity of barbers has remained constant. What should have happened to their real wage?

d. In what units is the real wage in part (c) measured?

e. Suppose workers can move freely between being farmers and being barbers. What does this mobility imply for the wages of farmers and barbers?

f. What do your previous answers imply for the price of haircuts relative to the price of food?

g. Who benefits from technological progress in farming—farmers or barbers?

4. The government raises taxes by $100 billion. If the marginal propensity to consume is 0.6, what happens to the following? Do they rise or fall? By what amounts?

a. Public saving.

b. Private saving.

c. National saving.

d. Investment.

5. Suppose that an increase in consumer confidence raises consumers' expectations of future income and thus the amount they want to consume today. This might be interpreted as an upward shift in the consumption function. How does this shift affect investment and the interest rate?

6. Consider an economy described by the following equations:

$$Y = C + I + G,$$
$$Y = 5{,}000,$$
$$G = 1{,}000,$$
$$T = 1{,}000,$$
$$C = 250 + 0.75(Y - T),$$
$$I = 1{,}000 - 50r.$$

a. In this economy, compute private saving, public saving, and national saving.

b. Find the equilibrium interest rate.

c. Now suppose that G rises to 1,250. Compute private saving, public saving, and national saving.

d. Find the new equilibrium interest rate.

7. Suppose that the government increases taxes and government purchases by equal amounts. What happens to the interest rate and investment in response to this balanced-budget change? Does your answer depend on the marginal propensity to consume?

8. When the government subsidizes investment, such as with an investment tax credit, the subsidy often applies to only some types of investment. This question asks you to consider the effect of such a change. Suppose there are two types of investment in the economy: business investment and residential investment. And suppose that the government institutes an investment tax credit only for business investment.

a. How does this policy affect the demand curve for business investment? The demand curve for residential investment?

b. Draw the economy's supply and demand for loanable funds. How does this policy affect the supply and demand for loanable funds? What happens to the equilibrium interest rate?

c. Compare the old and the new equilibrium. How does this policy affect the total quantity of investment? The quantity of business investment? The quantity of residential investment?

9. If consumption depended on the interest rate, how would that affect the conclusions reached in this chapter about the effects of fiscal policy?

APPENDIX

The Cobb–Douglas Production Function

What production function describes how actual economies turn capital and labor into GDP? The answer to this question came from a historic collaboration between a U.S. senator and a mathematician.

Paul Douglas was a U.S. senator from Illinois from 1949 to 1966. In 1927, however, when he was still a professor of economics, he noticed a surprising fact: the division of national income between capital and labor had been roughly constant over a long period. In other words, as the economy grew more prosperous over time, the total income of workers and the total income of capital owners grew at almost exactly the same rate. This observation caused Douglas to wonder what conditions lead to constant factor shares.

Douglas asked Charles Cobb, a mathematician, what production function, if any, would produce constant factor shares if factors always earned their marginal products. The production function would need to have the property that

$$\text{Capital Income} = MPK \times K = \alpha Y$$

and

$$\text{Labor Income} = MPL \times L = (1 - \alpha)\, Y,$$

where α is a constant between zero and one that measures capital's share of income. That is, α determines what share of income goes to capital and what share goes to labor. Cobb showed that the function with this property is

$$Y = F(K, L) = AK^{\alpha}L^{1-\alpha},$$

where A is a parameter greater than zero that measures the productivity of the available technology. This function became known as the Cobb–Douglas production function.

Let's take a closer look at some of the properties of this production function. First, the Cobb–Douglas production function has constant returns to scale. That is, if capital and labor are increased by the same proportion, then output increases by that proportion as well.[5]

---

[5] Mathematical note: To prove that the Cobb–Douglas production function has constant returns to scale, examine what happens when we multiply capital and labor by a constant z:

$$F(zK, zL) = A(zK)^{\alpha}(zL)^{1-\alpha}.$$

Expanding terms on the right,

$$F(zK, zL) = Az^{\alpha}K^{\alpha}z^{1-\alpha}L^{1-\alpha}.$$

(footnote continues)

Next, consider the marginal products for the Cobb–Douglas production function. The marginal product of labor[6] is

$$MPL = (1 - \alpha) AK^{\alpha}L^{-\alpha},$$

and the marginal product of capital is

$$MPK = \alpha AK^{\alpha-1}L^{1-\alpha}.$$

From these equations, recalling that α is between zero and one, we can see what causes the marginal products of the two factors to change. An increase in the amount of capital raises the MPL and reduces the MPK. Similarly, an increase in the amount of labor reduces the MPL and raises the MPK. A technological advance that increases the parameter A raises the marginal product of both factors proportionately.

The marginal products for the Cobb–Douglas production function can also be written as[7]

$$MPL = (1 - \alpha)Y/L.$$
$$MPK = \alpha Y/K.$$

The MPL is proportional to output per worker, and the MPK is proportional to output per unit of capital. Y/L is called average labor productivity, and Y/K is called average capital productivity. If the production function is Cobb–Douglas, then the marginal productivity of a factor is proportional to its average productivity.

We can now verify that if factors earn their marginal products, then the parameter α indeed tells us how much income goes to labor and how much goes to capital. The total wage bill, which we have seen is $MPL \times L$, is simply $(1 - \alpha)Y$.

Rearranging to bring like terms together, we get

$$F(zK, zL) = z^{\alpha}z^{1-\alpha}AK^{\alpha}L^{1-\alpha}.$$

Since $z^{\alpha}z^{1-\alpha} = z$, our function becomes

$$F(zK, zL) = zAK^{\alpha}L^{1-\alpha}.$$

But $AK^{\alpha}L^{1-\alpha} = F(K, L)$. Thus,

$$F(zK, zL) = zF(K, L) = zY.$$

Hence, the amount of output Y increases by the same factor z, which implies that this production function has constant returns to scale.

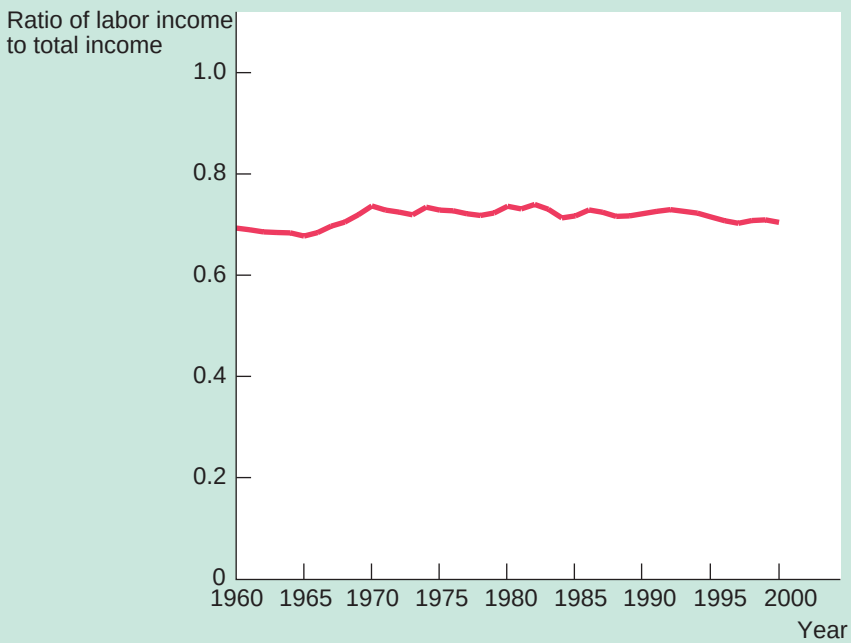
[6] Mathematical note: Obtaining the formulas for the marginal products from the production function requires a bit of calculus. To find the MPL, differentiate the production function with respect to L. This is done by multiplying by the exponent $(1 - \alpha)$, and then subtracting 1 from the old exponent to obtain the new exponent $-\alpha$. Similarly, to obtain the MPK, differentiate the production function with respect to K.

[7] Mathematical note: To check these expressions for the marginal products, substitute in the production function for Y to show that these expressions are equivalent to the earlier formulas for the marginal products.

Therefore, $(1 - \alpha)$ is labor's share of output. Similarly, the total return to capital, $MPK \times K$, is αY, and α is capital's share of output. The ratio of labor income to capital income is a constant, $(1 - \alpha)/\alpha$, just as Douglas observed. The factor shares depend only on the parameter α, not on the amounts of capital or labor or on the state of technology as measured by the parameter A.

More recent U.S. data are also consistent with the Cobb–Douglas production function. Figure 3-13 shows the ratio of labor income to total income in the United States from 1960 to 2000. Despite the many changes in the economy over the past four decades, this ratio has remained about 0.7. This division of income is easily explained by a Cobb–Douglas production function in which the parameter α is about 0.3.

figure 3-13



The Ratio of Labor Income to Total Income Labor income has remained about 0.7 of total income over a long period of time. This approximate constancy of factor shares is evidence for the Cobb–Douglas production function. (This figure is produced from U.S. national income accounts data. Labor income is compensation of employees. Total income is the sum of labor income, corporate profits, net interest, rental income, and depreciation. Proprietors' income is excluded from these calculations, because it is a combination of labor income and capital income.)

Source: U.S. Department of Commerce.

MORE PROBLEMS AND APPLICATIONS

1. Suppose that the production function is Cobb–Douglas with parameter $\alpha = 0.3$.
   a. What fractions of income do capital and labor receive?
   b. Suppose that immigration raises the labor force by 10 percent. What happens to total output (in percent)? The rental price of capital? The real wage?
   c. Suppose that a gift of capital from abroad raises the capital stock by 10 percent. What happens to total output (in percent)? The rental price of capital? The real wage?
   d. Suppose that a technological advance raises the value of the parameter A by 10 percent. What happens to total output (in percent)? The rental price of capital? The real wage?
2. (This problem requires the use of calculus.) Consider a Cobb–Douglas production function with three inputs. K is capital (the number of machines), L is labor (the number of workers), and H is human capital (the number of college degrees among the workers). The production function is

$$Y = K^{1/3}L^{1/3}H^{1/3}.$$

   a. Derive an expression for the marginal product of labor. How does an increase in the amount of human capital affect the marginal product of labor?
   b. Derive an expression for the marginal product of human capital. How does an increase in the amount of human capital affect the marginal product of human capital?
   c. What is the income share paid to labor? What is the income share paid to human capital? In the national income accounts of this economy, what share of total income do you think workers would appear to receive? (Hint: Consider where the return to human capital shows up.)
   d. An unskilled worker earns the marginal product of labor, whereas a skilled worker earns the marginal product of labor plus the marginal product of human capital. Using your answers to (a) and (b), find the ratio of the skilled wage to the unskilled wage. How does an increase in the amount of human capital affect this ratio? Explain.
   e. Some people advocate government funding of college scholarships as a way of creating a more egalitarian society. Others argue that scholarships help only those who are able to go to college. Do your answers to the preceding questions shed light on this debate?

CHAPTER 4 FOUR

Money and Inflation

> There is no subtler, no surer means of overturning the existing basis of society than to debauch the currency. The process engages all the hidden forces of economic law on the side of destruction, and does it in a manner which not one man in a million is able to diagnose.
>
> — John Maynard Keynes

In 1970 the New York Times cost 15 cents, the median price of a single-family home was $23,400, and the average wage in manufacturing was $3.36 per hour. In 2000 the Times cost 75 cents, the price of a home was $166,000, and the average wage was $14.26 per hour. This overall increase in prices is called inflation, and it is the subject of this chapter.

The rate of inflation—the percentage change in the overall level of prices—varies greatly over time and across countries. In the United States, according to the consumer price index, prices rose an average of 2.4 percent per year in the 1960s, 7.1 percent per year in the 1970s, 5.5 percent per year in the 1980s, and 3.0 percent in the 1990s. Even when the U.S inflation problem became severe during the 1970s, it was nothing compared to the episodes of extraordinarily high inflation, called hyperinflation, that other countries have experienced from time to time. A classic example is Germany in 1923, when prices rose an average of 500 percent per month.

In this chapter we examine the classical theory of the causes, effects, and social costs of inflation. The theory is "classical" in the sense that it assumes that prices are flexible. As we first discussed in Chapter 1, most economists believe this assumption describes the behavior of the economy in the long run. By contrast, many prices are thought to be sticky in the short run, and beginning in Chapter 9, we incorporate this fact into our analysis. Yet, for now, we ignore short-run price stickiness. As we will see, the classical theory of inflation not only provides a good description of the long run, it also provides a useful foundation for the short-run analysis we develop later.

The "hidden forces of economic law" that lead to inflation are not as mysterious as Keynes claims in the quotation that opens this chapter. Inflation is simply an increase in the average level of prices, and a price is the rate at which money is exchanged for a good or a service. To understand inflation, therefore,

we must understand money—what it is, what affects its supply and demand, and what influence it has on the economy. Thus, Section 4-1 begins our analysis of inflation by discussing the economist's concept of "money" and how, in most modern economies, the government controls the quantity of money in the hands of the public. Section 4-2 shows that the quantity of money determines the price level and that the rate of growth in the quantity of money determines the rate of inflation.

Inflation in turn has numerous effects on the economy. Section 4-3 discusses the revenue that the government raises by printing money, sometimes called the inflation tax. Section 4-4 examines how inflation affects the nominal interest rate. Section 4-5 discusses how the nominal interest rate in turn affects the quantity of money people wish to hold and, thereby, the price level.

After completing our analysis of the causes and effects of inflation, in Section 4-6 we address what is perhaps the most important question about inflation: Is it a major social problem? Does inflation amount to "overturning the existing basis of society," as the chapter's opening quotation suggests?

Finally, in Section 4-7, we discuss the extreme case of inflation: hyperinflation. Hyperinflations are interesting to examine because they show clearly the causes, effects, and costs of inflation. Just as seismologists learn much by studying earthquakes, economists learn much by studying how hyperinflations begin and end.

4-1 What Is Money?

When we say that a person has a lot of money, we usually mean that he or she is wealthy. By contrast, economists use the term money in a more specialized way. To an economist, money does not refer to all wealth but only to one type of it: money is the stock of assets that can be readily used to make transactions. Roughly speaking, the dollars in the hands of the public make up the nation's stock of money.

The Functions of Money

Money has three purposes. It is a store of value, a unit of account, and a medium of exchange.

As a *store of value*, money is a way to transfer purchasing power from the present to the future. If I work today and earn $100, I can hold the money and spend it tomorrow, next week, or next month. Of course, money is an imperfect store of value: if prices are rising, the amount you can buy with any given quantity of money is falling. Even so, people hold money because they can trade the money for goods and services at some time in the future.

As a *unit of account*, money provides the terms in which prices are quoted and debts are recorded. Microeconomics teaches us that resources are allocated according to relative prices—the prices of goods relative to other goods. In order

stores post their prices in dollars and cents. A car dealer tells you that a car costs $20,000, not 400 shirts (even though it may amount to the same thing). Similarly, most debts require the debtor to deliver a specific number of dollars in the future, not a specified amount of some commodity. Money is the yardstick with which we measure economic transactions.

As a *medium of exchange*, money is what we use to buy goods and services. "This note is legal tender for all debts, public and private," is printed on the U.S. dollar. When we walk into stores, we are confident that the shopkeepers will accept our money in exchange for the items they are selling. The ease with which money is converted into other things—goods and services—is sometimes called money's *liquidity*.

To better understand the functions of money, try to imagine an economy without it: a barter economy. In such a world, trade requires a double coincidence of wants—the unlikely happenstance of two people each having a good that the other wants at the right time and place to make an exchange. A barter economy permits only simple transactions.

Money makes more indirect transactions possible. A professor uses her salary to buy books; the book publisher uses its revenue from the sale of books to buy paper; the paper company uses its revenue from the sale of paper to pay the lumberjack; the lumberjack uses his income to send his child to college; and the college uses its tuition receipts to pay the salary of the professor. In a complex, modern economy, trade is often indirect and requires the use of money.

The Types of Money

Money takes many forms. In the U.S. economy we make transactions with an item whose sole function is to act as money: dollar bills. These pieces of green paper with small portraits of famous Americans would have little value if they were not widely accepted as money. Money that has no intrinsic value is called *fiat money* because it is established as money by government decree, or fiat.

Although fiat money is the norm in most economies today, most societies in the past have used for money a commodity with some intrinsic value. Money of this sort is called *commodity money*. The most widespread example of commodity money is gold. When people use gold as money (or use paper money that is redeemable for gold), the economy is said to be on a *gold standard*. Gold is a form of commodity money because it can be used for various purposes—jewelry, dental fillings, and so on—as well as for transactions. The gold standard was common throughout the world during the late nineteenth century.

Drawing by Bernard Schoenbaum; © 1979 The New Yorker Magazine, Inc.

"And how would you like your funny money?"

CASE STUDY

Money in a POW Camp

An unusual form of commodity money developed in some Nazi prisoner of war (POW) camps during World War II. The Red Cross supplied the prisoners with various goods—food, clothing, cigarettes, and so on. Yet these rations were allocated without close attention to personal preferences, so the allocations were often inefficient. One prisoner may have preferred chocolate, while another may have preferred cheese, and a third may have wanted a new shirt. The differing tastes and endowments of the prisoners led them to trade with one another.

Barter proved to be an inconvenient way to allocate these resources, however, because it required the double coincidence of wants. In other words, a barter system was not the easiest way to ensure that each prisoner received the goods he valued most. Even the limited economy of the POW camp needed some form of money to facilitate transactions.

Eventually, cigarettes became the established "currency" in which prices were quoted and with which trades were made. A shirt, for example, cost about 80 cigarettes. Services were also quoted in cigarettes: some prisoners offered to do other prisoners' laundry for 2 cigarettes per garment. Even nonsmokers were happy to accept cigarettes in exchange, knowing they could trade the cigarettes in the future for some good they did enjoy. Within the POW camp the cigarette became the store of value, the unit of account, and the medium of exchange.[1]

How Fiat Money Evolves

It is not surprising that some form of commodity money arises to facilitate exchange: people are willing to accept a commodity currency such as gold because it has intrinsic value. The development of fiat money, however, is more perplexing. What would make people begin to value something that is intrinsically useless?

To understand how the evolution from commodity money to fiat money takes place, imagine an economy in which people carry around bags of gold. When a purchase is made, the buyer measures out the appropriate amount of gold. If the seller is convinced that the weight and purity of the gold are right, the buyer and seller make the exchange.

The government might first get involved in the monetary system to help people reduce transaction costs. Using raw gold as money is costly because it takes time to verify the purity of the gold and to measure the correct quantity. To reduce these costs, the government can mint gold coins of known purity and weight. The coins are easier to use than gold bullion because their values are widely recognized.

The next step is for the government to accept gold from the public in exchange for gold certificates—pieces of paper that can be redeemed for a certain

[1] R. A. Radford, "The Economic Organisation of a P.O.W. Camp," *Economica* (November 1945): 189–201. The use of cigarettes as money is not limited to this example. In the Soviet Union in the late 1980s, packs of Marlboros were preferred to the ruble in the large underground economy.

quantity of gold. If people believe the government's promise to redeem the paper bills for gold, the bills are just as valuable as the gold itself. In addition, because the bills are lighter than gold (and gold coins), they are easier to use in transactions. Eventually, no one carries gold around at all, and these gold-backed government bills become the monetary standard.

Finally, the gold backing becomes irrelevant. If no one ever bothers to redeem the bills for gold, no one cares if the option is abandoned. As long as everyone continues to accept the paper bills in exchange, they will have value and serve as money. Thus, the system of commodity money evolves into a system of fiat money. Notice that in the end, the use of money in exchange is a social convention: everyone values fiat money because they expect everyone else to value it.

CASE STUDY

Money and Social Conventions on the Island of Yap

The economy of Yap, a small island in the Pacific, had a type of money that was something between commodity and fiat money. The traditional medium of exchange in Yap was large stone wheels up to 12 feet in diameter. These stones had holes in the center so that they could be carried on poles and used for exchange.

Large stone wheels are not a convenient form of money. The stones were heavy, so it took substantial effort for a new owner to take home his purchase after completing a transaction. Although the monetary system facilitated exchange, it did so at great cost.

Eventually, it became common practice for the new owner of a stone not to bother to take physical possession of the stone. Instead, the new owner accepted a claim to the stone without moving it. In future bargains, he traded this claim for goods that he wanted. Having physical possession of the stone became less important than having legal claim to it.

This practice was put to a test when a valuable stone was lost at sea during a storm. Because the owner lost his money by accident rather than through negligence, everyone agreed that his claim to the stone remained valid. Even generations later, when no one alive had ever seen this stone, the claim to it was still valued in exchange.[2]

How the Quantity of Money Is Controlled

The quantity of money available is called the **money supply**. In an economy that uses commodity money, the money supply is the quantity of that commodity. In an economy that uses fiat money, such as most economies today, the government controls the supply of money: legal restrictions give the government a monopoly on the printing of money. Just as the level of taxation and the level of government purchases are policy instruments of the government, so is the supply of money. The control over the money supply is called **monetary policy**.

[2] Norman Angell, *The Story of Money* (New York: Frederick A. Stokes Company, 1929), 88.

In the United States and many other countries, monetary policy is delegated to a partially independent institution called the **central bank**. The central bank of the United States is the **Federal Reserve**—often called the **Fed**. If you look at a U.S. dollar bill, you will see that it is called a **Federal Reserve Note**. Decisions over monetary policy are made by the Federal Open Market Committee. This committee is made up of members of the Federal Reserve Board, who are appointed by the president and confirmed by Congress, together with the presidents of the regional Federal Reserve Banks. The Federal Open Market Committee meets about every six weeks to discuss and set monetary policy.

The primary way in which the Fed controls the supply of money is through **open-market operations**—the purchase and sale of government bonds. When the Fed wants to increase the money supply, it uses some of the dollars it has to buy government bonds from the public. Because these dollars leave the Fed and enter into the hands of the public, the purchase increases the quantity of money in circulation. Conversely, when the Fed wants to decrease the money supply, it sells some government bonds from its own portfolio. This open-market sale of bonds takes some dollars out of the hands of the public and, thus, decreases the quantity of money in circulation.

In Chapter 18 we discuss in detail how the Fed controls the supply of money. For our current discussion, these details are not crucial. It is enough to assume that the Fed (or any other central bank) directly controls the supply of money.

How the Quantity of Money Is Measured

One goal of this chapter is to determine how the money supply affects the economy; we turn to that problem in the next section. As background for that analysis, let's first discuss how economists measure the quantity of money.

Because money is the stock of assets used for transactions, the quantity of money is the quantity of those assets. In simple economies, this quantity is easy to measure. In the POW camp, the quantity of money was the quantity of cigarettes in the camp. But how can we measure the quantity of money in more complex economies such as ours? The answer is not obvious, because no single asset is used for all transactions. People can use various assets, such as cash or checks, to make transactions, although some assets are more convenient than others. This ambiguity leads to numerous measures of the quantity of money.

The most obvious asset to include in the quantity of money is **currency**, the sum of outstanding paper money and coins. Most day-to-day transactions use currency as the medium of exchange.

A second type of asset used for transactions is **demand deposits**, the funds people hold in their checking accounts. If most sellers accept personal checks, assets in a checking account are almost as convenient as currency. In both cases, the assets are in a form ready to facilitate a transaction. Demand deposits are, therefore, added to currency when measuring the quantity of money.

Once we admit the logic of including demand deposits in the measured money stock, many other assets become candidates for inclusion. Funds in savings accounts, for example, can be easily transferred into checking accounts; these

assets are almost as convenient for transactions. Money market mutual funds allow investors to write checks against their accounts, although restrictions sometimes apply with regard to the size of the check or the number of checks written. Since these assets can be easily used for transactions, they should arguably be included in the quantity of money.

table 4-1

The Measures of Money

Symbol	Assets Included	Amount in March 2001 (billions of dollars)
C	Currency	$ 539
M1	Currency plus demand deposits, travelers checks, and other checkable deposits	1,111
M2	M1 plus retail money market mutual fund balances, saving deposits (including money market deposit accounts), and small time deposits	5,100
M3	M2 plus large time deposits, repurchase agreements, Eurodollars, and institution-only money market mutual fund balances	7,326

Source: Federal Reserve.

Because it is hard to judge which assets should be included in the money stock, various measures are available. Table 4-1 presents the four measures of the money stock that the Federal Reserve calculates for the U.S. economy, together with a list of which assets are included in each measure. From the smallest to the largest, they are designated C, M1, M2, and M3. The most common measures for studying the effects of money on the economy are M1 and M2. There is no consensus, however, about which measure of the money stock is best. Disagreements about monetary policy sometimes arise because different measures of money are moving in different directions.

4-2 The Quantity Theory of Money

Having defined what money is and described how it is controlled and measured, we can now examine how the quantity of money affects the economy. To do this, we must see how the quantity of money is related to other economic variables, such as prices and incomes.

Transactions and the Quantity Equation

People hold money to buy goods and services. The more money they need for such transactions, the more money they hold. Thus, the quantity of money in the economy is related to the number of dollars exchanged in transactions.

The link between transactions and money is expressed in the following equation, called the **quantity equation**:

$$\text{Money} \times \text{Velocity} = \text{Price} \times \text{Transactions}$$
$$M \times V = P \times T.$$

Let's examine each of the four variables in this equation.

The right-hand side of the quantity equation tells us about transactions. T represents the total number of transactions during some period of time, say, a year. In other words, T is the number of times in a year that goods or services are exchanged for money. P is the price of a typical transaction—the number of dollars exchanged. The product of the price of a transaction and the number of transactions, PT, equals the number of dollars exchanged in a year.

The left-hand side of the quantity equation tells us about the money used to make the transactions. M is the quantity of money. V is called the **transactions velocity of money** and measures the rate at which money circulates in the economy. In other words, velocity tells us the number of times a dollar bill changes hands in a given period of time.

For example, suppose that 60 loaves of bread are sold in a given year at $0.50 per loaf. Then T equals 60 loaves per year, and P equals $0.50 per loaf. The total number of dollars exchanged is

$$PT = \$0.50/\text{loaf} \times 60 \text{ loaves/year} = \$30/\text{year}.$$

The right-hand side of the quantity equation equals $30 per year, which is the dollar value of all transactions.

Suppose further that the quantity of money in the economy is $10. By rearranging the quantity equation, we can compute velocity as

$$\begin{aligned} V &= PT/M \\ &= (\$30/\text{year})/(\$10) \\ &= 3 \text{ times per year}. \end{aligned}$$

That is, for $30 of transactions per year to take place with $10 of money, each dollar must change hands 3 times per year.

The quantity equation is an **identity**: the definitions of the four variables make it true. The equation is useful because it shows that if one of the variables changes, one or more of the others must also change to maintain the equality. For example, if the quantity of money increases and the velocity of money stays unchanged, then either the price or the number of transactions must rise.

From Transactions to Income

When studying the role of money in the economy, economists usually use a slightly different version of the quantity equation than the one just introduced. The problem with the first equation is that the number of transactions is difficult to measure. To solve this problem, the number of transactions T is replaced by the total output of the economy Y.

Transactions and output are related, because the more the economy produces, the more goods are bought and sold. They are not the same, however. When one person sells a used car to another person, for example, they make a transaction using money, even though the used car is not part of current output. Nonetheless, the dollar value of transactions is roughly proportional to the dollar value of output.

If Y denotes the amount of output and P denotes the price of one unit of output, then the dollar value of output is PY. We encountered measures for these variables when we discussed the national income accounts in Chapter 2: Y is real GDP, P is the GDP deflator, and PY is nominal GDP. The quantity equation becomes

$$\text{Money} \times \text{Velocity} = \text{Price} \times \text{Output}$$
$$M \times V = P \times Y.$$

Because Y is also total income, V, in this version of the quantity equation is called the **income velocity of money**. The income velocity of money tells us the number of times a dollar bill enters someone's income in a given period of time. This version of the quantity equation is the most common, and it is the one we use from now on.

The Money Demand Function and the Quantity Equation

When we analyze how money affects the economy, it is often useful to express the quantity of money in terms of the quantity of goods and services it can buy. This amount, M/P, is called **real money balances**.

Real money balances measure the purchasing power of the stock of money. For example, consider an economy that produces only bread. If the quantity of money is $10, and the price of a loaf is $0.50, then real money balances are 20 loaves of bread. That is, at current prices, the stock of money in the economy is able to buy 20 loaves.

A **money demand function** is an equation that shows what determines the quantity of real money balances people wish to hold. A simple money demand function is

$$(M/P)^d = kY,$$

where k is a constant that tells us how much money people want to hold for every dollar of income. This equation states that the quantity of real money balances demanded is proportional to real income.

The money demand function is like the demand function for a particular good. Here the "good" is the convenience of holding real money balances. Just as owning an automobile makes it easier for a person to travel, holding money makes it easier to make transactions. Therefore, just as higher income leads to a greater demand for automobiles, higher income also leads to a greater demand for real money balances.

This money demand function offers another way to view the quantity equation. To see this, add to the money demand function the condition that the demand for real money balances (M/P)^d must equal the supply M/P. Therefore,

$$M/P = kY.$$

A simple rearrangement of terms changes this equation into

$$M(1/k) = PY,$$

which can be written as

$$MV = PY,$$

where $V = 1/k$. This simple mathematics shows the link between the demand for money and the velocity of money. When people want to hold a lot of money for each dollar of income (k is large), money changes hands infrequently (V is small). Conversely, when people want to hold only a little money (k is small), money changes hands frequently (V is large). In other words, the money demand parameter k and the velocity of money V are opposite sides of the same coin.

The Assumption of Constant Velocity

The quantity equation can be viewed as a definition: it defines velocity V as the ratio of nominal GDP, PY, to the quantity of money M. Yet if we make the additional assumption that the velocity of money is constant, then the quantity equation becomes a useful theory of the effects of money, called the **quantity theory of money**.

As with many of the assumptions in economics, the assumption of constant velocity is only an approximation to reality. Velocity does change if the money demand function changes. For example, when automatic teller machines were introduced, people could reduce their average money holdings, which meant a fall in the money demand parameter k and an increase in velocity V. Nonetheless, experience shows that the assumption of constant velocity provides a good approximation in many situations. Let us therefore assume that velocity is constant and see what this assumption implies about the effects of the money supply on the economy.

Once we assume that velocity is constant, the quantity equation can be seen as a theory of what determines nominal GDP. The quantity equation says

$$M\bar{V} = PY,$$

where the bar over V means that velocity is fixed. Therefore, a change in the quantity of money (M) must cause a proportionate change in nominal GDP (PY). That is, if velocity is fixed, the quantity of money determines the dollar value of the economy's output.

Money, Prices, and Inflation

We now have a theory to explain what determines the economy's overall level of prices. The theory has three building blocks:

1. The factors of production and the production function determine the level of output Y. We borrow this conclusion from Chapter 3.
2. The money supply determines the nominal value of output PY. This conclusion follows from the quantity equation and the assumption that the velocity of money is fixed.
3. The price level P is then the ratio of the nominal value of output PY, to the level of output Y.

In other words, the productive capability of the economy determines real GDP, the quantity of money determines nominal GDP, and the GDP deflator is the ratio of nominal GDP to real GDP.

This theory explains what happens when the Fed changes the supply of money. Because velocity is fixed, any change in the supply of money leads to a proportionate change in nominal GDP. Because the factors of production and the production function have already determined real GDP, the change in nominal GDP must represent a change in the price level. Hence, the quantity theory implies that the price level is proportional to the money supply.

Because the inflation rate is the percentage change in the price level, this theory of the price level is also a theory of the inflation rate. The quantity equation, written in percentage-change form, is

$$\%\ \text{Change in } M + \%\ \text{Change in } V = \%\ \text{Change in } P + \%\ \text{Change in } Y.$$

Consider each of these four terms. First, the percentage change in the quantity of money M is under the control of the central bank. Second, the percentage change in velocity V reflects shifts in money demand; we have assumed that velocity is constant, so the percentage change in velocity is zero. Third, the percentage change in the price level P is the rate of inflation; this is the variable in the equation that we would like to explain. Fourth, the percentage change in output Y depends on growth in the factors of production and on technological progress, which for our present purposes we can take as given. This analysis tells us that (except for a constant that depends on exogenous growth in output) the growth in the money supply determines the rate of inflation.

Thus, the quantity theory of money states that the central bank, which controls the money supply, has ultimate control over the rate of inflation. If the central bank keeps the money supply stable, the price level will be stable. If the central bank increases the money supply rapidly, the price level will rise rapidly

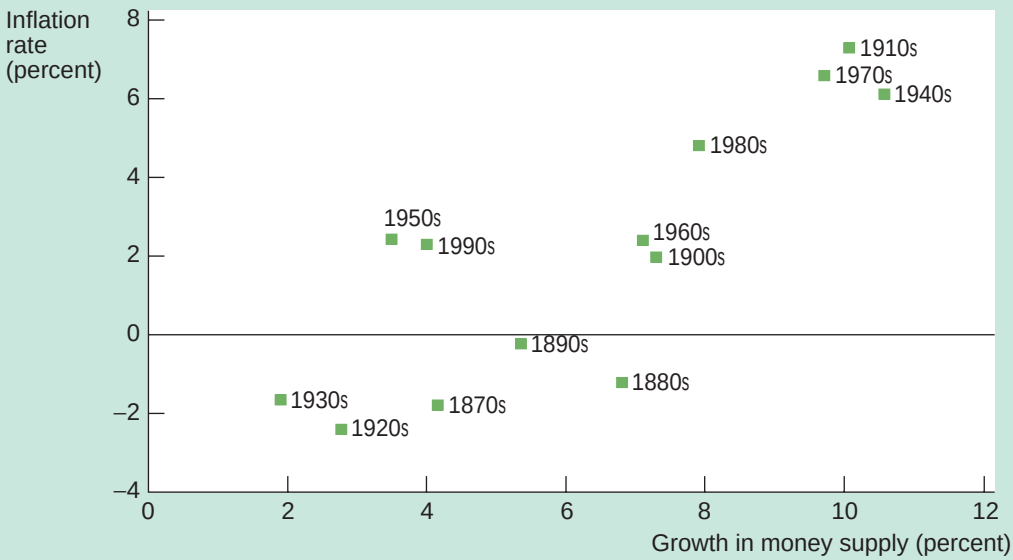
CASE STUDY

Inflation and Money Growth

"Inflation is always and everywhere a monetary phenomenon," wrote Milton Friedman, the great economist who won the Nobel Prize in economics in 1976. The quantity theory of money leads us to agree that the growth in the quantity of money is the primary determinant of the inflation rate. Yet Friedman's claim is empirical, not theoretical. To evaluate his claim, and to judge the usefulness of our theory, we need to look at data on money and prices.

Friedman, together with fellow economist Anna Schwartz, wrote two treatises on monetary history that documented the sources and effects of changes in the quantity of money over the past century.[3] Figure 4-1 uses some of their data and

figure 4-1



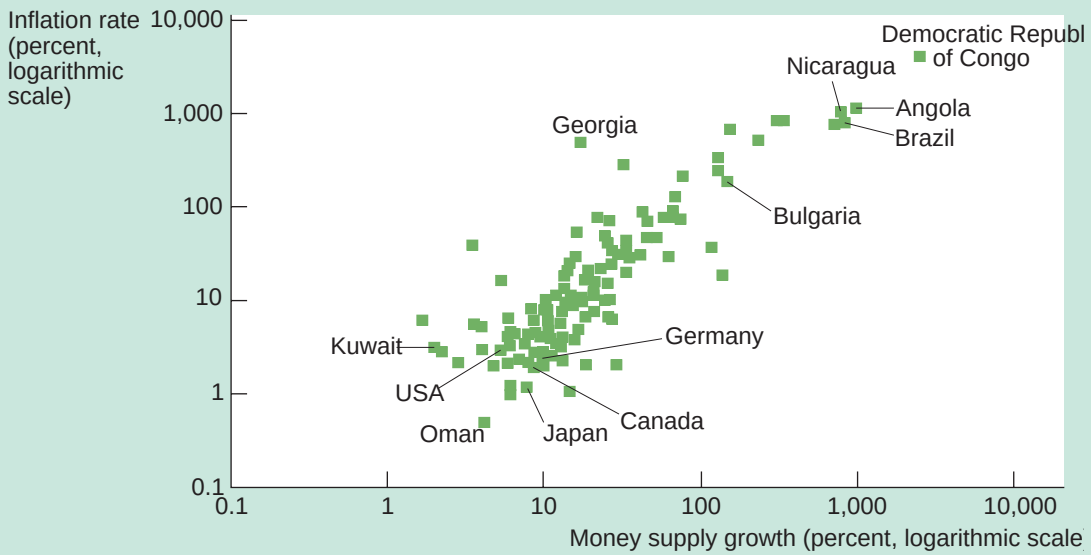
Historical Data on U.S. Inflation and Money Growth In this scatterplot of money growth and inflation, each point represents a decade. The horizontal axis shows the average growth in the money supply (as measured by M2) over the decade, and the vertical axis shows the average rate of inflation (as measured by the GDP deflator). The positive correlation between money growth and inflation is evidence for the quantity theory's prediction that high money growth leads to high inflation.

Source: For the data through the 1960s: Milton Friedman and Anna J. Schwartz, Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975 (Chicago: University of Chicago Press, 1982). For recent data: U.S. Department of Commerce, Federal Reserve Board.

[3] Milton Friedman and Anna J. Schwartz, A Monetary History of the United States, 1867–1960 (Princeton, NJ: Princeton University Press, 1963); Milton Friedman and Anna J. Schwartz, Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975 (Chicago: University of Chicago Press, 1982).

plots the average rate of money growth and the average rate of inflation in the United States over each decade since the 1870s. The data verify the link between inflation and growth in the quantity of money. Decades with high money growth tend to have high inflation, and decades with low money growth tend to have low inflation.

figure 4-2



International Data on Inflation and Money Growth In this scatterplot, each point represents a country. The horizontal axis shows the average growth in the money supply (as measured by currency plus demand deposits) during the 1990s, and the vertical axis shows the average rate of inflation (as measured by the GDP deflator). Once again, the positive correlation is evidence for the quantity theory's prediction that high money growth leads to high inflation.

Source: International Financial Statistics.

Figure 4-2 examines the same question with international data. It shows the average rate of inflation and the average rate of money growth in more than 100 countries during the 1990s. Again, the link between money growth and inflation is clear. Countries with high money growth tend to have high inflation, and countries with low money growth tend to have low inflation.

If we looked at monthly data on money growth and inflation, rather than data for 10-year periods, we would not see as close a connection between these two variables. This theory of inflation works best in the long run, not in the short run. We examine the short-run impact of changes in the quantity of money when we turn to economic fluctuations in Part IV of this book.

4-3 Seigniorage: The Revenue From Printing Money

So far, we have seen how growth in the money supply causes inflation. But what might ever induce the government to increase the money supply? Here we examine one answer to this question.

Let's start with an indisputable fact: all governments spend money. Some of this spending is to buy goods and services (such as roads and police), and some is to provide transfer payments (for the poor and elderly, for example). A government can finance its spending in three ways. First, it can raise revenue through taxes, such as personal and corporate income taxes. Second, it can borrow from the public by selling government bonds. Third, it can print money.

The revenue raised through the printing of money is called **seigniorage**. The term comes from *seigneur*, the French word for "feudal lord." In the Middle Ages, the lord had the exclusive right on his manor to coin money. Today this right belongs to the central government, and it is one source of revenue.

When the government prints money to finance expenditure, it increases the money supply. The increase in the money supply, in turn, causes inflation. Printing money to raise revenue is like imposing an **inflation tax**.

At first it may not be obvious that inflation can be viewed as a tax. After all, no one receives a bill for this tax. The government merely prints the money it needs. Who then pays the inflation tax? The answer is the holders of money. As prices rise, the real value of the money in your wallet falls. When the government prints new money for its use, it makes the old money in the hands of the public less valuable. Thus, inflation is like a tax on holding money.

The amount raised by printing money varies from country to country. In the United States, the amount has been small: seigniorage has usually accounted for less than 3 percent of government revenue. In Italy and Greece, seigniorage has often been more than 10 percent of government revenue. In countries experiencing hyperinflation, seigniorage is often the government's chief source of revenue—indeed, the need to print money to finance expenditure is a primary cause of hyperinflation.

CASE STUDY

Paying for the American Revolution

Although seigniorage has not been a major source of revenue for the U.S. government in recent history, the situation was very different two centuries ago. Beginning in 1775 the Continental Congress needed to find a way to finance the Revolution, but it had limited ability to raise revenue through taxation. It, therefore, relied on the printing of fiat money to help pay for the war.

The Continental Congress's reliance on seigniorage increased over time. In 1775 new issues of continental currency were about $6 million. This amount

[4] Stanley Fischer, "Seigniorage and the Case for a National Money," *Journal of Political Economy* (April 1982): 295–313.

increased to $19 million in 1776, $13 million in 1777, $63 million in 1778, and $125 million in 1779.

Not surprisingly, this rapid growth in the money supply led to massive inflation. At the end of the war, the price of gold measured in continental dollars was more than 100 times its level of only a few years earlier. The large quantity of the continental currency made the continental dollar nearly worthless. This experience also gave birth to a once-popular expression: people used to say something was "not worth a continental," meaning that the item had little real value.

4-4 Inflation and Interest Rates

As we first discussed in Chapter 3, interest rates are among the most important macroeconomic variables. In essence, they are the prices that link the present and the future. Here we discuss the relationship between inflation and interest rates.

Two Interest Rates: Real and Nominal

Suppose you deposit your savings in a bank account that pays 8 percent interest annually. Next year, you withdraw your savings and the accumulated interest. Are you 8 percent richer than you were when you made the deposit a year earlier?

The answer depends on what "richer" means. Certainly, you have 8 percent more dollars than you had before. But if prices have risen, so that each dollar buys less, then your purchasing power has not risen by 8 percent. If the inflation rate was 5 percent, then the amount of goods you can buy has increased by only 3 percent. And if the inflation rate was 10 percent, then your purchasing power has fallen by 2 percent.

Economists call the interest rate that the bank pays the *nominal interest rate* and the increase in your purchasing power the *real interest rate*. If i denotes the nominal interest rate, r the real interest rate, and π the rate of inflation, then the relationship among these three variables can be written as

$$r = i - \pi.$$

The real interest rate is the difference between the nominal interest rate and the rate of inflation.[5]

The Fisher Effect

Rearranging the terms in our equation for the real interest rate, we can show that the nominal interest rate is the sum of the real interest rate and the inflation rate:

$$i = r + \pi.$$

[5] Mathematical note: This equation relating the real interest rate, nominal interest rate, and inflation rate is only an approximation. The exact formula is $1 + r = (1 + i)/(1 + \pi)$. The approximation in the text is reasonably accurate as long as r, i, and π are relatively small (say, less than 20 percent per year).

The equation written in this way is called the Fisher equation, after economist Irving Fisher (1867–1947). It shows that the nominal interest rate can change for two reasons: because the real interest rate changes or because the inflation rate changes.

Once we separate the nominal interest rate into these two parts, we can use this equation to develop a theory that explains the nominal interest rate. Chapter 3 showed that the real interest rate adjusts to equilibrate saving and investment. The quantity theory of money shows that the rate of money growth determines the rate of inflation. The Fisher equation then tells us to add the real interest rate and the inflation rate together to determine the nominal interest rate.

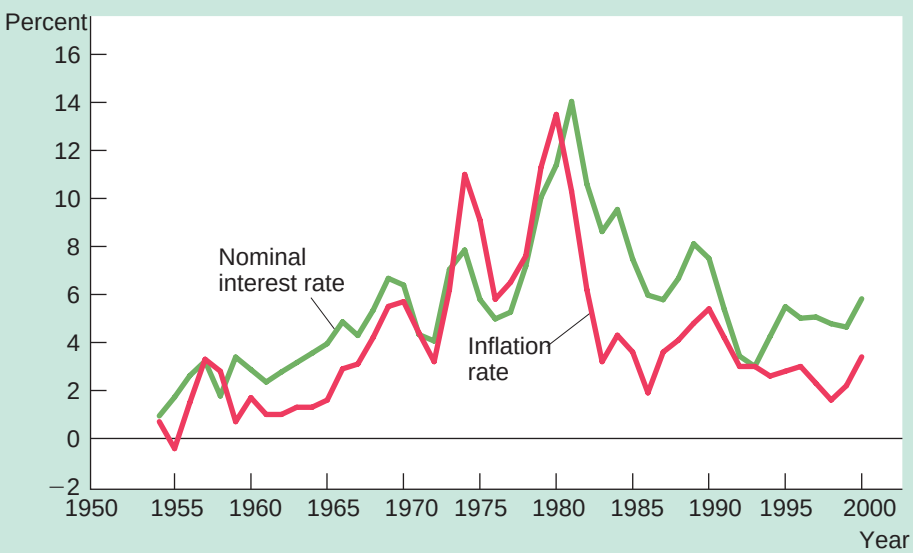
The quantity theory and the Fisher equation together tell us how money growth affects the nominal interest rate. According to the quantity theory, an increase in the rate of money growth of 1 percent causes a 1-percent increase in the rate of inflation. According to the Fisher equation, a 1-percent increase in the rate of inflation causes a 1-percent increase in the nominal interest rate. The one-for-one relation between the inflation rate and the nominal interest rate is called the Fisher effect.

CASE STUDY

Inflation and Nominal Interest Rates

How useful is the Fisher effect in explaining interest rates? To answer this question we look at two types of data on inflation and nominal interest rates.

figure 4-3

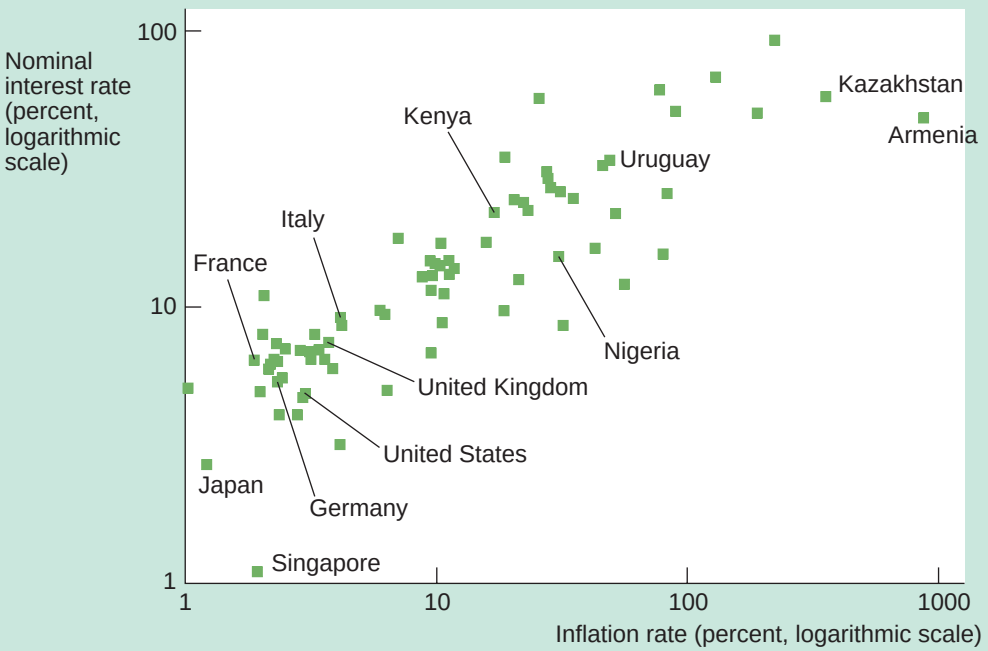


Inflation and Nominal Interest Rates Over Time This figure plots the nominal interest rate (on three-month Treasury bills) and the inflation rate (as measured by the CPI) in the United States since 1954. It shows the Fisher effect: higher inflation leads to a higher nominal interest rate.

Source: Federal Reserve and U.S. Department of Labor.

Figure 4-3 shows the variation over time in the nominal interest rate and the inflation rate in the United States. You can see that the Fisher effect has done a good job explaining fluctuations in the nominal interest rate over the past half century. When inflation is high, nominal interest rates are typically high, and when inflation is low, nominal interest rates are typically low as well.

figure 4-4



Inflation and Nominal Interest Rates Across Countries This scatterplot exhibits the average nominal interest rate on short-term Treasury bills and the average inflation rate in 77 countries during the 1990s. The positive correlation between the inflation rate and the nominal interest rate is evidence for the Fisher effect.

Source: International Financial Statistics.

Similar support for the Fisher effect comes from examining the variation across countries. As Figure 4-4 shows, a nation's inflation rate and its nominal interest rate are related. Countries with high inflation tend to have high nominal interest rates as well, and countries with low inflation tend to have low nominal interest rates.

The link between inflation and interest rates is well known to Wall Street investment firms. Because bond prices move inversely with interest rates, one can get rich by predicting correctly the direction in which interest rates will move. Many Wall Street firms hire Fed watchers to monitor monetary policy and news about inflation in order to anticipate changes in interest rates.

Two Real Interest Rates: *Ex Ante* and *Ex Post*

When a borrower and lender agree on a nominal interest rate, they do not know what the inflation rate over the term of the loan will be. Therefore, we must distinguish between two concepts of the real interest rate: the real interest rate the borrower and lender expect when the loan is made, called the *ex ante* real interest rate, and the real interest rate actually realized, called the *ex post* real interest rate.

Although borrowers and lenders cannot predict future inflation with certainty, they do have some expectation of the inflation rate. Let π denote actual future inflation and π^e the expectation of future inflation. The *ex ante* real interest rate is $r - \pi^e$, and the *ex post* real interest rate is $r - \pi$. The two real interest rates differ when actual inflation π differs from expected inflation π^e.

How does this distinction between actual and expected inflation modify the Fisher effect? Clearly, the nominal interest rate cannot adjust to actual inflation, because actual inflation is not known when the nominal interest rate is set. The nominal interest rate can adjust only to expected inflation. The Fisher effect is more precisely written as

$$i = r + \pi^e.$$

The *ex ante* real interest rate r is determined by equilibrium in the market for goods and services, as described by the model in Chapter 3. The nominal interest rate i moves one-for-one with changes in expected inflation π^e.

CASE STUDY

Nominal Interest Rates in the Nineteenth Century

Although recent data show a positive relationship between nominal interest rates and inflation rates, this finding is not universal. In data from the late nineteenth and early twentieth centuries, high nominal interest rates did not accompany high inflation. The apparent absence of any Fisher effect during this time puzzled Irving Fisher. He suggested that inflation 'caught merchants napping'.

How should we interpret the absence of an apparent Fisher effect in nineteenth-century data? Does this period of history provide evidence against the adjustment of nominal interest rates to inflation? Recent research suggests that this period has little to tell us about the validity of the Fisher effect. The reason is that the Fisher effect relates the nominal interest rate to expected inflation, and, according to this research, inflation at this time was largely unexpected.

Although expectations are not observable, we can draw inferences about them by examining the persistence of inflation. In recent experience, inflation has been very persistent: when it is high one year, it tends to be high the next year as well. Therefore, when people have observed high inflation, it has been rational for them to expect high inflation in the future. By contrast, during the nineteenth century, when the gold standard was in effect, inflation had

little persistence. High inflation in one year was just as likely to be followed the next year by low inflation as by high inflation. Therefore, high inflation did not imply high expected inflation and did not lead to high nominal interest rates. So, in a sense, Fisher was right to say that inflation "caught merchants napping."[6]

4-5 The Nominal Interest Rate and the Demand for Money

The quantity theory is based on a simple money demand function: it assumes that the demand for real money balances is proportional to income. Although the quantity theory is a good place to start when analyzing the effects of money on the economy, it is not the whole story. Here we add another determinant of the quantity of money demanded—the nominal interest rate.

The Cost of Holding Money

The money you hold in your wallet does not earn interest. If, instead of holding that money, you used it to buy government bonds or deposited it in a savings account, you would earn the nominal interest rate. The nominal interest rate is the opportunity cost of holding money: it is what you give up by holding money rather than bonds.

Another way to see that the cost of holding money equals the nominal interest rate is by comparing the real returns on alternative assets. Assets other than money, such as government bonds, earn the real return r. Money earns an expected real return of $-\pi^e$, because its real value declines at the rate of inflation. When you hold money, you give up the difference between these two returns. Thus, the cost of holding money is $i - (-\pi^e)$, which the Fisher equation tells us is the nominal interest rate i.

Just as the quantity of bread demanded depends on the price of bread, the quantity of money demanded depends on the price of holding money. Hence, the demand for real money balances depends both on the level of income and on the nominal interest rate. We write the general money demand function as

$$(M/P)^d = L(i, Y).$$

The letter L is used to denote money demand because money is the economy's most liquid asset (the asset most easily used to make transactions). This equation states that the demand for the liquidity of real money balances is a function of

[6] Robert B. Barsky, "The Fisher Effect and the Forecastability and Persistence of Inflation," Journal of Monetary Economics 19 (January 1987): 3–24.

income and the nominal interest rate. The higher the level of income, the greater the demand for real money balances. The higher the nominal interest rate, the lower the demand for real money balances.

Future Money and Current Prices

Money, prices, and interest rates are now related in several ways. Figure 4-5 illustrates the linkages we have discussed. As the quantity theory of money explains, money supply and money demand together determine the equilibrium price level. Changes in the price level are, by definition, the rate of inflation. Inflation, in turn, affects the nominal interest rate through the Fisher effect. But now, because the nominal interest rate is the cost of holding money, the nominal interest rate feeds back to affect the demand for money.

Consider how the introduction of this last link affects our theory of the price level. First, equate the supply of real money balances M/P to the demand $L(i, Y)$:

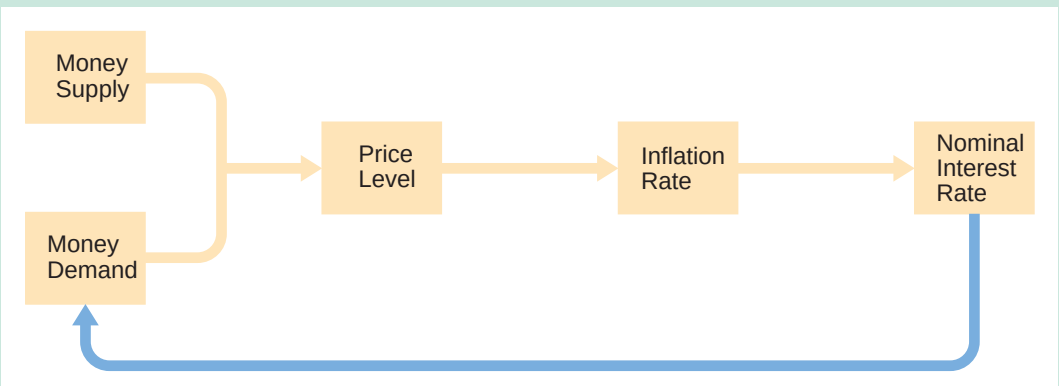
$$M/P = L(i, Y).$$

Next, use the Fisher equation to write the nominal interest rate as the sum of the real interest rate and expected inflation:

$$M/P = L(r + \pi^e, Y).$$

This equation states that the level of real money balances depends on the expected rate of inflation.

figure 4-5



The Linkages Among Money, Prices, and Interest Rates This figure illustrates the relationships among money, prices, and interest rates. Money supply and money demand determine the price level. Changes in the price level determine the inflation rate. The inflation rate influences the nominal interest rate. Because the nominal interest rate is the cost of holding money, it may affect money demand. This last link (shown as a blue line) is omitted from the basic quantity theory of money.

The last equation tells a more sophisticated story than the quantity theory about the determination of the price level. The quantity theory of money says that today's money supply determines today's price level. This conclusion remains partly true: if the nominal interest rate and the level of output are held constant, the price level moves proportionately with the money supply. Yet the nominal interest rate is not constant; it depends on expected inflation, which in turn depends on growth in the money supply. The presence of the nominal interest rate in the money demand function yields an additional channel through which money supply affects the price level.

This general money demand equation implies that the price level depends not only on today's money supply but also on the money supply expected in the future. To see why, suppose the Fed announces that it will raise the money supply in the future, but it does not change the money supply today. This announcement causes people to expect higher money growth and higher inflation. Through the Fisher effect, this increase in expected inflation raises the nominal interest rate. The higher nominal interest rate reduces the demand for real money balances. Because the quantity of money has not changed, the reduced demand for real money balances leads to a higher price level. Hence, higher expected money growth in the future leads to a higher price level today.

The effect of money on prices is complex. The appendix to this chapter works out the mathematics relating the price level to current and future money. The conclusion of the analysis is that the price level depends on a weighted average of the current money supply and the money supply expected to prevail in the future.

4-6 The Social Costs of Inflation

Our discussion of the causes and effects of inflation does not tell us much about the social problems that result from inflation. We turn to those problems now.

The Layman's View and the Classical Response

If you ask the average person why inflation is a social problem, he will probably answer that inflation makes him poorer. "Each year my boss gives me a raise, but prices go up and that takes some of my raise away from me." The implicit assumption in this statement is that if there were no inflation, he would get the same raise and be able to buy more goods.

This complaint about inflation is a common fallacy. As we know from Chapter 3, the purchasing power of labor—the real wage—depends on the marginal productivity of labor, not on how much money the government chooses to print. If the government reduced inflation by slowing the rate of money growth, workers would not see their real wage increasing more rapidly. Instead, when

inflation slowed, firms would increase the prices of their products less each year and, as a result, would give their workers smaller raises.

According to the classical theory of money, a change in the overall price level is like a change in the units of measurement. It is as if we switched from measuring distances in feet to measuring them in inches: numbers get larger, but nothing really changes. Imagine that tomorrow morning you wake up and find that, for some reason, all dollar figures in the economy have been multiplied by ten. The price of everything you buy has increased tenfold, but so has your wage and the value of your savings. What difference would this make? All numbers would have an extra zero at the end, but nothing else would change. Your economic well-being depends on relative prices, not the overall price level.

Why, then, is a persistent increase in the price level a social problem? It turns out that the costs of inflation are subtle. Indeed, economists disagree about the size of the social costs. To the surprise of many laymen, some economists argue that the costs of inflation are small—at least for the moderate rates of inflation that most countries have experienced in recent years.[7]

CASE STUDY

What Economists and the Public Say About Inflation

As we have been discussing, laymen and economists hold very different views about the costs of inflation. Economist Robert Shiller has documented this difference of opinion in a survey of the two groups. The survey results are striking, for they show how the study of economics changes a person's attitudes.

In one question, Shiller asked people whether they agreed that "the biggest gripe about inflation" was that "inflation hurts my real buying power, it makes me poorer." Of the general public, 77 percent agreed with this statement, compared to only 12 percent of economists. Shiller also asked people whether they agreed with the following statement: "When I see projections about how many times more a college education will cost, or how many times more the cost of living will be in coming decades, I feel a sense of uneasiness; these inflation projections really make me worry that my own income will not rise as much as such costs will." Among the general public, 66 percent said they fully agreed with this statement, while only 5 percent of economists agreed with it.

Survey respondents were asked to judge the seriousness of inflation as a policy problem: "Do you agree that preventing high inflation is an important national priority, as important as preventing drug abuse or preventing deterioration in the quality of our schools?" Fifty-two percent of laymen, but only 18 percent of economists, fully agreed with this view. Apparently, inflation worries the public much more than it does the economics profession.

---

[7] See, for example, Chapter 2 of Alan Blinder, Hard Heads, Soft Hearts: Tough-Minded Economics for a Just Society (Reading, MA: Addison Wesley, 1987).

The public's distaste for inflation may be psychological. Shiller asked those surveyed if they agreed with the following statement: "I think that if my pay went up I would feel more satisfaction in my job, more sense of fulfillment, even if prices went up just as much." Of the public, 49 percent fully or partly agreed with this statement, compared to 8 percent of economists.

Do these survey results mean that laymen are wrong and economists are right about the costs of inflation? Not necessarily. But economists do have the advantage of having given the issue more thought. So now consider what some of the costs of inflation might be.[8]

The Costs of Expected Inflation

Consider first the case of expected inflation. Suppose that every month the price level rose by 1 percent. What would be the social costs of such a steady and predictable 12-percent annual inflation?

One cost is the distortion of the inflation tax on the amount of money people hold. As we have already discussed, a higher inflation rate leads to a higher nominal interest rate, which in turn leads to lower real money balances. If people are to hold lower money balances on average, they must make more frequent trips to the bank to withdraw money—for example, they might withdraw $50 twice a week rather than $100 once a week. The inconvenience of reducing money holding is metaphorically called the *shoe-leather cost* of inflation, because walking to the bank more often causes one's shoes to wear out more quickly.

A second cost of inflation arises because high inflation induces firms to change their posted prices more often. Changing prices is sometimes costly: for example, it may require printing and distributing a new catalog. These costs are called *menu costs*, because the higher the rate of inflation, the more often restaurants have to print new menus.

A third cost of inflation arises because firms facing menu costs change prices infrequently; therefore, the higher the rate of inflation, the greater the variability in relative prices. For example, suppose a firm issues a new catalog every January. If there is no inflation, then the firm's prices relative to the overall price level are constant over the year. Yet if inflation is 1 percent per month, then from the beginning to the end of the year the firm's relative prices fall by 12 percent. Sales from this catalog will tend to be low early in the year (when its prices are relatively high) and high later in the year (when its prices are relatively low). Hence, when inflation induces variability in relative prices, it leads to microeconomic inefficiencies in the allocation of resources.

---

[8] Robert J. Shiller, "Why Do People Dislike Inflation?" in Christina D. Romer and David H. Romer, eds., *Reducing Inflation: Motivation and Strategy* (Chicago: University of Chicago Press, 1997).

A fourth cost of inflation results from the tax laws. Many provisions of the tax code do not take into account the effects of inflation. Inflation can alter individuals' tax liability, often in ways that lawmakers did not intend.

One example of the failure of the tax code to deal with inflation is the tax treatment of capital gains. Suppose you buy some stock today and sell it a year from now at the same real price. It would seem reasonable for the government not to levy a tax, because you have earned no real income from this investment. Indeed, if there is no inflation, a zero tax liability would be the outcome. But suppose the rate of inflation is 12 percent and you initially paid $100 per share for the stock; for the real price to be the same a year later, you must sell the stock for $112 per share. In this case the tax code, which ignores the effects of inflation, says that you have earned $12 per share in income, and the government taxes you on this capital gain. The problem, of course, is that the tax code measures income as the nominal rather than the real capital gain. In this example, and in many others, inflation distorts how taxes are levied.

A fifth cost of inflation is the inconvenience of living in a world with a changing price level. Money is the yardstick with which we measure economic transactions. When there is inflation, that yardstick is changing in length. To continue the analogy, suppose that Congress passed a law specifying that a yard would equal 36 inches in 2002, 35 inches in 2003, 34 inches in 2004, and so on. Although the law would result in no ambiguity, it would be highly inconvenient. When someone measured a distance in yards, it would be necessary to specify whether the measurement was in 2002 yards or 2003 yards; to compare distances measured in different years, one would need to make an "inflation" correction. Similarly, the dollar is a less useful measure when its value is always changing.

For example, a changing price level complicates personal financial planning. One important decision that all households face is how much of their income to consume today and how much to save for retirement. A dollar saved today and invested at a fixed nominal interest rate will yield a fixed dollar amount in the future. Yet the real value of that dollar amount—which will determine the retiree's living standard—depends on the future price level. Deciding how much to save would be much simpler if people could count on the price level in 30 years being similar to its level today.

The Costs of Unexpected Inflation

Unexpected inflation has an effect that is more pernicious than any of the costs of steady, anticipated inflation: it arbitrarily redistributes wealth among individuals. You can see how this works by examining long-term loans. Most loan agreements specify a nominal interest rate, which is based on the rate of inflation expected at the time of the agreement. If inflation turns out differently from what was expected, the *ex post* real return that the debtor pays to the creditor differs from what both parties anticipated. On the one hand, if inflation turns out to be higher than expected, the debtor wins and the creditor loses because the debtor repays the loan with less valuable dollars. On the other hand, if inflation turns out to be lower than expected, the creditor wins and the debtor loses because the repayment is worth more than the two parties anticipated.

Consider, for example, a person taking out a mortgage in 1960. At the time, a 30-year mortgage had an interest rate of about 6 percent per year. This rate was based on a low rate of expected inflation—inflation over the previous decade had averaged only 2.5 percent. The creditor probably expected to receive a real return of about 3.5 percent, and the debtor expected to pay this real return. In fact, over the life of the mortgage, the inflation rate averaged 5 percent, so the real return was only 1 percent. This unanticipated inflation benefited the debtor at the expense of the creditor.

Unanticipated inflation also hurts individuals on fixed pensions. Workers and firms often agree on a fixed nominal pension when the worker retires (or even earlier). Because the pension is deferred earnings, the worker is essentially providing the firm a loan: the worker provides labor services to the firm while young but does not get fully paid until old age. Like any creditor, the worker is hurt when inflation is higher than anticipated. Like any debtor, the firm is hurt when inflation is lower than anticipated.

These situations provide a clear argument against variable inflation. The more variable the rate of inflation, the greater the uncertainty that both debtors and creditors face. Because most people are risk averse—they dislike uncertainty—the unpredictability caused by highly variable inflation hurts almost everyone.

Given these effects of uncertain inflation, it is puzzling that nominal contracts are so prevalent. One might expect debtors and creditors to protect themselves from this uncertainty by writing contracts in real terms—that is, by indexing to some measure of the price level. In economies with high and variable inflation, indexation is often widespread; sometimes this indexation takes the form of writing contracts using a more stable foreign currency. In economies with moderate inflation, such as the United States, indexation is less common. Yet even in the United States, some long-term obligations are indexed. For example, Social Security benefits for the elderly are adjusted annually in response to changes in the consumer price index. And in 1997, the U.S. federal government issued inflation-indexed bonds for the first time.

Finally, in thinking about the costs of inflation, it is important to note a widely documented but little understood fact: high inflation is variable inflation. That is, countries with high average inflation also tend to have inflation rates that change greatly from year to year. The implication is that if a country decides to pursue a high-inflation monetary policy, it will likely have to accept highly variable inflation as well. As we have just discussed, highly variable inflation increases uncertainty for both creditors and debtors by subjecting them to arbitrary and potentially large redistributions of wealth.

CASE STUDY

The Free Silver Movement, the Election of 1896, and the Wizard of Oz

The redistributions of wealth caused by unexpected changes in the price level are often a source of political turmoil, as evidenced by the Free Silver movement in the late nineteenth century. From 1880 to 1896 the price level in the

United States fell 23 percent. This deflation was good for creditors, primarily the bankers of the Northeast, but it was bad for debtors, primarily the farmers of the South and West. One proposed solution to this problem was to replace the gold standard with a bimetallic standard, under which both gold and silver could be minted into coin. The move to a bimetallic standard would increase the money supply and stop the deflation.

The silver issue dominated the presidential election of 1896. William McKinley, the Republican nominee, campaigned on a platform of preserving the gold standard. William Jennings Bryan, the Democratic nominee, supported the bimetallic standard. In a famous speech, Bryan proclaimed, "You shall not press down upon the brow of labor this crown of thorns, you shall not crucify mankind upon a cross of gold." Not surprisingly, McKinley was the candidate of the conservative eastern establishment, while Bryan was the candidate of the southern and western populists.

This debate over silver found its most memorable expression in a children's book, *The Wizard of Oz*. Written by a midwestern journalist, L. Frank Baum, just after the 1896 election, it tells the story of Dorothy, a girl lost in a strange land far from her home in Kansas. Dorothy (representing traditional American values) makes three friends: a scarecrow (the farmer), a tin woodman (the industrial worker), and a lion whose roar exceeds his might (William Jennings Bryan). Together, the four of them make their way along a perilous yellow brick road (the gold standard), hoping to find the Wizard who will help Dorothy return home. Eventually they arrive in Oz (Washington), where everyone sees the world through green glasses (money). The Wizard (William McKinley) tries to be all things to all people but turns out to be a fraud. Dorothy's problem is solved only when she learns about the magical power of her silver slippers.[9]

Although the Republicans won the election of 1896 and the United States stayed on a gold standard, the Free Silver advocates got some of what they wanted. Around the time of the election, gold was discovered in Alaska, Australia, and South Africa. In addition, gold miners devised the cyanide process, which facilitated the extraction of gold from ore. These developments led to increases in the money supply and in prices. From 1896 to 1910 the price level rose 35 percent.

One Benefit of Inflation

So far, we have discussed the many costs of inflation. These costs lead many economists to conclude that monetary policymakers should aim for zero inflation. Yet there is another side to the story. Some economists believe that a little bit of inflation—say, 2 or 3 percent per year—can be a good thing.

---

[9] The movie made 40 years later hid much of the allegory by changing Dorothy's shoes from silver to ruby. For more on this topic, see Henry M. Littlefield, "The Wizard of Oz: Parable on Populism," *American Quarterly* 16 (Spring 1964): 47–58; and Hugh Rockoff, "The Wizard of Oz as a Monetary Allegory," *Journal of Political Economy* 98 (August 1990): 739–760.

FYI

Keynes (and Lenin) on the Cost of Inflation

The great economist John Maynard Keynes was no friend of inflation, as this chapter's opening quotation indicates. Here is the more complete passage from his famous book, *The Economic Consequences of the Peace*, in which Keynes predicted (correctly) that the treaty imposed on Germany after World War I would lead to economic hardship and renewed international tensions:

> Lenin is said to have declared that the best way to destroy the Capitalist System was to debauch the currency. By a continuing process of inflation, governments can confiscate, secretly and unobserved, an important part of the wealth of their citizens. By this method they not only confiscate, but they confiscate arbitrarily and, while the process impoverishes many, it actually enriches some. The sight of this arbitrary rearrangement of riches strikes not only at security, but at confidence in the equity of the existing distribution of wealth. Those to whom the system brings windfalls, beyond their deserts and even beyond their expectations or desires, become "profiteers," who are the object of the hatred of the bourgeoisie whom the inflationism has impoverished, not less than of the proletariat. As the inflation proceeds and the real value of the currency fluctuates wildly from month to month, all permanent relations between debtors and creditors, which form the ultimate foundation of capitalism, become so utterly disordered as to be almost meaningless; and the process of wealth-getting degenerates into a gamble and a lottery.
>
> Lenin was certainly right. There is no subtler, no surer means of overturning the existing basis of society than to debauch the currency. The process engages all the hidden forces of economic law on the side of destruction, and does it in a manner which not one man in a million is able to diagnose.[10]

History has given ample support to this assessment. A recent example occurred in Russia in 1998, where many citizens saw high rates of inflation wipe out their ruble-denominated savings. And, as Lenin would have predicted, this inflation put the country's burgeoning capitalist system in serious jeopardy.

The argument for moderate inflation starts with the observation that cuts in nominal wages are rare. Firms are reluctant to cut their workers' nominal wages, and workers are reluctant to accept such cuts. A 2-percent wage cut in a zero-inflation world is, in real terms, the same as a 3-percent raise with 5-percent inflation, but workers do not always see it that way. The 2-percent wage cut may seem like an insult, whereas the 3-percent raise is, after all, still a raise. Empirical studies confirm that nominal wages rarely fall.

This fact suggests that some inflation may make labor markets work better. The supply and demand for different kinds of labor are always changing. Sometimes an increase in supply or decrease in demand leads to a fall in the equilibrium real wage for a group of workers. If nominal wages don't be cut, then the only way to cut real wages is to allow inflation to do the job. Without inflation, the real wage will be stuck above the equilibrium level, resulting in higher unemployment.

For this reason, some economists argue that inflation "greases the wheels" of labor markets. Only a little inflation is needed: an inflation rate of 2 percent

[10] John Maynard Keynes, *The Economic Consequences of the Peace* (London: Macmillan, 1920): 219–220.

lets real wages fall by 2 percent per year, or 20 percent per decade, without cuts in nominal wages. Such automatic reductions in real wages are impossible with zero inflation.[11]

4-7 Hyperinflation

Hyperinflation is often defined as inflation that exceeds 50 percent per month, which is just over 1 percent per day. Compounded over many months, this rate of inflation leads to very large increases in the price level. An inflation rate of 50 percent per month implies a more than 100-fold increase in the price level over a year, and a more than 2-million-fold increase over three years. Here we consider the costs and causes of such extreme inflation.

The Costs of Hyperinflation

Although economists debate whether the costs of moderate inflation are large or small, no one doubts that hyperinflation extracts a high toll on society. The costs are qualitatively the same as those we discussed earlier. When inflation reaches extreme levels, however, these costs are more apparent because they are so severe.

The shoeleather costs associated with reduced money holding, for instance, are serious under hyperinflation. Business executives devote much time and energy to cash management when cash loses its value quickly. By diverting this time and energy from more socially valuable activities, such as production and investment decisions, hyperinflation makes the economy run less efficiently.

Menu costs also become larger under hyperinflation. Firms have to change prices so often that normal business practices, such as printing and distributing catalogs with fixed prices, become impossible. In one restaurant during the German hyperinflation of the 1920s, a waiter would stand up on a table every 30 minutes to call out the new prices.

Similarly, relative prices do not do a good job of reflecting true scarcity during hyperinflations. When prices change frequently by large amounts, it is hard for customers to shop around for the best price. Highly volatile and rapidly rising prices can alter behavior in many ways. According to one report, when patrons entered a pub during the German hyperinflation, they would often buy two pitchers of beer. Although the second pitcher would lose value by getting warm over time, it would lose value less rapidly than the money left sitting in the patron's wallet.

---

[11] For a recent paper examining this benefit of inflation, see George A. Akerlof, William T. Dickens, and George L. Perry, "The Macroeconomics of Low Inflation," Brookings Papers on Economic Activity 1996:1, pp. 1–76.

Tax systems are also distorted by hyperinflation—but in ways that are quite different than under moderate inflation. In most tax systems there is a delay between the time a tax is levied and the time the tax is paid to the government. In the United States, for example, taxpayers are required to make estimated income tax payments every three months. This short delay does not matter much under low inflation. By contrast, during hyperinflation, even a short delay greatly reduces real tax revenue. By the time the government gets the money it is due, the money has fallen in value. As a result, once hyperinflation starts, the real tax revenue of the government often falls substantially.

Finally, no one should underestimate the sheer inconvenience of living with hyperinflation. When carrying money to the grocery store is as burdensome as carrying the groceries back home, the monetary system is not doing its best to facilitate exchange. The government tries to overcome this problem by adding more and more zeros to the paper currency, but often it cannot keep up with the exploding price level.

Eventually, these costs of hyperinflation become intolerable. Over time, money loses its role as a store of value, unit of account, and medium of exchange. Barter becomes more common. And more stable foreign monies—cigarettes or the U.S. dollar—start to replace the official money.

CASE STUDY

Life During the Bolivian Hyperinflation

The following article from the *Wall Street Journal* shows what life was like during the Bolivian hyperinflation of 1985.[12] What costs of inflation does this article emphasize?

> Precarious Peso—Amid Wild Inflation, Bolivians Concentrate on Swapping Currency
>
> LA PAZ, Bolivia—When Edgar Miranda gets his monthly teacher's pay of 25 million pesos, he hasn't a moment to lose. Every hour, pesos drop in value. So, while his wife rushes to market to lay in a month's supply of rice and noodles, he is off with the rest of the pesos to change them into black-market dollars.
>
> Mr. Miranda is practicing the First Rule of Survival amid the most out-of-control inflation in the world today. Bolivia is a case study of how runaway inflation undermines a society. Price increases are so huge that figures build up almost beyond comprehension. In one six-month period, for example, prices soared at an annual rate of 38,000%. By official count, however, last year's inflation reached 2,000%, and this year's is expected to hit 8,000%—although other estimates range many times higher. In any event, Bolivia's rate dwarfs Israel's 370% and Argentina's 1,100%—two other cases of severe inflation.

[12] Reprinted by permission of the *Wall Street Journal*, © August 13, 1985, page 1, Dow Jones & Company, Inc. All rights reserved worldwide.

It is easier to comprehend what happens to the 38-year-old Mr. Mosqueira if he doesn't quickly change it into dollars. The day he was paid 25 million pesos, a dollar cost 500,000 pesos. So he received $50. Just days later, with the rate at 900,000 pesos, he would have received $27.

"We think only about today and converting every peso into dollars," says Ronald MacLean, the manager of a gold-mining firm. "We have become myopic."

And intent on survival. Civil servants won't hand out a form without a bribe. Lawyers, accountants, hairdressers, even prostitutes have almost given up working to become money-changers in the streets. Workers stage repeated strikes and steal from their bosses. The bosses smuggle production abroad, take out phony loans, duck taxes—anything to get dollars for speculation.

The production at the state mines, for example, dropped to 12,000 tons last year from 18,000. The miners pad their wages by smuggling out the richest ore in their lunch pails, and the ore goes by a contraband network into neighboring Peru. Without a major tin mine, Peru now exports some 4,000 metric tons of tin a year.

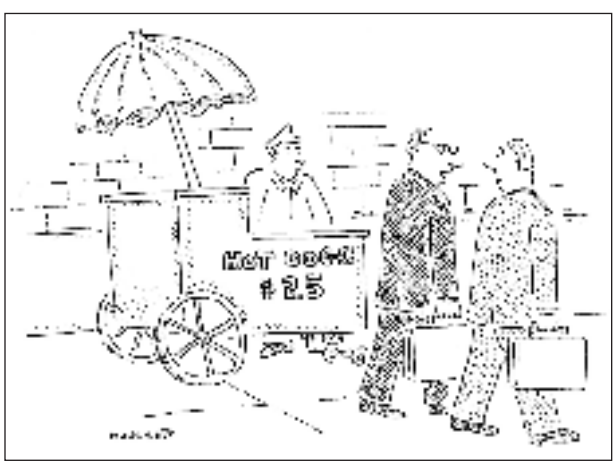
"We don't produce anything. We are all currency speculators," says a heavy-equipment dealer in La Paz. "People don't know what's good and bad anymore. We have become an amoral society."

It is an open secret that practically all of the black-market dollars come from the illegal cocaine trade with the U.S. Cocaine traffickers earn an estimated $1 billion a year. . . .

But meanwhile the country is suffering from inflation largely because the government's revenues cover a mere 15% of its expenditures and its deficit has widened to nearly 25% of the country's total annual output. The revenues are hurt by a lag in tax payments, and taxes aren't being collected largely because of widespread theft and bribery.

The Causes of Hyperinflation

Why do hyperinflations start, and how do they end? This question can be answered at different levels.



© The New Yorker Collection 1997 Robert Mankoff from cartoonbank.com. All Rights Reserved.

"I told you the Fed should have tightened."

The most obvious answer is that hyperinflations are caused by excessive growth in the supply of money. When the central bank prints money, the price level rises. When it prints money rapidly enough, the result is hyperinflation. To stop the hyperinflation, the central bank must reduce the rate of money growth.

This answer is incomplete, however, for it leaves open the question of why central banks in hyperinflating economies choose to print so much money. To address this deeper question, we must turn our attention from monetary to fiscal policy. Most hyperinflations begin when the government has inadequate tax

revenue to pay for its spending. Although the government might prefer to finance this budget deficit by issuing debt, it may find itself unable to borrow, perhaps because lenders view the government as a bad credit risk. To cover the deficit, the government turns to the only mechanism at its disposal—printing press. The result is rapid money growth and hyperinflation.

Once the hyperinflation is under way, the fiscal problems become even more severe. Because of the delay in collecting tax payments, real tax revenue falls as inflation rises. Thus, the government's need to rely on seigniorage is self-reinforcing. Rapid money creation leads to hyperinflation, which leads to a larger budget deficit, which leads to even more rapid money creation.

The ends of hyperinflations almost always coincide with fiscal reforms. Once the magnitude of the problem becomes apparent, the government musters the political will to reduce government spending and increase taxes. These fiscal reforms reduce the need for seigniorage, which allows a reduction in money growth. Hence, even if inflation is always and everywhere a monetary phenomenon, the end of hyperinflation is often a fiscal phenomenon as well.[13]

CASE STUDY

Hyperinflation in Interwar Germany

After World War I, Germany experienced one of history's most spectacular examples of hyperinflation. At the war's end, the Allies demanded that Germany pay substantial reparations. These payments led to budget deficits in Germany, which the German government eventually financed by printing large quantities of money.

Panel (a) of Figure 4-6 shows the quantity of money and the general price level in Germany from January 1922 to December 1924. During this period both money and prices rose at an amazing rate. For example, the price of a daily newspaper rose from 0.30 marks in January 1921 to 1 mark in May 1922, to 8 marks in October 1922, to 100 marks in February 1923, and to 1,000 marks in September 1923. Then, in the fall of 1923, prices took off: the newspaper sold for 2,000 marks on October 1, 20,000 marks on October 15, 1 million marks on October 29, 15 million marks on November 9, and 70 million marks on November 17. In December 1923 the money supply and prices abruptly stabilized.[14]

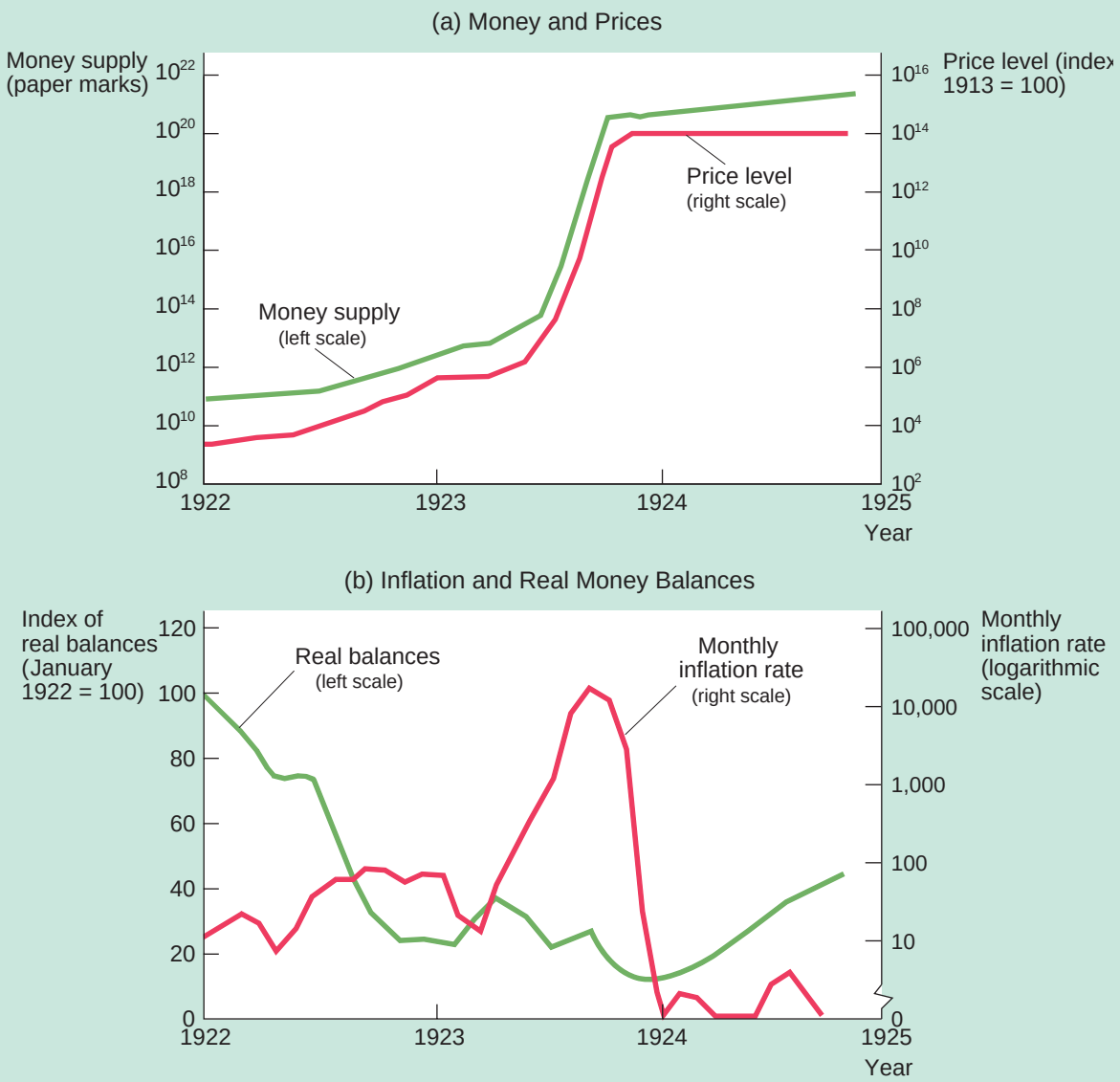
Just as fiscal problems caused the German hyperinflation, a fiscal reform ended it. At the end of 1923, the number of government employees was cut by one-third, and the reparations payments were temporarily suspended and eventually

---

[13] For more on these issues, see Thomas J. Sargent, "The End of Four Big Inflations," in Robert Hall, ed., *Inflation* (Chicago: University of Chicago Press, 1983), 41–93; and Rudiger Dornbusch and Stanley Fischer, "Stopping Hyperinflations: Past and Present," *Weltwirtschaftliches Archiv* 122 (April 1986): 1–47.

[14] The data on newspaper prices are from Michael M. Bomberger and Gail J. Makinen, "Inflation, Money Growth, and Stock Prices in Weimar Germany," *Journal of Political Economy* 91 (October 1983): 801–824. The data on individual prices are from Michael Mussa, "Sticky Individual Prices and the Dynamics of the General Price Level," *Carnegie-Rochester Conference on Public Policy* 15 (Autumn 1981): 261–296.

figure 4-6



Money and Prices in Interwar Germany Panel (a) shows the money supply and the price level in Germany from January 1922 to December 1924. The immense increases in the money supply and the price level provide a dramatic illustration of the effects of printing large amounts of money. Panel (b) shows inflation and real money balances. As inflation rose, real money balances fell. When the inflation ended at the end of 1923, real money balances rose.

Source: Adapted from Thomas J. Sargent, "The End of Four Big Inflations," in Robert Hall, ed., *Inflation* (Chicago: University of Chicago Press, 1983): 41–93.

reduced. At the same time, a new central bank, the Rentenbank, replaced the old central bank, the Reichsbank. The Rentenbank was committed to not financing the government by printing money.

According to our theoretical analysis of money demand, an end to a hyperinflation should lead to an increase in real money balances as the cost of holding money falls. Panel (b) of Figure 4-6 shows that real money balances in Germany did fall as inflation increased, and then increased again as inflation fell. Yet the increase in real money balances was not immediate. Perhaps the adjustment of real money balances to the cost of holding money is a gradual process. Or perhaps it took time for people in Germany to believe that the inflation had ended, so that expected inflation fell more gradually than actual inflation.

4-8 Conclusion: The Classical Dichotomy

We have finished our discussion of money and inflation. Let's now step back and examine a key assumption that has been implicit in our discussion.

In Chapter 3, we explained many macroeconomic variables. Some of these variables were quantities, such as real GDP and the capital stock; others were relative prices, such as the real wage and the real interest rate. But all of these variables had one thing in common—they measured a physical (rather than a monetary) quantity. Real GDP is the quantity of goods and services produced in a given year, and the capital stock is the quantity of machines and structures available at a given time. The real wage is the quantity of output a worker earns for each hour of work, and the real interest rate is the quantity of output a person earns in the future by lending one unit of output today. All variables measured in physical units, such as quantities and relative prices, are called **real variables**.

In this chapter we examined **nominal variables**—variables expressed in terms of money. The economy has many nominal variables, such as the price level, the inflation rate, and the dollar wage a person earns.

At first it may seem surprising that we were able to explain real variables without introducing nominal variables or the existence of money. In Chapter 3 we studied the level and allocation of the economy's output without mentioning the price level or the rate of inflation. Our theory of the labor market explained the real wage without explaining the nominal wage.

Economists call this theoretical separation of real and nominal variables the **classical dichotomy**. It is the hallmark of classical macroeconomic theory. The classical dichotomy is an important insight, because it simplifies economic theory. In particular, it allows us to examine real variables, as we have done, while ignoring nominal variables. The classical dichotomy arises because, in classical economic theory, changes in the money supply do not influence real variables. This irrelevance of money for real variables is called **monetary neutrality**. For many purposes—in particular for studying long-run issues—monetary neutrality is approximately correct.

Yet monetary neutrality does not fully describe the world in which we live. Beginning in Chapter 9, we discuss departures from the classical model and monetary neutrality. These departures are crucial for understanding many macroeconomic phenomena, such as short-run economic fluctuations.

Summary

1. Money is the stock of assets used for transactions. It serves as a store of value, a unit of account, and a medium of exchange. Different sorts of assets are used as money: commodity money systems use an asset with intrinsic value, whereas fiat money systems use an asset whose sole function is to serve as money. In modern economies, a central bank such as the Federal Reserve is responsible for controlling the supply of money.
2. The quantity theory of money assumes that the velocity of money is stable and concludes that nominal GDP is proportional to the stock of money. Because the factors of production and the production function determine real GDP, the quantity theory implies that the price level is proportional to the quantity of money. Therefore, the rate of growth in the quantity of money determines the inflation rate.
3. Seigniorage is the revenue that the government raises by printing money. It is a tax on money holding. Although seigniorage is quantitatively small in most economies, it is often a major source of government revenue in economies experiencing hyperinflation.
4. The nominal interest rate is the sum of the real interest rate and the inflation rate. The Fisher effect says that the nominal interest rate moves one-for-one with expected inflation.
5. The nominal interest rate is the opportunity cost of holding money. Thus, one might expect the demand for money to depend on the nominal interest rate. If it does, then the price level depends on both the current quantity of money and the quantities of money expected in the future.
6. The costs of expected inflation include shoeleather costs, menu costs, the cost of relative price variability, tax distortions, and the inconvenience of making inflation corrections. In addition, unexpected inflation causes arbitrary redistributions of wealth between debtors and creditors. One possible benefit of inflation is that it improves the functioning of labor markets by allowing real wages to reach equilibrium levels without cuts in nominal wages.
7. During hyperinflations, most of the costs of inflation become severe. Hyperinflations typically begin when governments finance large budget deficits by printing money. They end when fiscal reforms eliminate the need for seigniorage.
8. According to classical economic theory, money is neutral: the money supply does not affect real variables. Therefore, classical theory allows us to study

how real variables are determined without any reference to the money supply. The equilibrium in the money market then determines the price level and, as a result, all other nominal variables. This theoretical separation of real and nominal variables is called the classical dichotomy.

KEY CONCEPTS

Inflation
Hyperinflation
Money
Store of value
Unit of account
Medium of exchange
Fiat money
Commodity money
Gold standard
Money supply
Monetary policy
Central bank
Federal Reserve
Open-market operations
Currency
Demand deposits
Quantity equation
Transactions velocity of money
Income velocity of money
Real money balances
Money demand function
Quantity theory of money
Seigniorage
Nominal and real interest rates
Fisher equation and Fisher effect
Ex ante and ex post real interest rates
Shoeleather costs
Menu costs
Real and nominal variables
Classical dichotomy
Monetary neutrality

QUESTIONS FOR REVIEW

1. Describe the functions of money.
2. What is fiat money? What is commodity money?
3. Who controls the money supply and how?
4. Write the quantity equation and explain it.
5. What does the assumption of constant velocity imply?
6. Who pays the inflation tax?
7. If inflation rises from 6 to 8 percent, what happens to real and nominal interest rates according to the Fisher effect?
8. List all the costs of inflation you can think of, and rank them according to how important you think they are.
9. Explain the roles of monetary and fiscal policy in causing and ending hyperinflations.
10. Define the terms "real variable" and "nominal variable," and give an example of each.

PROBLEMS AND APPLICATIONS

1. What are the three functions of money? Which of the functions do the following items satisfy? Which do they not satisfy?
   a. A credit card
   b. A painting by Rembrandt
   c. A subway token
2. In the country of Wiknam, the velocity of money is constant. Real GDP grows by 5 percent per year, the money stock grows by 14 percent per year, and the nominal interest rate is 11 percent. What is the real interest rate?
3. A newspaper article written by the Associated Press in 1994 reported that the U.S. economy

was experiencing a low rate of inflation. It said that "low inflation has a downside: 45 million recipients of Social Security and other benefits will see their checks go up by just 2.8 percent next year".

a. Why does inflation affect the increase in Social Security and other benefits?

b. Is this effect a cost of inflation as the article suggests? Why or why not?

4. Suppose you are advising a small country (such as Bermuda) on whether to print its own money or to use the money of its larger neighbor (such as the United States). What are the costs and benefits of a national money? Does the relative political stability of the two countries have any role in this decision?

5. During World War II, both Germany and England had plans for a paper weapon: they each printed the other's currency, with the intention of dropping large quantities by airplane. Why might this have been an effective weapon?

6. Calvin Coolidge once said that "inflation is repudiation." What might he have meant by this? Do you agree? Why or why not? Does it matter whether the inflation is expected or unexpected?

7. Some economic historians have noted that during the period of the gold standard, gold discoveries were most likely to occur after a long deflation. (The discoveries of 1896 are an example.) Why might this be true?

8. Suppose that consumption depends on the level of real money balances (on the grounds that real money balances are part of wealth). Show that if real money balances depend on the nominal interest rate, then an increase in the rate of money growth affects consumption, investment, and the real interest rate. Does the nominal interest rate adjust more than one-for-one or less than one-for-one to expected inflation?

   This deviation from the classical dichotomy and the Fisher effect is called the "Mundell–Tobin effect". How might you decide whether the Mundell–Tobin effect is important in practice?

9. Use the Internet to identify a country that has had high inflation over the past year and another country that has had low inflation. (Hint: One useful Web site is www.economist.com) For these two countries, find the rate of money growth and the current level of the nominal interest rate. Relate your findings to the theories presented in this chapter.

APPENDIX

The Cagan Model: How Current and Future Money Affect the Price Level

In this chapter we showed that if the quantity of real money balances demanded depends on the cost of holding money, the price level depends on both the current money supply and the future money supply. This appendix develops the Cagan model[15] to show more explicitly how this works.

To keep the math as simple as possible, we posit a money demand function that is linear in the natural logarithms of all the variables. The money demand function is

$$m_t - p_t = -\gamma(p_{t+1} - p_t), \tag{A1}$$

where m_t is the log of the quantity of money at time t, p_t is the log of the price level at time t, and γ is a parameter that governs the sensitivity of money demand to the rate of inflation. By the property of logarithms, $m_t - p_t$ is the log of real money balances, and $p_{t+1} - p_t$ is the inflation rate between period t and period $t + 1$. This equation states that if inflation goes up by 1 percentage point, real money balances fall by γ percent.

We have made a number of assumptions in writing the money demand function in this way. First, by excluding the level of output as a determinant of money demand, we are implicitly assuming that it is constant. Second, by including the rate of inflation rather than the nominal interest rate, we are assuming that the real interest rate is constant. Third, by including actual inflation rather than expected inflation, we are assuming perfect foresight. All of these assumptions are to keep the analysis as simple as possible.

We want to solve Equation A1 to express the price level as a function of current and future money. To do this, note that Equation A1 can be rewritten as

$$p_t = \left(\frac{1}{1+\gamma}\right) m_t + \left(\frac{\gamma}{1+\gamma}\right) p_{t+1}. \tag{A2}$$

This equation states that the current price level, p_t, is a weighted average of the current money supply, m_t, and the next period's price level, p_{t+1}. The next period's price level will be determined the same way as this period's price level:

$$p_{t+1} = \left(\frac{1}{1+\gamma}\right) m_{t+1} + \left(\frac{\gamma}{1+\gamma}\right) p_{t+2}. \tag{A3}$$

Now substitute Equation A3 for p_{t+1} in Equation A2 to obtain

$$p_t = \frac{1}{1+\gamma}\, m_t + \frac{\gamma}{(1+\gamma)^2}\, m_{t+1} + \frac{\gamma^2}{(1+\gamma)^2}\, p_{t+2}. \tag{A4}$$

[15] This model is derived from Phillip Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *Studies in the Quantity Theory of Money* (Chicago: University of Chicago Press, 1956).

Equation A4 states that the current price level is a weighted average of the current money supply, the next period's money supply, and the following period's price level. Once again, the price level in $t + 2$ is determined as in Equation A2:

$$p_{t+2} = \left(\frac{1}{1+\gamma}\right) m_{t+2} + \left(\frac{\gamma}{1+\gamma}\right) p_{t+3}. \tag{A5}$$

Now substitute Equation A5 into Equation A4 to obtain

$$p_t = \frac{1}{1+\gamma}\, m_t + \frac{\gamma}{(1+\gamma)^2}\, m_{t+1} + \frac{\gamma^2}{(1+\gamma)^3}\, m_{t+2} + \frac{\gamma^3}{(1+\gamma)^3}\, p_{t+3}. \tag{A6}$$

By now you see the pattern. We can continue to use Equation A2 to substitute for the future price level. If we do this an infinite number of times, we find

$$p_t = \left(\frac{1}{1+\gamma}\right)\left[m_t + \left(\frac{\gamma}{1+\gamma}\right) m_{t+1} + \left(\frac{\gamma}{1+\gamma}\right)^2 m_{t+2} + \left(\frac{\gamma}{1+\gamma}\right)^3 m_{t+3} + \cdots\right], \tag{A7}$$

where $\cdots$ indicates an infinite number of analogous terms. According to Equation A7, the current price level is a weighted average of the current money supply and all future money supplies.

Note the importance of γ, the parameter governing the sensitivity of real money balances to inflation. The weights on the future money supplies decline geometrically at rate $\gamma/(1 + \gamma)$. If γ is small, then $\gamma/(1 + \gamma)$ is small, and the weights decline quickly. In this case, the current money supply is the primary determinant of the price level. (Indeed, if γ equals zero, then we obtain the quantity theory of money: the price level is proportional to the current money supply, and the future money supplies do not matter at all.) If γ is large, then $\gamma/(1 + \gamma)$ is close to 1, and the weights decline slowly. In this case, the future money supplies play a key role in determining today's price level.

Finally, let's relax the assumption of perfect foresight. If the future is not known with certainty, then we should write the money demand function as

$$m_t - p_t = -\gamma(Ep_{t+1} - p_t), \tag{A8}$$

where Ep_{t+1} is the expected price level. Equation A8 states that real money balances depend on expected inflation. By following steps similar to those preceding, we can show that

$$p_t = \left(\frac{1}{1+\gamma}\right)\left[m_t + \left(\frac{\gamma}{1+\gamma}\right) Em_{t+1} + \left(\frac{\gamma}{1+\gamma}\right)^2 Em_{t+2} + \left(\frac{\gamma}{1+\gamma}\right)^3 Em_{t+3} + \cdots\right]. \tag{A9}$$

Equation A9 states that the price level depends on the current money supply and expected future money supplies.

Some economists use this model to argue that credibility is important for ending hyperinflation. Because the price level depends on both current and expected future money, inflation depends on both current and expected future money growth. Therefore, to end high inflation, both money growth and expected money growth must fall. Expectations, in turn, depend on credibility—the perception that the central bank is committed to a new, more stable policy.

How can a central bank achieve credibility in the midst of hyperinflation? Credibility is often achieved by removing the underlying cause of the hyperinflation—the need for seigniorage. Thus, a credible fiscal reform is often necessary for a credible change in monetary policy. This fiscal reform might take the form of reducing government spending and making the central bank more independent from the government. Reduced spending decreases the need for seigniorage, while increased independence allows the central bank to resist government demands for seigniorage.

MORE PROBLEMS AND APPLICATIONS

1. In the Cagan model, if the money supply is expected to grow at some constant rate μ (so that $E_t m_{t+s} = m_t + s\mu$), then Equation A9 can be shown to imply that $p_t = m_t + \gamma\mu$.
   a. Interpret this result.
   b. What happens to the price level p_t when the money supply m_t changes, holding the money growth rate μ constant?
   c. What happens to the price level p_t when the money growth rate μ changes, holding the current money supply m_t constant?
   d. If a central bank is about to reduce the rate of money growth μ but wants to hold the price level p_t constant, what should it do with m_t? Can you see any practical problems that might arise in following such a policy?
   e. How do your previous answers change in the special case where money demand does not depend on the expected rate of inflation (so that $\gamma = 0$)?

CHAPTER 5 FIVE

The Open Economy

> No nation was ever ruined by trade.
>
> — Benjamin Franklin

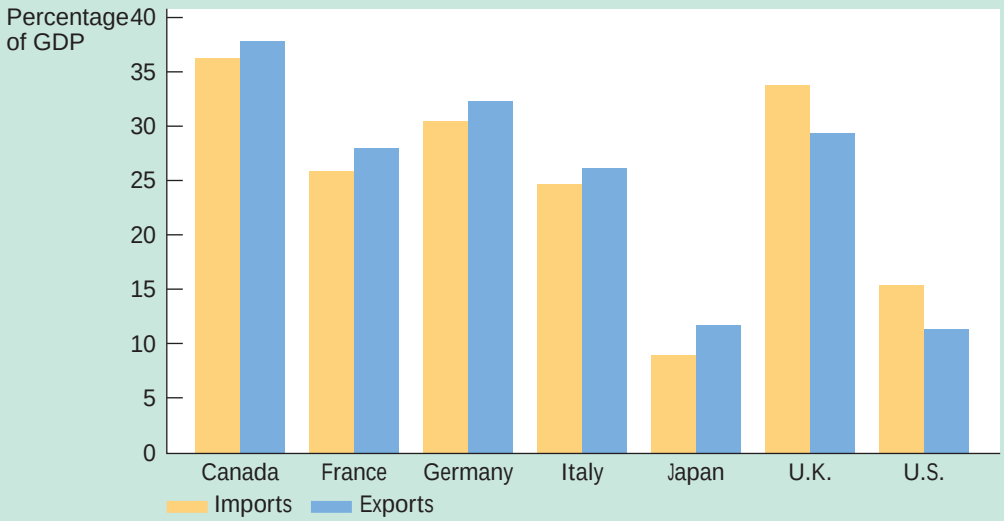
Even if you never leave your home town, you are an active participant in a global economy. When you go to the grocery store, for instance, you might choose between apples grown locally and grapes grown in Chile. When you make a deposit into your local bank, the bank might lend those funds to your next-door neighbor or to a Japanese company building a factory outside Tokyo. Because our economy is integrated with many others around the world, consumers have more goods and services from which to choose, and savers have more opportunities to invest their wealth.

In previous chapters we simpliÞed our analysis by assuming a closed economy. In actuality, however, most economies are open: they export goods and services abroad, they import goods and services from abroad, and they borrow and lend in world Þnancial markets. Figure 5-1 gives some sense of the importance of these international interactions by showing imports and exports as a percentage of GDP for seven major industrial countries. As the Þgure shows, imports and exports in the United States are more than 10 percent of GDP. Trade is even more important for many other countriesÑin Canada and the United Kingdom, for instance, imports and exports are about a third of GDP. In these countries, international trade is central to analyzing economic developments and formulating economic policies.

This chapter begins our study of open-economy macroeconomics. We begin in Section 5-1 with questions of measurement. To understand how the open economy works, we must understand the key macroeconomic variables that measure the interactions among countries. Accounting identities reveal a key insight: the ßow of goods and services across national borders is always matched by an equivalent ßow of funds to Þnance capital accumulation.

In Section 5-2 we examine the determinants of these international ßows. We develop a model of the small open economy that corresponds to our model of the closed economy in Chapter 3. The model shows the factors that determine whether a country is a borrower or a lender in world markets, and how policies at home and abroad affect the ßows of capital and goods.

figure 5-1



Imports and Exports as a Percentage of Output: 2000 While international trade is important for the United States, it is even more vital for other countries.

Source: OECD.

In Section 5-3 we extend the model to discuss the prices at which a country makes exchanges in world markets. We examine what determines the price of domestic goods relative to foreign goods. We also examine what determines the rate at which the domestic currency trades for foreign currencies. Our model shows how protectionist trade policies—policies designed to protect domestic industries from foreign competition—influence the amount of international trade and the exchange rate.

5-1 The International Flows of Capital and Goods

The key macroeconomic difference between open and closed economies is that, in an open economy, a country's spending in any given year need not equal its output of goods and services. A country can spend more than it produces by borrowing from abroad, or it can spend less than it produces and lend the difference to foreigners. To understand this more fully, let's take another look at national income accounting, which we first discussed in Chapter 2.

The Role of Net Exports

Consider the expenditure on an economy's output of goods and services. In a closed economy, all output is sold domestically, and expenditure is divided into three components: consumption, investment, and government purchases. In an open economy, some output is sold domestically and some is exported to be sold abroad. We can divide expenditure on an open economy's output Y into four components:

- C^d, consumption of domestic goods and services,
- I^d, investment in domestic goods and services,
- G^d, government purchases of domestic goods and services,
- EX, exports of domestic goods and services.

The division of expenditure into these components is expressed in the identity

$$Y = C^d + I^d + G^d + EX.$$

The sum of the first three terms, $C^d + I^d + G^d$, is domestic spending on domestic goods and services. The fourth term, EX, is foreign spending on domestic goods and services.

We now want to make this identity more useful. To do this, note that domestic spending on all goods and services is the sum of domestic spending on domestic goods and services and on foreign goods and services. Hence, total consumption C equals consumption of domestic goods and services C^d plus consumption of foreign goods and services C^f; total investment I equals investment in domestic goods and services I^d plus investment in foreign goods and services I^f; and total government purchases G equals government purchases of domestic goods and services G^d plus government purchases of foreign goods and services G^f. Thus,

$$\begin{aligned} C &= C^d + C^f, \\ I &= I^d + I^f, \\ G &= G^d + G^f. \end{aligned}$$

We substitute these three equations into the identity above:

$$Y = (C - C^f) + (I - I^f) + (G - G^f) + EX.$$

We can rearrange to obtain

$$Y = C + I + G + EX - (C^f + I^f + G^f).$$

The sum of domestic spending on foreign goods and services $(C^f + I^f + G^f)$ is expenditure on imports (IM). We can thus write the national income accounts identity as

$$Y = C + I + G + EX - IM.$$

Because spending on imports is included in domestic spending $(C + I + G)$, and because goods and services imported from abroad are not part of a country

output, this equation subtracts spending on imports. Defining net exports to be exports minus imports ($NX = EX - IM$), the identity becomes

$$Y = C + I + G + NX.$$

This equation states that expenditure on domestic output is the sum of consumption, investment, government purchases, and net exports. This is the most common form of the national income accounts identity; it should be familiar from Chapter 2.

The national income accounts identity shows how domestic output, domestic spending, and net exports are related. In particular,

$$NX = Y - (C + I + G)$$
$$\text{Net Exports} = \text{Output} - \text{Domestic Spending}.$$

This equation shows that in an open economy, domestic spending need not equal the output of goods and services. If output exceeds domestic spending, we export the difference: net exports are positive. If output falls short of domestic spending, we import the difference: net exports are negative.

International Capital Flows and the Trade Balance

In an open economy, as in the closed economy we discussed in Chapter 3, financial markets and goods markets are closely related. To see the relationship, we must rewrite the national income accounts identity in terms of saving and investment. Begin with the identity

$$Y = C + I + G + NX.$$

Subtract C and G from both sides to obtain

$$Y - C - G = I + NX.$$

Recall from Chapter 3 that $Y - C - G$ is national saving S, the sum of private saving, $Y - T - C$, and public saving, $T - G$. Therefore,

$$S = I + NX.$$

Subtracting I from both sides of the equation, we can write the national income accounts identity as

$$S - I = NX.$$

This form of the national income accounts identity shows that an economy's net exports must always equal the difference between its saving and its investment.

Let's look more closely at each part of this identity. The easy part is the right-hand side, NX, which is our net export of goods and services. Another name for net exports is the **trade balance**, because it tells us how our trade in goods and services departs from the benchmark of equal imports and exports.

The left-hand side of the identity is the difference between domestic saving and domestic investment, $S - I$, which we call **net capital outflow**. (It's sometimes called **net foreign investment**.) If net capital outflow is positive, our saving exceeds our investment, and we are lending the excess to foreigners. If the net capital outflow is negative, our investment exceeds our saving, and we are financing this extra investment by borrowing from abroad. Thus, net capital outflow equals the amount that domestic residents are lending abroad minus the amount that foreigners are lending to us. It reflects the international flow of funds to finance capital accumulation.

The national income accounts identity shows that net capital outflow always equals the trade balance. That is,

$$\begin{aligned} \text{Net Capital Outflow} &= \text{Trade Balance} \\ S - I &= NX. \end{aligned}$$

If $S - I$ and NX are positive, we have a **trade surplus**. In this case, we are net lenders in world financial markets, and we are exporting more goods than we are importing. If $S - I$ and NX are negative, we have a **trade deficit**. In this case, we are net borrowers in world financial markets, and we are importing more goods than we are exporting. If $S - I$ and NX are exactly zero, we are said to have **balanced trade** because the value of imports equals the value of exports.

The national income accounts identity shows that the international flow of funds to finance capital accumulation and the international flow of goods and services are two sides of the same coin. On the one hand, if our saving exceeds our investment, the saving that is not invested domestically is used to make loans to foreigners. Foreigners require these loans because we are providing them with more goods and services than they are providing us. That is, we are running a trade surplus. On the other hand, if our investment exceeds our saving, the extra investment must be financed by borrowing from abroad. These foreign loans enable us to import more

table 5-1

International Flows of Goods and Capital: Summary

This table shows the three outcomes that an open economy can experience.

Trade Surplus	Balanced Trade	Trade Deficit
Exports > Imports	Exports = Imports	Exports < Imports
Net Exports > 0	Net Exports = 0	Net Exports < 0
$Y > C + I + G$	$Y = C + I + G$	$Y < C + I + G$
Savings > Investment	Saving = Investment	Saving < Investment
Net Capital Outflow > 0	Net Capital Outflow = 0	Net Capital Outflow < 0

FYI

International Flows of Goods and Capital: An Example

The equality of net exports and net capital outflow is an identity: it must hold by the way the numbers are added up. But it is easy to miss the intuition behind this important relationship. The best way to understand it is to consider an example.

Imagine that Bill Gates sells a copy of the Windows operating system to a Japanese consumer for 5,000 yen. Because Mr. Gates is a U.S. resident, the sale represents an export of the United States. Other things equal, U.S. net exports rise. What else happens to make the identity hold? It depends on what Mr. Gates does with the 5,000 yen.

Suppose Mr. Gates decides to stuff the 5,000 yen in his mattress. In this case, Mr. Gates has allocated some of his saving to an investment in the Japanese economy (in the form of the Japanese currency) rather than to an investment in the U.S. economy. Thus, U.S. saving exceeds U.S. investment. The rise in U.S. net exports is matched by a rise in the U.S. net capital outflow.

If Mr. Gates wants to invest in Japan, however, he is unlikely to make currency his asset of choice. He might use the 5,000 yen to buy some stock in, say, the Sony Corporation, or he might buy a bond issued by the Japanese government. In either case, some of U.S. saving flows abroad. Once again, the U.S. net capital outflow exactly balances U.S. net exports.

The opposite situation occurs in Japan. When the Japanese consumer buys a copy of the Windows operating system, Japan's purchases of goods and services ($C + I + G$) rise, but there is no change in what Japan has produced (Y). The transaction reduces Japan's saving ($S = Y - C - G$) for a given level of investment (I). While the U.S. experiences a net capital outflow, Japan experiences a net capital inflow.

Now let's change the example. Suppose that instead of investing his 5,000 yen in a Japanese asset, Mr. Gates uses it to buy something made in Japan, such as a Sony Walkman. In this case, imports into the United State rise. Together, the Windows export and the Walkman import represent balanced trade between Japan and the United States. Because exports and imports rise equally, net exports and net capital outflow are both unchanged.

A final possibility is that Mr. Gates exchanges his 5,000 yen for U.S. dollars at a local bank. But this doesn't change the situation: the bank now has to do something with the 5,000 yen. It can buy Japanese assets (a U.S. net capital outflow); it can buy a Japanese good (a U.S. import); or it can sell the yen to another American who wants to make such a transaction. If you follow the money, you can see that, in the end, U.S. net exports must equal U.S. net capital outflow.

goods and services than we export. That is, we are running a trade deficit. Table 5-1 summarizes these lessons.

Note that the international flow of capital can take many forms. It is easiest to assume, as we have done so far, that when we run a trade deficit, foreigners make loans to us. This happens, for example, when the Japanese buy the debt issued by U.S. corporations or by the U.S. government. But the flow of capital can also take the form of foreigners buying domestic assets, such as when a citizen of Germany buys stock from an American on the New York Stock Exchange. Whether foreigners are buying domestically issued debt or domestically owned assets, they are obtaining a claim to the future returns to domestic capital. In both cases, foreigners end up owning some of the domestic capital stock.

5-2 Saving and Investment in a Small Open Economy

So far in our discussion of the international flows of goods and capital, we have merely rearranged accounting identities. That is, we have defined some of the variables that measure transactions in an open economy, and we have shown the links among these variables that follow from their definitions. Our next step is to develop a model that explains the behavior of these variables. We can then use the model to answer questions such as how the trade balance responds to changes in policy.

Capital Mobility and the World Interest Rate

In a moment we present a model of the international flows of capital and goods. Because the trade balance equals the net capital outflow, which in turn equals saving minus investment, our model focuses on saving and investment. To develop this model, we use some elements that should be familiar from Chapter 3, but in contrast to the Chapter 3 model, we do not assume that the real interest rate equilibrates saving and investment. Instead, we allow the economy to run a trade deficit and borrow from other countries, or to run a trade surplus and lend to other countries.

If the real interest rate does not adjust to equilibrate saving and investment in this model, what does determine the real interest rate? We answer this question here by considering the simple case of a small open economy with perfect capital mobility. By "small" we mean that this economy is a small part of the world market and thus, by itself, can have only a negligible effect on the world interest rate. By "perfect capital mobility" we mean that residents of the country have full access to world financial markets. In particular, the government does not impede international borrowing or lending.

Because of this assumption of perfect capital mobility, the interest rate in our small open economy, r, must equal the world interest rate r^*, the real interest rate prevailing in world financial markets:

$$r = r^*.$$

Residents of the small open economy need never borrow at any interest rate above r^*, because they can always get a loan from abroad. Similarly, residents of this economy need never lend at any interest rate below r^*, because they can always earn r^* by lending abroad. Thus, the world interest rate determines the interest rate in our small open economy.

Let us discuss for a moment what determines the world real interest rate. In a closed economy, the equilibrium of domestic saving and domestic investment determines the interest rate. Barring interplanetary trade, the world economy is a closed economy. Therefore, the equilibrium of world saving and world investment determines the world interest rate. Our small open economy has a negligible effect on the world real interest rate because, being a small part of the world,

it has a negligible effect on world saving and world investment. Hence, our small open economy takes the world interest rate as exogenously given.

The Model

To build the model of the small open economy, we take three assumptions from Chapter 3:

- The economy's output Y is fixed by the factors of production and the production function. We write this as

$$Y = \bar{Y} = F(\bar{K}, \bar{L}).$$

- Consumption C is positively related to disposable income $Y - T$. We write the consumption function as

$$C = C(Y - T).$$

- Investment I is negatively related to the real interest rate r. We write the investment function as

$$I = I(r).$$

These are the three key parts of our model. If you do not understand these relationships, review Chapter 3 before continuing.

We can now return to the accounting identity and write it as

$$NX = (Y - C - G) - I$$
$$NX = S - I.$$

Substituting our three assumptions from Chapter 3 and the condition that the interest rate equals the world interest rate, we obtain

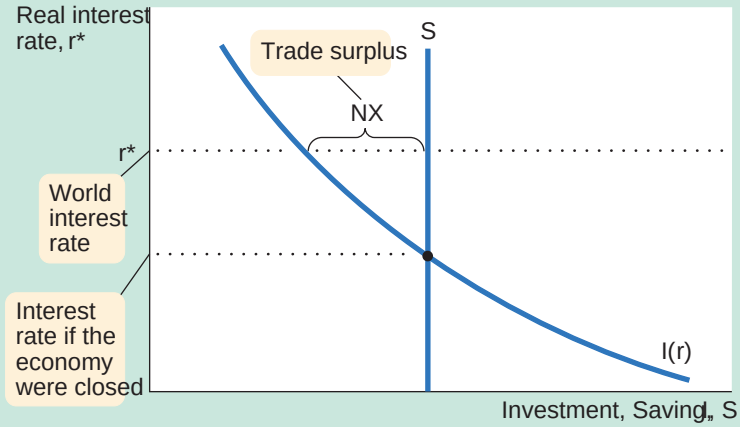
$$NX = [\bar{Y} - C(\bar{Y} - T) - G] - I(r^*)$$
$$= \bar{S} - I(r^*).$$

This equation shows what determines saving S and investment I and thus the trade balance NX. Remember that saving depends on fiscal policy: lower government purchases G or higher taxes T raise national saving. Investment depends on the world real interest rate r^*: high interest rates make some investment projects unprofitable. Therefore, the trade balance depends on these variables as well.

In Chapter 3 we graphed saving and investment as in Figure 5-2. In the closed economy studied in that chapter, the real interest rate adjusts to equilibrate saving and investment—that is, the real interest rate is found where the saving and investment curves cross. In the small open economy, however, the real interest rate equals the world real interest rate. The trade balance is determined by the difference between saving and investment at the world interest rate.

At this point, you might wonder about the mechanism that causes the trade balance to equal the net capital outflow. The determinants of the capital flows are

figure 5-2



Saving and Investment in a Small Open Economy In a closed economy, the real interest rate adjusts to equilibrate saving and investment. In a small open economy, the interest rate is determined in world financial markets. The difference between saving and investment determines the trade balance. Here there is a trade surplus, because at the world interest rate, saving exceeds investment.

easy to understand. When saving falls short of investment, investors borrow from abroad; when saving exceeds investment, the excess is lent to other countries. But what causes those who import and export to behave in a way that ensures that the international flow of goods exactly balances this international flow of capital? For now we leave this question unanswered, but we return to it in Section 5-3 when we discuss the determination of exchange rates.

How Policies Influence the Trade Balance

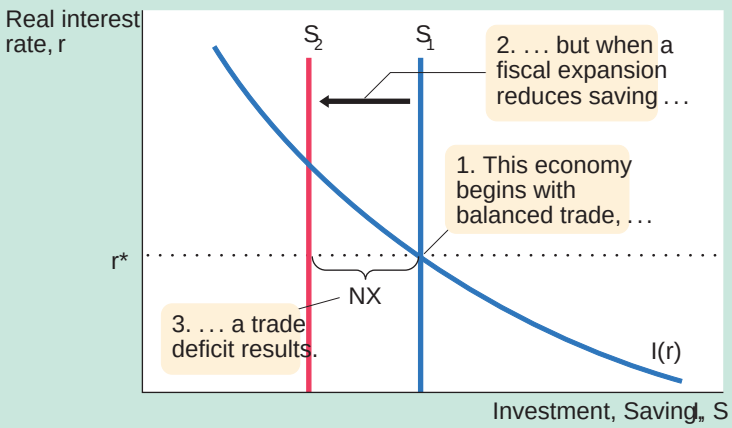
Suppose that the economy begins in a position of balanced trade. That is, at the world interest rate, investment equals saving and net exports equal zero. Let us use our model to predict the effects of government policies at home and abroad.

Fiscal Policy at Home Consider first what happens to the small open economy if the government expands domestic spending by increasing government purchases. The increase in reduces national saving, because With an unchanged world real interest rate, investment remains the same. Therefore, saving falls below investment, and some investment must now be financed by borrowing from abroad. Because implies a fall in The economy now runs a trade deficit.

The same logic applies to a decrease in taxes. A tax cut raises disposable income stimulates consumption, and reduces national saving. (Even though some of the tax cut finds its way into private saving, public saving falls by the full amount of the tax cut; in total, saving falls.) Because the reduction in national saving in turn lowers

Figure 5-3 illustrates these effects. Any fiscal-policy change that increases private consumption or public consumption reduces national saving and, therefore, shifts the vertical line that represents saving from to Because is the distance between the saving schedule and the investment schedule at the world interest rate, this shift reduces Hence, starting from balanced trade, a change in fiscal policy that reduces national saving leads to a trade de

figure 5-3



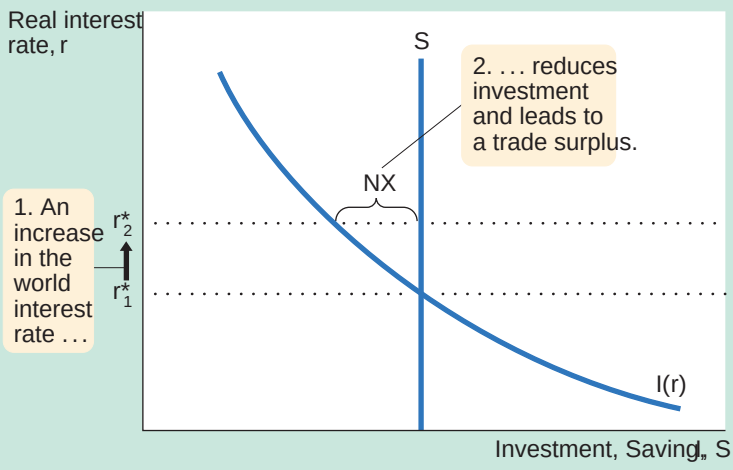
A Fiscal Expansion at Home in a Small Open Economy An increase in government purchases or a reduction in taxes reduces national saving and thus shifts the saving schedule to the left, from S_1 to S_2. The result is a trade deficit.

Fiscal Policy Abroad Consider now what happens to a small open economy when foreign governments increase their government purchases. If these foreign countries are a small part of the world economy, then this fiscal change has a negligible impact on other countries. But if these foreign countries are a large part of the world economy, their increase in government purchases reduces world saving and causes the world interest rate to rise.

The increase in the world interest rate raises the cost of borrowing and, thus, reduces investment in our small open economy. Because there has been no change in domestic saving, saving now exceeds investment, and some of our saving begins to flow abroad. Since $NX = S - I$, the reduction in I must also increase NX. Hence, reduced saving abroad leads to a trade surplus at home.

Figure 5-4 illustrates how a small open economy starting from balanced trade responds to a foreign fiscal expansion. Because the policy change is occurring

figure 5-4

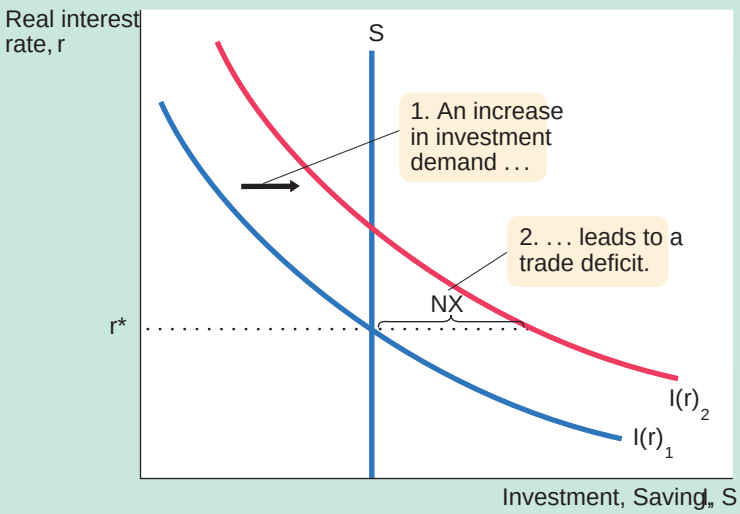


A Fiscal Expansion Abroad in a Small Open Economy A fiscal expansion in a foreign economy large enough to influence world saving and investment raises the world interest rate from r^*_1 to r^*_2. The higher world interest rate reduces investment in this small open economy, causing a trade surplus.

abroad, the domestic saving and investment schedules remain the same. The only change is an increase in the world interest rate from r_1^* to r_2^*. The trade balance is the difference between the saving and investment schedules; because saving exceeds investment at r_2^*, there is a trade surplus. Hence, an increase in the world interest rate due to fiscal expansion abroad leads to a trade surplus.

Shifts in Investment Demand Consider what happens to our small open economy if its investment schedule shifts outward—that is, if the demand for investment goods at every interest rate increases. This shift would occur if, for example, the government changed the tax laws to encourage investment by providing an investment tax credit. Figure 5-5 illustrates the impact of a shift in the investment schedule. At a given world interest rate, investment is now higher. Because saving is unchanged, some investment must now be financed by borrowing from abroad, which means the net capital outflow is negative. Put differently, because $NX = S - I$, the increase in I implies a decrease in NX. Hence, an outward shift in the investment schedule causes a trade deficit.

figure 5-5



A Shift in the Investment Schedule in a Small Open Economy An outward shift in the investment schedule from $I(r)_1$ to $I(r)_2$ increases the amount of investment at the world interest rate r^*. As a result, investment now exceeds saving, which means the economy is borrowing from abroad and running a trade deficit.

Evaluating Economic Policy

Our model of the open economy shows that the flow of goods and services measured by the trade balance is inextricably connected to the international flow of funds for capital accumulation. The net capital outflow is the difference between domestic saving and domestic investment. Thus, the impact of economic policies on the trade balance can always be found by examining their impact on domestic saving and domestic investment. Policies that increase investment or decrease saving tend to cause a trade deficit, and policies that decrease investment or increase saving tend to cause a trade surplus.

Our analysis of the open economy has been positive, not normative. That is, our analysis of how economic policies influence the international flows of capital and goods has not told us whether these policies are desirable. Evaluating economic policies and their impact on the open economy is a frequent topic of debate among economists and policymakers.

When a country runs a trade deficit, policymakers must confront the question of whether it represents a national problem. Most economists view a trade deficit not as a problem in itself, but perhaps as a symptom of a problem. A trade deficit could be a reflection of low saving. In a closed economy, low saving leads to low investment and a smaller future capital stock. In an open economy, low saving leads to a trade deficit and a growing foreign debt, which eventually must be repaid. In both cases, high current consumption leads to lower future consumption, implying that future generations bear the burden of low national saving.

Yet trade deficits are not always a reflection of economic malady. When poor rural economies develop into modern industrial economies, they sometimes finance their high levels of investment with foreign borrowing. In these cases, trade deficits are a sign of economic development. For example, South Korea ran large trade deficits throughout the 1970s, and it became one of the success stories of economic growth. The lesson is that one cannot judge economic performance from the trade balance alone. Instead, one must look at the underlying causes of the international flows.

CASE STUDY

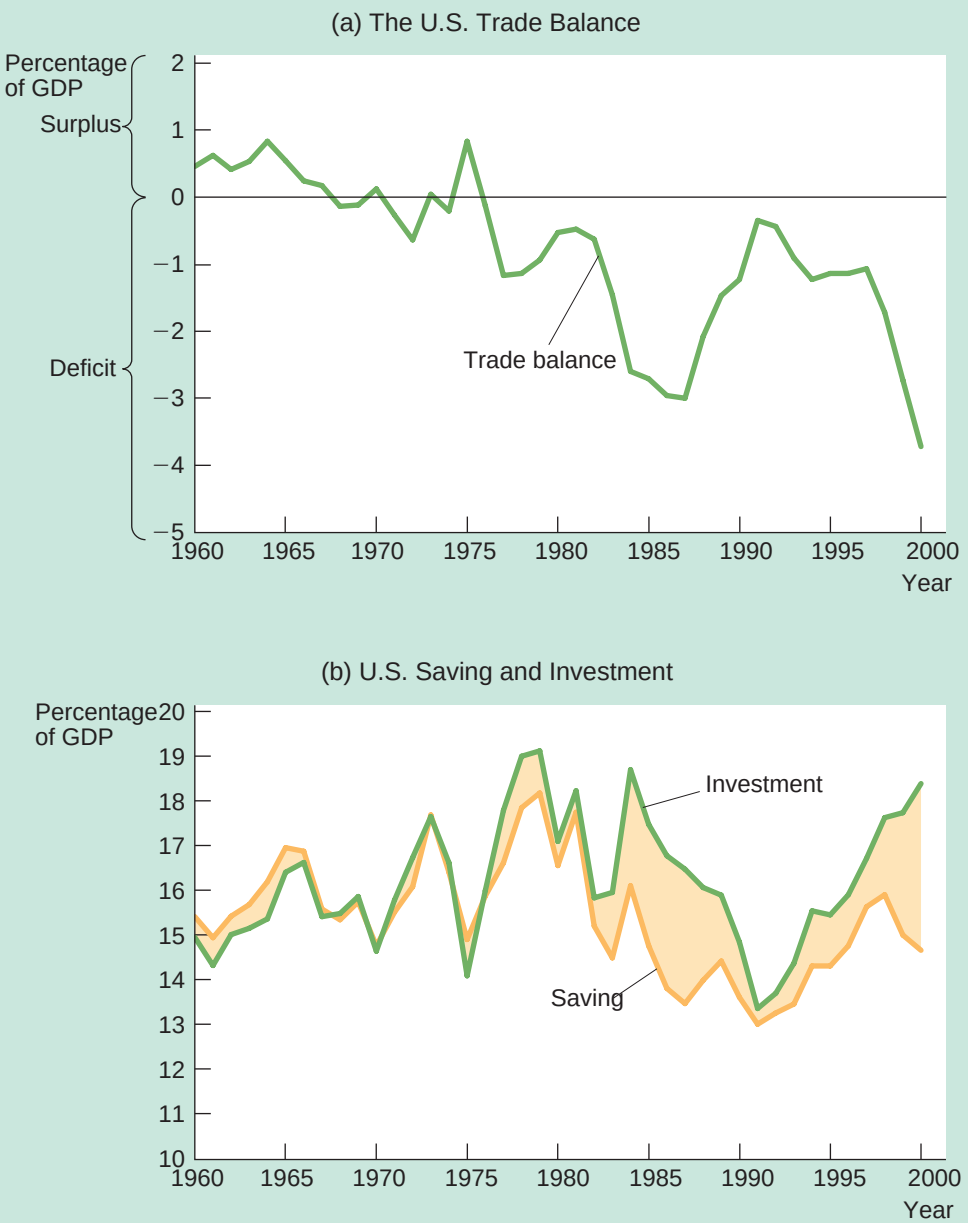
The U.S. Trade Deficit

During the 1980s and 1990s, the United States ran large trade deficits. Panel (a) of Figure 5-6 documents this experience by showing net exports as a percentage of GDP. The exact size of the trade deficit fluctuated over time, but it was large throughout these two decades. In 2000, the trade deficit was $371 billion, or 3.7 percent of GDP. As accounting identities require, this trade deficit had to be financed by borrowing from abroad (or, equivalently, by selling U.S. assets abroad). During this period, the United States went from being the world's largest creditor to the world's largest debtor.

What caused the U.S. trade deficit? There is no single explanation. But to understand some of the forces at work, it helps to look at national saving and domestic investment, as shown in panel (b) of the figure. Keep in mind that the trade deficit is the difference between saving and investment.

The start of the trade deficit coincided with a fall in national saving. This development can be explained by the expansionary fiscal policy in the 1980s. With the support of President Reagan, the U.S. Congress passed legislation in 1981 that substantially cut personal income taxes over the next three years. Because these tax cuts were not met with equal cuts in government spending, the federal budget went into deficit. These budget deficits were among the largest ever experienced in a period of peace and prosperity, and they continued long after Reagan left office. According to our model, such a policy should reduce national

figure 5-6



The Trade Balance, Saving, and Investment: The U.S. Experience Panel (a) shows the trade balance as a percentage of GDP. Positive numbers represent a surplus, and negative numbers represent a deficit. Panel (b) shows national saving and investment as a percentage of GDP since 1960. The trade balance equals saving minus investment.

Source: U.S. Department of Commerce.

saving, thereby causing a trade deficit. And, in fact, that is exactly what happened. Because the government budget and trade balance went into deficit at roughly the same time, these shortfalls were called the twin deficits.

Things started to change in the 1990s, when the U.S. federal government got its fiscal house in order. The first President Bush and President Clinton both signed tax increases, while Congress kept a lid on spending. In addition to these policy changes, rapid productivity growth in the late 1990s raised incomes and, thus, further increased tax revenue. These developments moved the U.S. federal budget from deficit to surplus, which in turn caused national saving to rise.

In contrast to what our model predicts, the increase in national saving did not coincide with a shrinking trade deficit, because domestic investment rose at the same time. The likely explanation is that the boom in information technology caused an expansionary shift in the U.S. investment function. Even though policy was pushing the trade deficit toward surplus, the investment boom was an even stronger force pushing the trade balance toward deficit.

The history of the U.S. trade deficit shows that this statistic, by itself, does not tell us much about what is happening in the economy. We have to look deeper at saving, investment, and the policies and events that cause them to change over time.

5-3 Exchange Rates

Having examined the international flows of capital and of goods and services, we now extend the analysis by considering the prices that apply to these transactions. The **exchange rate** between two countries is the price at which residents of those countries trade with each other. In this section we examine precisely what the exchange rate measures, and we then discuss how exchange rates are determined.

Nominal and Real Exchange Rates

Economists distinguish between two exchange rates: the nominal exchange rate and the real exchange rate. Let's discuss each in turn and see how they are related.

The Nominal Exchange Rate The **nominal exchange rate** is the relative price of the currency of two countries. For example, if the exchange rate between the U.S. dollar and the Japanese yen is 120 yen per dollar, then you can exchange one dollar for 120 yen in world markets for foreign currency. A Japanese who wants to obtain dollars would pay 120 yen for each dollar he bought. An American who wants to obtain yen would get 120 yen for each dollar he paid. When people refer to "the exchange rate" between two countries, they usually mean the nominal exchange rate.

FYI How Newspapers Report the Exchange Rate

You can find nominal exchange rates reported daily in many newspapers. Here's how they are reported in the *Wall Street Journal*.

Notice that each exchange rate is reported in two ways. On this Thursday, 1 dollar bought 116.29 yen, and 1 yen bought 0.008599 dollars. We can say the exchange rate is 116.29 yen per dollar, or we can say the exchange rate is 0.008599 dollars per yen. Because 0.008599 equals 1/116.29, these two ways of expressing the exchange rate are equivalent. This book always expresses the exchange rate in units of foreign currency per dollar.

The exchange rate on this Thursday of 116.29 yen per dollar was down from 117.67 yen per dollar on Wednesday. Such a fall in the exchange rate is called a *depreciation* of the dollar; a rise in the exchange rate is called an *appreciation*. When the domestic currency depreciates, it buys less of the foreign currency; when it appreciates, it buys more.

CURRENCY TRADING

Thursday, September 20, 2001

EXCHANGE RATES

The foreign exchange mid-range rates below apply to trading among banks in amounts of $1 million and more, as quoted at 4 p.m. Eastern time by Rueters and other sources. Retail transactions provide fewer units of foreign currency per dollar. Rates for the 12 Euro currency countries are derived from the latest dollar-euro rate using the exchange ratios set 1/1/99.

	U.S. $ EQUIV.		CURRENCY PER U.S. $	
Country	Thu	Wed	Thu	Wed
Argentina (Peso)	1.0002	1.0003	.9998	.9997
Australia (Dollar)	.4932	.4931	2.0274	2.0278
Austria (Schilling)	.06740	.06735	14.837	14.848
Bahrain (Dinar)	2.6525	2.6518	.3770	.3771
Belgium (Franc)	.0230	.0230	43.4955	43.5284
Brazil (Real)	.3612	.3699	2.7685	2.7035
Britain (Pound)	1.4671	1.4684	.6816	.6810
1-month forward	1.4648	1.4658	.6827	.6822
3-months forward	1.4603	1.4615	.6848	.6842
6-months forward	1.4540	1.4549	.6878	.6873
Canada (Dollar)	.6364	.6376	1.5713	1.5683
1-month forward	.6359	.6371	1.5725	1.5696
3-months forward	.6354	.6365	1.5739	1.5712
6-months forward	.6348	.6359	1.5752	1.5725
Chile (Peso)	.001440	.001449	694.55	690.15
China (Renminbi)	.1208	.1208	8.2766	8.2766
Colombia (Peso)	.0004269	.0004264	2342.50	2345.00
Czech. Rep. (Koruna)				
Commercial rate	.02704	.02716	36.985	36.822
Denmark (Krone)	.1246	.1245	8.0260	8.0315
Ecuador (US Dollar)-e	1.0000	1.0000	1.0000	1.0000
Finland (Markka)	.1560	.1559	6.4108	6.4157
France (Franc)	.1414	.1413	7.0727	7.0780
1-month forward	.1413	.1412	7.0794	7.0846
3-months forward	.1410	.1409	7.0921	7.0970
6-months forward	.1407	.1406	7.1088	7.1146
Germany (Mark)	.4742	.4738	2.1088	2.1104
1-month forward	.4738	.4734	2.1108	2.1124
3-months forward	.4729	.4726	2.1146	2.1161
6-months forward	.4718	.4714	2.1196	2.1214
Greece (Drachma)	.002722	.002720	367.41	367.70
Hong Kong (Dollar)	.1282	.1282	7.7995	7.7992
Hungary (Forint)	.003589	.003592	278.67	278.40
India (Rupee)	.02084	.02083	47.980	48.010
Indonesia (Rupiah)	.0001059	.0001041	9445	9605
Ireland (Punt)	1.1776	1.1767	.8492	.8498
Israel (Shekel)	.2297	.2304	4.3541	4.3400
Italy (Lira)	.0004790	.0004786	2087.74	2089.31
Japan (Yen)	.008599	.008498	116.29	117.67
1-month forward	.008618	.008517	116.04	117.41
3-months forward	.008655	.008553	115.53	116.92
6-months forward	.008707	.008603	114.85	116.23
Jordan (Dinar)	1.4069	1.4069	.7108	.7108
Kuwait (Dinar)	3.2830	3.2841	.3046	.3045
Lebanon (Pound)	.0006604	.0006604	1514.25	1514.25
Malaysia (Ringgit)-b	.2632	.2632	3.8000	3.8000
Malta (Lira)	2.779	2.2774	.4390	.4391
Mexico (Peso)				
Floating rate	.1055	.1060	9.4825	9.4325
Netherlands (Guilder)	.4209	.4205	2.3761	2.3779
New Zealand (Dollar)	.4128	.4122	2.4225	2.4260
Norway (Krone)	.1169	.1169	8.5530	8.5551
Pakistan (Rupee)	.01559	.01559	64.150	64.150
Peru (new Sol)	.2857	.2857	3.5005	3.5003
Philippines (Peso)	.01949	.01946	51.300	51.375
Poland (Zloty)-d	.2404	.2382	4.1600	4.1975
Portugal (Escudo)	.004626	.004623	216.16	216.33
Russia (Ruble)-a	.03399	.03397	29.422	29.442
Saudi Arabia (Riyal)	.2666	.2666	3.7508	3.7505
Singapore (Dollar)	.5758	.5749	1.7368	1.7395
Slovak Rep. (Koruna)	.02116	.02118	47.253	47.219
South Africa (Rand)	.1149	.1156	8.6998	8.6478
South Korea (Won)	.0007734	.0007740	1293.00	1292.00
Spain (Peseta)	.005574	.005570	179.40	179.54
Sweden (Krona)	.0942	.0952	10.6178	10.5030
Switzerland (Franc)	.6303	.6260	1.5865	1.5974
1-month forward	.6302	.6259	1.5867	1.5977
3-months forward	.6301	.6258	1.5870	1.5980
6-months forward	.6303	.6259	1.5866	1.5977
Taiwan (Dollar)	.02894	.02895	34.560	34.540
Thailand (Baht)	.02266	.02265	44.140	44.150
Turkey (Lira)-f	.00000066	.00000066	1520000	1504000
United Arab (Dirham)	.2723	.2723	3.6730	3.6728
Uruguay (New Peso)				
Financial	.07278	.07278	13.740	13.740
Venezuela (Bolivar)	.001345	.001345	743.25	743.25
SDR	1.2945	1.2950	.7725	.7722
Euro	.9275	.9268	1.0782	1.0790

Special Drawing Rights (SDR) are based on exchange rates for the U.S., German, British, French, and Japanese currencies. Source: International Monetary Fund.
a-Russian Central Bank rate. b-Government rate. d-Floating rate; trading band suspended on 4/11/00. e-Adopted U.S. dollar as of 9/11/00. f-Floating rate, eff. Feb. 22.

Source: *Wall Street Journal*, Friday, September 21, 2001. Reprinted by permission of the *Wall Street Journal*.

© 2001 Dow Jones & Company, Inc. All Rights Reserved Worldwide.

The Real Exchange Rate The real exchange rate is the relative price of the goods of two countries. That is, the real exchange rate tells us the rate at which we can trade the goods of one country for the goods of another. The real exchange rate is sometimes called the *terms of trade*.

To see the relation between the real and nominal exchange rates, consider a single good produced in many countries: cars. Suppose an American car costs $10,000 and a similar Japanese car costs 2,400,000 yen. To compare the prices of the two cars, we must convert them into a common currency. If a dollar is worth 120 yen, then the American car costs 1,200,000 yen. Comparing the price of the American car (1,200,000 yen) and the price of the Japanese car (2,400,000 yen), we conclude that the American car costs one-half of what the Japanese car costs. In other words, at current prices, we can exchange two American cars for one Japanese car.

We can summarize our calculation as follows:

$$\text{Real Exchange Rate} = \frac{(120 \text{ yen/dollar}) \times (10{,}000 \text{ dollars/American Car})}{(2{,}400{,}000 \text{ yen/Japanese Car})}$$

$$= 0.5 \frac{\text{Japanese Car}}{\text{American Car}}.$$

At these prices and this exchange rate, we obtain one-half of a Japanese car per American car. More generally, we can write this calculation as

$$\text{Real Exchange Rate} = \frac{\text{Nominal Exchange Rate} \times \text{Price of Domestic Good}}{\text{Price of Foreign Good}}.$$

The rate at which we exchange foreign and domestic goods depends on the prices of the goods in the local currencies and on the rate at which the currencies are exchanged.

This calculation of the real exchange rate for a single good suggests how we should define the real exchange rate for a broader basket of goods. Let e be the nominal exchange rate (the number of yen per dollar), P the price level in the United States (measured in dollars), and P^* the price level in Japan (measured in yen). Then the real exchange rate is

$$\begin{array}{ccccc} \text{Real} & & \text{Nominal} & & \text{Ratio of} \\ \text{Exchange} & = & \text{Exchange} & \times & \text{Price} \\ \text{Rate} & & \text{Rate} & & \text{Levels} \end{array}$$

$$\epsilon = e \times (P/P^*).$$

The real exchange rate between two countries is computed from the nominal exchange rate and the price levels in the two countries. If the real exchange rate is high, foreign goods are relatively cheap, and domestic goods are relatively expensive. If the real exchange rate is low, foreign goods are relatively expensive, and domestic goods are relatively cheap

The Real Exchange Rate and the Trade Balance

What macroeconomic influence does the real exchange rate exert? To answer this question, remember that the real exchange rate is nothing more than a relative price. Just as the relative price of hamburgers and pizza determines which you choose for lunch, the relative price of domestic and foreign goods affects the demand for these goods.

Drawing by Matteson; © 1988 The New Yorker Magazine, Inc.

"How about Nebraska? The dollar is still strong in Nebraska."

Suppose first that the real exchange rate is low. In this case, because domestic goods are relatively cheap, domestic residents will want to purchase few imported goods: they will buy Fords rather than Toyotas, drink Coors rather than Heineken, and vacation in Florida rather than Europe. For the same reason, foreigners will want to buy many of our goods. As a result of both of these actions, the quantity of our net exports demanded will be high.

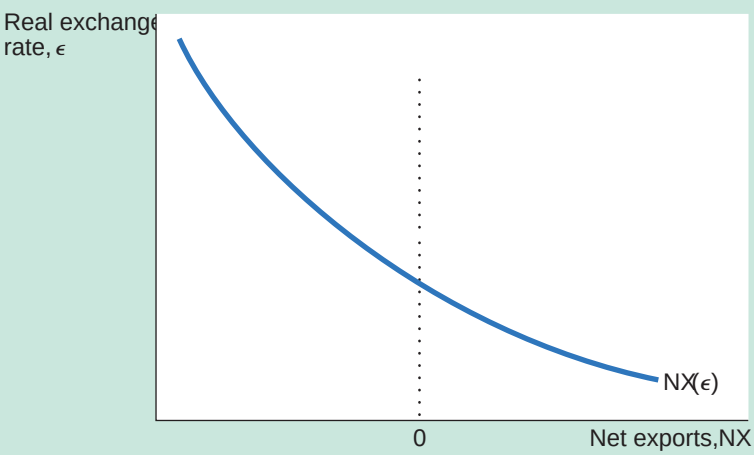
The opposite occurs if the real exchange rate is high. Because domestic goods are expensive relative to foreign goods, domestic residents will want to buy many imported goods, and foreigners will want to buy few of our goods. Therefore, the quantity of our net exports demanded will be low.

We write this relationship between the real exchange rate and net exports as

$$NX = NX(\epsilon).$$

This equation states that net exports are a function of the real exchange rate. Figure 5-7 illustrates this negative relationship between the trade balance and the real exchange rate.

figure 5-7



Net Exports and the Real Exchange Rate The figure shows the relationship between the real exchange rate and net exports: the lower the real exchange rate, the less expensive are domestic goods relative to foreign goods, and thus the greater are our net exports. Note that a portion of the horizontal axis measures negative values of NX: because imports can exceed exports, net exports can be less than zero.

The Determinants of the Real Exchange Rate

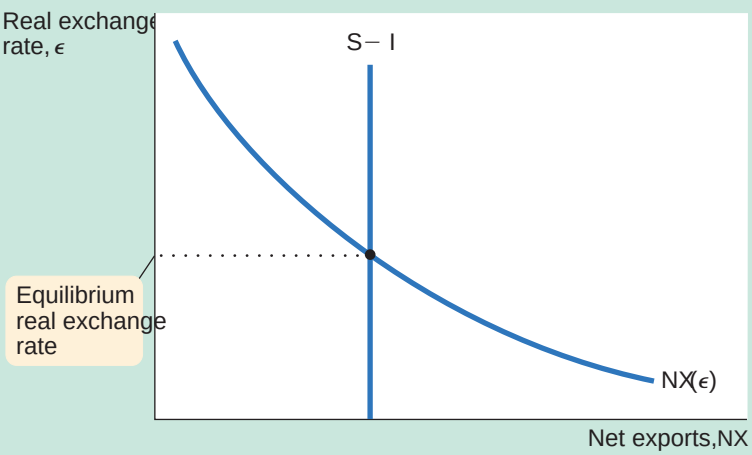
We now have all the pieces needed to construct a model that explains what factors determine the real exchange rate. In particular, we combine the relationship between net exports and the real exchange rate we just discussed with the model of the trade balance we developed earlier in the chapter. We can summarize the analysis as follows:

- The real exchange rate is related to net exports. When the real exchange rate is lower, domestic goods are less expensive relative to foreign goods, and net exports are greater.
- The trade balance (net exports) must equal the net capital outflow, which in turn equals saving minus investment. Saving is fixed by the consumption function and fiscal policy; investment is fixed by the investment function and the world interest rate.

Figure 5-8 illustrates these two conditions. The line showing the relationship between net exports and the real exchange rate slopes downward because a low real exchange rate makes domestic goods relatively inexpensive. The line representing the excess of saving over investment, $S - I$, is vertical because neither saving nor investment depends on the real exchange rate. The crossing of these two lines determines the equilibrium exchange rate.

Figure 5-8 looks like an ordinary supply-and-demand diagram. In fact, you can think of this diagram as representing the supply and demand for foreign-currency exchange. The vertical line, $S - I$, represents the net capital outflow and thus the supply of dollars to be exchanged into foreign currency and invested abroad. The downward-sloping line, NX, represents the net demand for dollars coming from foreigners who want dollars to buy our goods. At the equilibrium real exchange rate, the supply of dollars available from the net capital outflow balances the demand for dollars by foreigners buying our net exports.

figure 5-8



How the Real Exchange Rate Is Determined The real exchange rate is determined by the intersection of the vertical line representing saving minus investment and the downward-sloping net-exports schedule. At this intersection, the quantity of dollars supplied for the flow of capital abroad equals the quantity of dollars demanded for the net export of goods and services.

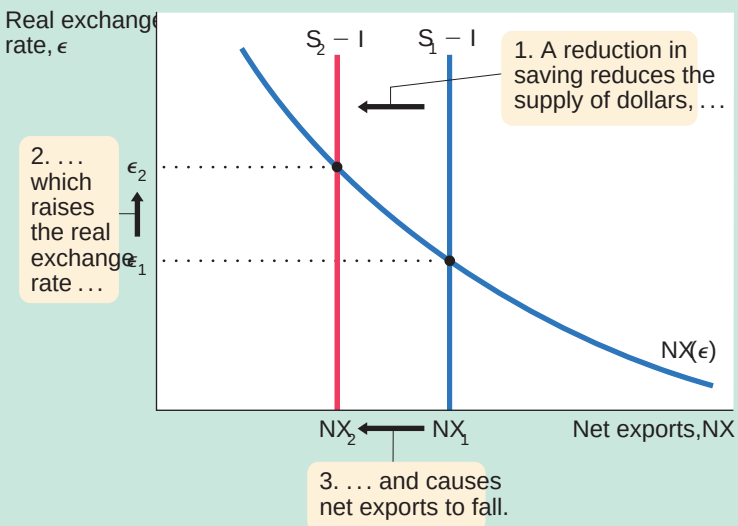
How Policies Influence the Real Exchange Rate

We can use this model to show how the changes in economic policy we discussed earlier affect the real exchange rate.

Fiscal Policy at Home What happens to the real exchange rate if the government reduces national saving by increasing government purchases or cutting taxes? As we discussed earlier, this reduction in saving lowers S and thus NX. That is, the reduction in saving causes a trade deficit.

Figure 5-9 shows how the equilibrium real exchange rate adjusts to ensure that NX falls. The change in policy shifts the vertical line to the left, lowering the supply of dollars to be invested abroad. The lower supply causes the equilibrium real exchange rate to rise from ϵ_1 to ϵ_2—that is, the dollar becomes more valuable. Because of the rise in the value of the dollar, domestic goods become more expensive relative to foreign goods, which causes exports to fall and imports to rise. The change in exports and the change in imports both act to reduce net exports.

figure 5-9

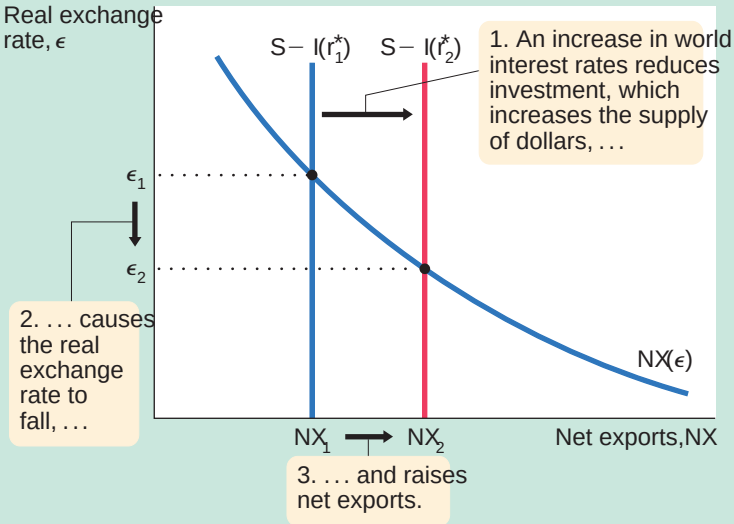


The Impact of Expansionary Fiscal Policy at Home on the Real Exchange Rate Expansionary fiscal policy at home, such as an increase in government purchases or a cut in taxes, reduces national saving. The fall in saving reduces the supply of dollars to be exchanged into foreign currency, from $S_1 - I$ to $S_2 - I$. This shift raises the equilibrium real exchange rate from ϵ_1 to ϵ_2.

Fiscal Policy Abroad What happens to the real exchange rate if foreign governments increase government purchases or cut taxes? This change in fiscal policy reduces world saving and raises the world interest rate. The increase in the world interest rate reduces domestic investment, which raises $S - I$ and thus NX. That is, the increase in the world interest rate causes a trade surplus.

Figure 5-10 shows that this change in policy shifts the vertical $S - I$ line to the right, raising the supply of dollars to be invested abroad. The equilibrium real exchange rate falls. That is, the dollar becomes less valuable, and domestic goods become less expensive relative to foreign goods.

figure 5-10

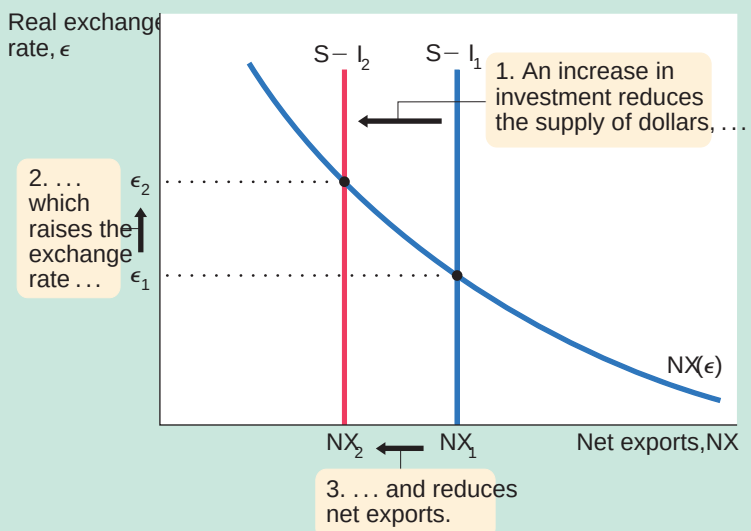


The Impact of Expansionary Fiscal Policy Abroad on the Real Exchange Rate Expansionary fiscal policy abroad reduces world saving and raises the world interest rate from r_1^* to r_2^*. The increase in the world interest rate reduces investment at home, which in turn raises the supply of dollars to be exchanged into foreign currencies. As a result, the equilibrium real exchange rate falls from ϵ_1 to ϵ_2.

Shifts in Investment Demand What happens to the real exchange rate if investment demand at home increases, perhaps because Congress passes an investment tax credit? At the given world interest rate, the increase in investment demand leads to higher investment. A higher value of I means lower values of $S - I$ and NX. That is, the increase in investment demand causes a trade deficit.

Figure 5-11 shows that the increase in investment demand shifts the vertical $S - I$ line to the left, reducing the supply of dollars to be invested abroad. The

figure 5-11



The Impact of an Increase in Investment Demand on the Real Exchange Rate An increase in investment demand raises the quantity of domestic investment from I_1 to I_2. As a result, the supply of dollars to be exchanged into foreign currencies falls from $S - I_1$ to $S - I_2$. This fall in supply raises the equilibrium real exchange rate from ϵ_1 to ϵ_2.

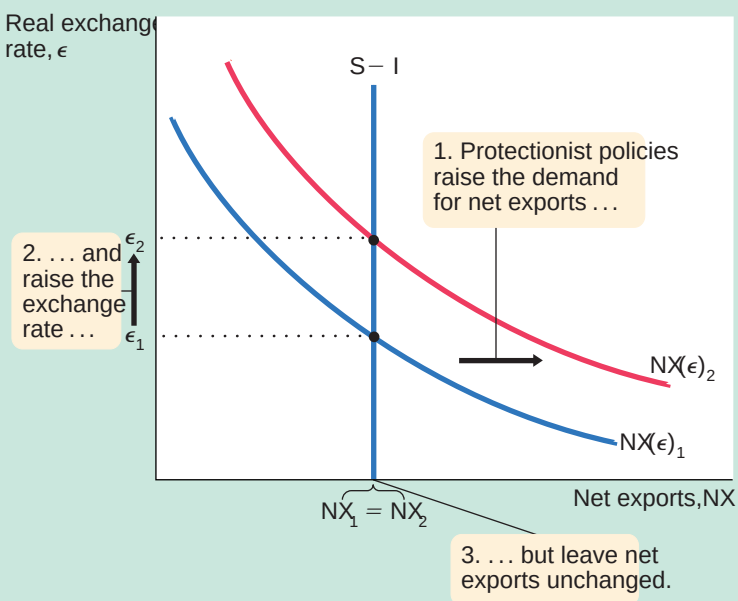
equilibrium real exchange rate rises. Hence, when the investment tax credit makes investing in the United States more attractive, it also increases the value of the U.S. dollars necessary to make these investments. When the dollar appreciates, domestic goods become more expensive relative to foreign goods, and net exports fall.

The Effects of Trade Policies

Now that we have a model that explains the trade balance and the real exchange rate, we have the tools to examine the macroeconomic effects of trade policies. Trade policies, broadly defined, are policies designed to influence directly the amount of goods and services exported or imported. Most often, trade policies take the form of protecting domestic industries from foreign competition either by placing a tax on foreign imports (a tariff) or restricting the amount of goods and services that can be imported (a quota).

As an example of a protectionist trade policy, consider what would happen if the government prohibited the import of foreign cars. For any given real exchange rate, imports would now be lower, implying that net exports (exports minus imports) would be higher. Thus, the net-exports schedule shifts outward, as in Figure 5-12. To see the effects of the policy, we compare the old equilibrium and the new equilibrium. In the new equilibrium, the real exchange rate is higher, and net exports are unchanged. Despite the shift in the net-exports schedule, the equilibrium level of net exports remains the same, because the protectionist policy does not alter either saving or investment.

figure 5-12



The Impact of Protectionist Trade Policies on the Real Exchange Rate A protectionist trade policy, such as a ban on imported cars, shifts the net-exports schedule from $NX(\epsilon)_1$ to $NX(\epsilon)_2$, which raises the real exchange rate from ϵ_1 to ϵ_2. Notice that, despite the shift in the net-exports schedule, the equilibrium level of net exports is unchanged.

This analysis shows that protectionist trade policies do not affect the trade balance. This surprising conclusion is often overlooked in the popular debate over trade policies. Because a trade deficit reflects an excess of imports over exports, one might guess that reducing imports—such as by prohibiting the import of foreign cars—would reduce a trade deficit. Yet our model shows that protectionist policies lead only to an appreciation of the real exchange rate. The increase in the price of domestic goods relative to foreign goods tends to lower net exports by stimulating imports and depressing exports. Thus, the appreciation offsets the increase in net exports that is directly attributable to the trade restriction.

Although protectionist trade policies do not alter the trade balance, they do affect the amount of trade. As we have seen, because the real exchange rate appreciates, the goods and services we produce become more expensive relative to foreign goods and services. We therefore export less in the new equilibrium. Because net exports are unchanged, we must import less as well. (The appreciation of the exchange rate does stimulate imports to some extent, but this only partly offsets the decrease in imports caused by the trade restriction.) Thus, protectionist policies reduce both the quantity of imports and the quantity of exports.

This fall in the total amount of trade is the reason economists almost always oppose protectionist policies. International trade benefits all countries by allowing each country to specialize in what it produces best and by providing each country with a greater variety of goods and services. Protectionist policies diminish these gains from trade. Although these policies benefit certain groups within society—for example, a ban on imported cars helps domestic car producers—society on average is worse off when policies reduce the amount of international trade.

The Determinants of the Nominal Exchange Rate

Having seen what determines the real exchange rate, we now turn our attention to the nominal exchange rate—the rate at which the currencies of two countries trade. Recall the relationship between the real and the nominal exchange rate:

$$\begin{array}{ccccc} \text{Real} & & \text{Nominal} & & \text{Ratio of} \\ \text{Exchange} & = & \text{Exchange} & \times & \text{Price} \\ \text{Rate} & & \text{Rate} & & \text{Levels} \end{array}$$

$$\epsilon = e \times (P/P^*).$$

We can write the nominal exchange rate as

$$e = \epsilon \times (P^*/P).$$

This equation shows that the nominal exchange rate depends on the real exchange rate and the price levels in the two countries. Given the value of the real

exchange rate, if the domestic price level rises, then the nominal exchange rate e will fall: because a dollar is worth less, a dollar will buy fewer yen. However, if the Japanese price level P^* rises, then the nominal exchange rate will increase: because the yen is worth less, a dollar will buy more yen.

It is instructive to consider changes in exchange rates over time. The exchange rate equation can be written

$$\% \text{ Change in } e = \% \text{ Change in } \varepsilon + \% \text{ Change in } P^* - \% \text{ Change in } P.$$

The percentage change in ε is the change in the real exchange rate. The percentage change in P is the domestic inflation rate, π, and the percentage change in P^* is the foreign country's inflation rate π^*. Thus, the percentage change in the nominal exchange rate is

$$\% \text{ Change in } e = \% \text{ Change in } \varepsilon + (\pi^* - \pi)$$

$$\begin{array}{c} \text{Percentage Change in} \\ \text{Nominal Exchange Rate} \end{array} = \begin{array}{c} \text{Percentage Change in} \\ \text{Real Exchange Rate} \end{array} + \begin{array}{c} \text{Difference in} \\ \text{Inflation Rates.} \end{array}$$

This equation states that the percentage change in the nominal exchange rate between the currencies of two countries equals the percentage change in the real exchange rate plus the difference in the inflation rates. If a country has a high rate of inflation relative to the United States, a dollar will buy an increasing amount of the foreign currency over time. If a country has a low rate of inflation relative to the United States, a dollar will buy a decreasing amount of the foreign currency over time.

This analysis shows how monetary policy affects the nominal exchange rate. We know from Chapter 4 that high growth in the money supply leads to high inflation. Here, we have just seen that one consequence of high inflation is a depreciating currency: high π implies falling e. In other words, just as growth in the amount of money raises the price of goods measured in terms of money, it also tends to raise the price of foreign currencies measured in terms of the domestic currency.

CASE STUDY

Inflation and Nominal Exchange Rates

If we look at data on exchange rates and price levels of different countries, we quickly see the importance of inflation for explaining changes in the nominal exchange rate. The most dramatic examples come from periods of very high inflation. For example, the price level in Mexico rose by 2,300 percent from 1983 to 1988. Because of this inflation, the number of pesos a person could buy with a U.S. dollar rose from 144 in 1983 to 2,281 in 1988.

The same relationship holds true for countries with more moderate inflation. Figure 5-13 is a scatterplot showing the relationship between inflation and the exchange rate for 15 countries. On the horizontal axis is the difference between each country's average inflation rate and the average inflation rate of the United

figure 5-13

Inflation Differentials and the Exchange Rate This scatterplot shows the relationship between inflation and the nominal exchange rate. The horizontal axis shows the country's average inflation rate minus the U.S. average inflation rate over the period 1972–2000. The vertical axis is the average percentage change in the country's exchange rate (per U.S. dollar) over that period. This figure shows that countries with relatively high inflation tend to have depreciating currencies and that countries with relatively low inflation tend to have appreciating currencies.

Source: International Financial Statistics.

States ($\pi^* - \pi$). On the vertical axis is the average percentage change in the exchange rate between each country's currency and the U.S. dollar (% change in e). The positive relationship between these two variables is clear in this data. Countries with relatively high inflation tend to have depreciating currencies (you can buy more of them for your dollars over time), and countries with relatively low inflation tend to have appreciating currencies (you can buy less of them for your dollars over time).

As an example, consider the exchange rate between German marks and U.S. dollars. Both Germany and the United States have experienced inflation over the past twenty years, so both the mark and the dollar buy fewer goods than they once did. But, as Figure 5-13 shows, inflation in Germany has been lower than inflation in the United States. This means that the value of the mark has fallen less than the value of the dollar. Therefore, the number of German marks you can buy with a U.S. dollar has been falling over time.

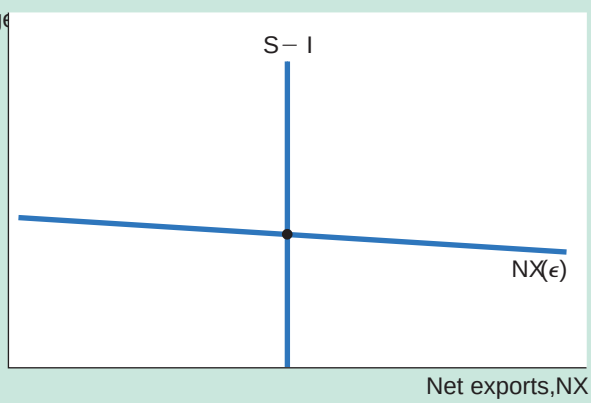
The Special Case of Purchasing-Power Parity

A famous hypothesis in economics, called the *law of one price*, states that the same good cannot sell for different prices in different locations at the same time. If a bushel of wheat sold for less in New York than in Chicago, it would be profitable to buy wheat in New York and then sell it in Chicago. Astute arbitrageurs would take advantage of such an opportunity and, thereby, would increase the demand for wheat in New York and increase the supply in Chicago. This would drive the price up in New York and down in Chicago, thereby ensuring that prices are equalized in the two markets.

The law of one price applied to the international marketplace is called *purchasing-power parity*. It states that if international arbitrage is possible, then a dollar (or any other currency) must have the same purchasing power in every country. The argument goes as follows. If a dollar could buy more wheat domestically than abroad, there would be opportunities to profit by buying wheat domestically and selling it abroad. Profit-seeking arbitrageurs would drive up the domestic price of wheat relative to the foreign price. Similarly, if a dollar could buy more wheat abroad than domestically, the arbitrageurs would buy wheat abroad and sell it domestically, driving down the domestic price relative to the foreign price. Thus, profit-seeking by international arbitrageurs causes wheat prices to be the same in all countries.

We can interpret the doctrine of purchasing-power parity using our model of the real exchange rate. The quick action of these international arbitrageurs implies that net exports are highly sensitive to small movements in the real exchange rate. A small decrease in the price of domestic goods relative to foreign goods—that is, a small decrease in the real exchange rate—causes arbitrageurs to buy goods domestically and sell them abroad. Similarly, a small increase in the relative price of domestic goods causes arbitrageurs to import goods from abroad. Therefore, as in Figure 5-14, the net-exports schedule is at the real exchange rate that equalizes purchasing power among countries: any small

figure 5-14



Purchasing-Power Parity The law of one price applied to the international marketplace suggests that net exports are highly sensitive to small movements in the real exchange rate. This high sensitivity is reflected here with a very flat net-exports schedule.

movement in the real exchange rate leads to a large change in net exports. This extreme sensitivity of net exports guarantees that the equilibrium real exchange rate is always close to the level ensuring purchasing-power parity.

Purchasing-power parity has two important implications. First, because the net-exports schedule is flat, changes in saving or investment do not influence the real or nominal exchange rate. Second, because the real exchange rate is fixed, all changes in the nominal exchange rate result from changes in price levels.

Is this doctrine of purchasing-power parity realistic? Most economists believe that, despite its appealing logic, purchasing-power parity does not provide a completely accurate description of the world. First, many goods are not easily traded. A haircut can be more expensive in Tokyo than in New York, yet there is no room for international arbitrage because it is impossible to transport haircuts. Second, even tradable goods are not always perfect substitutes. Some consumers prefer Toyotas, and others prefer Fords. Thus, the relative price of Toyotas and Fords can vary to some extent without leaving any profit opportunities. For these reasons, real exchange rates do in fact vary over time.

Although the doctrine of purchasing-power parity does not describe the world perfectly, it does provide a reason why movement in the real exchange rate will be limited. There is much validity to its underlying logic: the farther the real exchange rate drifts from the level predicted by purchasing-power parity, the greater the incentive for individuals to engage in international arbitrage in goods. Although we cannot rely on purchasing-power parity to eliminate all changes in the real exchange rate, this doctrine does provide a reason to expect that fluctuations in the real exchange rate will typically be small or temporary.[1]

CASE STUDY

The Big Mac Around the World

The doctrine of purchasing-power parity says that after we adjust for exchange rates, we should find that goods sell for the same price everywhere. Conversely, it says that the exchange rate between two currencies should depend on the price levels in the two countries.

To see how well this doctrine works, *The Economist*, an international news-magazine, regularly collects data on the price of a good sold in many countries: the McDonald's Big Mac hamburger. According to purchasing-power parity, the price of a Big Mac should be closely related to the country's nominal exchange rate. The higher the price of a Big Mac in the local currency, the higher the exchange rate (measured in units of local currency per U.S. dollar) should be.

Table 5-2 presents the international prices in 2000, when a Big Mac sold for $2.51 in the United States. With these data we can use the doctrine of purchasing-power parity to predict nominal exchange rates. For example, because a Big Mac

[1] To learn more about purchasing-power parity, see Kenneth A. Froot and Kenneth Rogoff, "Perspectives on PPP and Long-Run Real Exchange Rates," in Gene M. Grossman and Kenneth Rogoff, eds., *Handbook of International Economics*, vol. 3 (Amsterdam: North-Holland, 1995).

table 5-2

Big Mac Prices and the Exchange Rate:
An Application of Purchasing-Power Parity

Country	Currency	Price of a Big Mac	Exchange Rate (per U.S. dollar)	
			Predicted	Actual
Indonesia	Rupiah	14,500	5,777	7,945
Italy	Lira	4,500	1,793	2,088
South Korea	Won	3,000	1,195	1,108
Chile	Peso	1,260	502	514
Spain	Peseta	375	149	179
Hungary	Forint	339	135	279
Japan	Yen	294	117	106
Taiwan	Dollar	70.0	27.9	30.6
Thailand	Baht	55.0	21.9	38.0
Czech Rep.	Crown	54.37	21.7	39.1
Russia	Ruble	39.50	15.7	28.5
Denmark	Crown	24.75	9.86	8.04
Sweden	Crown	24.0	9.56	8.84
Mexico	Peso	20.9	8.33	9.41
France	Franc	18.5	7.37	7.07
Israel	Shekel	14.5	5.78	4.05
China	Yuan	9.90	3.94	8.28
South Africa	Rand	9.0	3.59	6.72
Switzerland	Franc	5.90	2.35	1.70
Poland	Zloty	5.50	2.19	4.30
Germany	Mark	4.99	1.99	2.11
Malaysia	Dollar	4.52	1.80	3.80
New Zealand	Dollar	3.40	1.35	2.01
Singapore	Dollar	3.20	1.27	1.70
Brazil	Real	2.95	1.18	1.79
Canada	Dollar	2.85	1.14	1.47
Australia	Dollar	2.59	1.03	1.68
United States	Dollar	2.51	1.00	1.00
Argentina	Peso	2.50	1.00	1.00
Britain	Pound	1.90	0.76	0.63

Note The predicted exchange rate is the exchange rate that would make the price of a Big Mac in that country equal to its price in the United States.
Source The Economist, April 29, 2000, 75.

cost 294 yen in Japan, we would predict that the exchange rate between the dollar and the yen was 294/2.51, or 117, yen per dollar. At this exchange rate, a Big Mac would have cost the same in Japan and the United States.

Table 5-2 shows the predicted and actual exchange rates for 30 countries, ranked by the predicted exchange rate. You can see that the evidence on purchasing-power

parity is mixed. As the last two columns show, the actual and predicted exchange rate are usually in the same ballpark. Our theory predicts, for instance, that a U.S. dollar should buy the greatest number of Indonesian rupiahs and fewest British pounds, and this turns out to be true. In the case of Japan, the predicted exchange rate of 117 yen per dollar is close to the actual exchange rate of 106. Yet the theory's predictions are far from exact and, in many cases, are off by 30 percent or more. Hence, although the theory of purchasing-power parity provides a rough guide to the level of exchange rates, it does not explain exchange rates completely.

5-4 Conclusion: The United States as a Large Open Economy

In this chapter we have seen how a small open economy works. We have examined the determinants of the international flow of funds for capital accumulation and the international flow of goods and services. We have also examined the determinants of a country's real and nominal exchange rates. Our analysis shows how various policies—monetary policies, fiscal policies, and trade policies—affect the trade balance and the exchange rate.

The economy we have studied is "small" in the sense that its interest rate is fixed by world financial markets. That is, we have assumed that this economy does not affect the world interest rate, and that the economy can borrow and lend at the world interest rate in unlimited amounts. This assumption contrasts with the assumption we made when we studied the closed economy in Chapter 3. In the closed economy, the domestic interest rate equilibrates domestic saving and domestic investment, implying that policies that influence saving or investment alter the equilibrium interest rate.

Which of these analyses should we apply to an economy such as the United States? The answer is a little of both. The United States is neither so large nor so isolated that it is immune to developments occurring abroad. The large trade deficits of the 1980s and 1990s show the importance of international financial markets for funding U.S. investment. Hence, the closed-economy analysis of Chapter 3 cannot by itself fully explain the impact of policies on the U.S. economy.

Yet the U.S. economy is not so small and so open that the analysis of this chapter applies perfectly either. First, the United States is large enough that it can influence world financial markets. For example, large U.S. budget deficits were often blamed for the high real interest rates that prevailed throughout the world in the 1980s. Second, capital may not be perfectly mobile across countries. If individuals prefer holding their wealth in domestic rather than foreign assets, funds for capital accumulation will not flow freely to equate interest rates in all countries. For these two reasons, we cannot directly apply our model of the small open economy to the United States.

When analyzing policy for a country such as the United States, we need to combine the closed-economy logic of Chapter 3 and the small-open-economy logic of this chapter. The appendix to this chapter builds a model of an economy between these two extremes. In this intermediate case, there is international borrowing and lending, but the interest rate is not fixed by world financial markets. Instead, the more the economy borrows from abroad, the higher the interest rate it must offer foreign investors. The results, not surprisingly, are a mixture of the two polar cases we have already examined.

Consider, for example, a reduction in national saving caused by a fiscal expansion. As in the closed economy, this policy raises the real interest rate and crowds out domestic investment. As in the small open economy, it also reduces the net capital outflow, leading to a trade deficit and an appreciation of the exchange rate. Hence, although the model of the small open economy examined here does not precisely describe an economy such as the United States, it does provide approximately the right answer to how policies affect the trade balance and the exchange rate.

Summary

1. Net exports are the difference between exports and imports. They are equal to the difference between what we produce and what we demand for consumption, investment, and government purchases.
2. The net capital outflow is the excess of domestic saving over domestic investment. The trade balance is the amount received for our net exports of goods and services. The national income accounts identity shows that the net capital outflow always equals the trade balance.
3. The impact of any policy on the trade balance can be determined by examining its impact on saving and investment. Policies that raise saving or lower investment lead to a trade surplus, and policies that lower saving or raise investment lead to a trade deficit.
4. The nominal exchange rate is the rate at which people trade the currency of one country for the currency of another country. The real exchange rate is the rate at which people trade the goods produced by the two countries. The real exchange rate equals the nominal exchange rate multiplied by the ratio of the price levels in the two countries.
5. Because the real exchange rate is the price of domestic goods relative to foreign goods, an appreciation of the real exchange rate tends to reduce net exports. The equilibrium real exchange rate is the rate at which the quantity of net exports demanded equals the net capital outflow.
6. The nominal exchange rate is determined by the real exchange rate and the price levels in the two countries. Other things equal, a high rate of inflation leads to a depreciating currency.

KEY CONCEPTS

Net exports
Trade balance
Net capital outflow
Trade surplus and trade deficit
Balanced trade
Small open economy
World interest rate
Nominal exchange rate
Real exchange rate
Purchasing-power parity

QUESTIONS FOR REVIEW

1. What are the net capital outflow and the trade balance? Explain how they are related.
2. Define the nominal exchange rate and the real exchange rate.
3. If a small open economy cuts defense spending, what happens to saving, investment, the trade balance, the interest rate, and the exchange rate?
4. If a small open economy bans the import of Japanese VCRs, what happens to saving, investment, the trade balance, the interest rate, and the exchange rate?
5. If Germany has low inflation and Italy has high inflation, what will happen to the exchange rate between the German mark and the Italian lira?

PROBLEMS AND APPLICATIONS

1. Use the model of the small open economy to predict what would happen to the trade balance, the real exchange rate, and the nominal exchange rate in response to each of the following events.
   a. A fall in consumer confidence about the future induces consumers to spend less and save more.
   b. The introduction of a stylish line of Toyotas makes some consumers prefer foreign cars over domestic cars.
   c. The introduction of automatic teller machines reduces the demand for money.
2. Consider an economy described by the following equations:

$$Y = C + I + G + NX,$$
$$Y = 5{,}000,$$
$$G = 1{,}000,$$
$$T = 1{,}000,$$
$$C = 250 + 0.75(Y - T),$$
$$I = 1{,}000 - 50r,$$
$$NX = 500 - 500\varepsilon,$$
$$r = r^* = 5.$$

   a. In this economy, solve for national saving, investment, the trade balance, and the equilibrium exchange rate.
   b. Suppose now that G rises to 1,250. Solve for national saving, investment, the trade balance, and the equilibrium exchange rate. Explain what you find.
   c. Now suppose that the world interest rate rises from 5 to 10 percent (G is again 1,000). Solve for national saving, investment, the trade balance, and the equilibrium exchange rate. Explain what you find.
3. The country of Leverett is a small open economy. Suddenly, a change in world fashions makes the exports of Leverett unpopular.
   a. What happens in Leverett to saving, investment, net exports, the interest rate, and the exchange rate?
   b. The citizens of Leverett like to travel abroad. How will this change in the exchange rate affect them?
   c. The fiscal policymakers of Leverett want to adjust taxes to maintain the exchange rate at

its previous level. What should they do? If they do this, what are the overall effects on saving, investment, net exports, and the interest rate?

4. What will happen to the trade balance and the real exchange rate of a small open economy when government purchases increase, such as during a war? Does your answer depend on whether this is a local war or a world war?

5. In 1995, President Clinton considered placing a 100-percent tariff on the import of Japanese luxury cars. Discuss the economics and politics of such a policy. In particular, how would the policy affect the U.S. trade deficit? How would it affect the exchange rate? Who would be hurt by such a policy? Who would benefit?

6. Suppose that some foreign countries begin to subsidize investment by instituting an investment tax credit.

   a. What happens to world investment demand as a function of the world interest rate?

   b. What happens to the world interest rate?

   c. What happens to investment in our small open economy?

   d. What happens to our trade balance?

   e. What happens to our real exchange rate?

7. "Traveling in Italy is much cheaper now than it was ten years ago," says a friend. "Ten years ago, a dollar bought 1,000 lire; this year, a dollar buys 1,500 lire."

   Is your friend right or wrong? Given that total inflation over this period was 25 percent in the United States and 100 percent in Italy, has it become more or less expensive to travel in Italy? Write your answer using a concrete example such as a cup of American coffee versus a cup of Italian espresso that will convince your friend.

8. You read in a newspaper that the nominal interest rate is 12 percent per year in Canada and 8 percent per year in the United States. Suppose that the real interest rates are equalized in the two countries and that purchasing-power parity holds.

   a. Using the Fisher equation (discussed in Chapter 4), what can you infer about expected inflation in Canada and in the United States?

   b. What can you infer about the expected change in the exchange rate between the Canadian dollar and the U.S. dollar?

   c. A friend proposes a get-rich-quick scheme: borrow from a U.S. bank at 8 percent, deposit the money in a Canadian bank at 12 percent, and make a 4 percent profit. What's wrong with this scheme?

APPENDIX

The Large Open Economy

When analyzing policy for a country such as the United States, we need to combine the closed-economy logic of Chapter 3 and the small-open-economy logic of this chapter. This appendix presents a model of an economy between these two extremes, called the *large open economy*.

Net Capital Outflow

The key difference between the small and large open economies is the behavior of the net capital outflow. In the model of the small open economy, capital flows freely into or out of the economy at a fixed world interest rate, r^*. The model of the large open economy makes a different assumption about international capital flows. To understand that assumption, keep in mind that the net capital outflow is the amount that domestic investors lend abroad minus the amount that foreign investors lend here.

Imagine that you are a domestic investor—such as the portfolio manager of a university endowment—deciding where to invest your funds. You could invest domestically (for example, by making loans to U.S. companies), or you could invest abroad (by making loans to foreign companies). Many factors may affect your decision, but surely one of them is the interest rate you can earn. The higher the interest rate you can earn domestically, the less attractive you would find foreign investment.

Investors abroad face a similar decision. They have a choice between investing in their home country or lending to someone in the United States. The higher the interest rate in the United States, the more willing foreigners are to lend to U.S. companies and to buy U.S. assets.

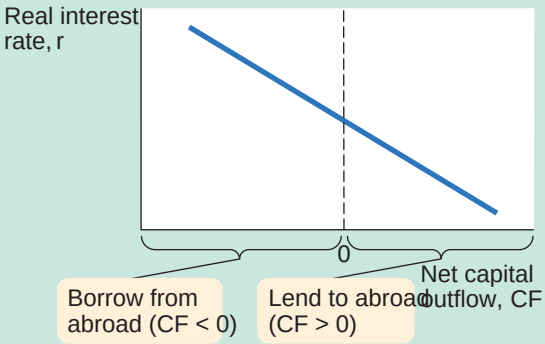
Thus, because of the behavior of both domestic and foreign investors, the net flow of capital to other countries, which we denote as CF, is negatively related to the domestic real interest rate, r. As the interest rate rises, less of our saving flows abroad, and more funds for capital accumulation flow in from other countries. We write this as

$$CF = CF(r).$$

This equation states that the net capital outflow is a function of the domestic interest rate. Figure 5-15 on page 146 illustrates this relationship. Notice that CF can be either positive or negative, depending on whether the economy is a lender or borrower in world financial markets.

To see how this CF function relates to our previous models, consider Figure 5-16 on page 146. This figure shows two special cases: a vertical CF function and a horizontal CF function.

figure 5-15

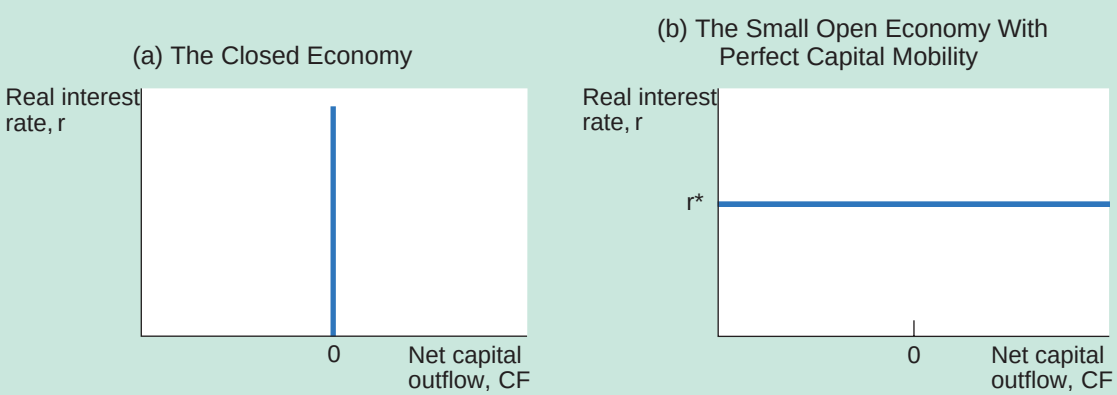


How the Net Capital Outflow Depends on the Interest Rate A higher domestic interest rate discourages domestic investors from lending abroad and encourages foreign investors to lend here. Therefore, net capital outflow CF is negatively related to the interest rate.

The closed economy is the special case shown in panel (a) of Figure 5-16. In the closed economy, there is no international borrowing or lending, and the interest rate adjusts to equilibrate domestic saving and investment. This means that CF = 0 at all interest rates. This situation would arise if investors here and abroad were unwilling to hold foreign assets, regardless of the return. It might also arise if the government prohibited its citizens from transacting in foreign financial markets, as some governments do.

The small open economy with perfect capital mobility is the special case shown in panel (b) of Figure 5-16. In this case, capital flows freely into and out of the country at the fixed world interest rate r^*. This situation would arise if investors here and abroad bought whatever asset yielded the highest

figure 5-16



Two Special Cases In the closed economy, shown in panel (a), the net capital outflow is zero for all interest rates. In the small open economy with perfect capital mobility, shown in panel (b), the net capital outflow is perfectly elastic at the world interest rate r^*.

return, and if this economy were too small to affect the world interest rate. The economy's interest rate would be fixed at the interest rate prevailing in world financial markets.

Why isn't the interest rate of a large open economy such as the United States fixed by the world interest rate? There are two reasons. The first is that the United States is large enough to influence world financial markets. The more the United States lends abroad, the greater the supply of loans in the world economy is, and the lower interest rates become around the world. The more the United States borrows from abroad (that is, the more negative CF becomes), the higher world interest rates are. We use the "large open economy" because this model applies to an economy large enough to affect world interest rates.

There is, however, a second reason that the interest rate in an economy may not be fixed by the world interest rate: capital may not be perfectly mobile. That is, investors here and abroad may prefer to hold their wealth in domestic rather than foreign assets. Such a preference for domestic assets could arise because of imperfect information about foreign assets or because of government impediments to international borrowing and lending. In either case, funds for capital accumulation will not flow freely to equalize interest rates in all countries. Instead, the net capital outflow will depend on domestic interest rates relative to foreign interest rates. U.S. investors will lend abroad only if U.S. interest rates are comparatively low, and foreign investors will lend in the United States only if U.S. interest rates are comparatively high. The large-open-economy model, therefore, may apply even to a small economy if capital does not flow freely into and out of the economy.

Hence, either because the large open economy affects world interest rates, or because capital is imperfectly mobile, or perhaps for both reasons, the net capital outflow function slopes downward. Except for this new downward-sloping CF function, the model of the large open economy resembles the model of the small open economy. We put all the pieces together in the next section.

The Model

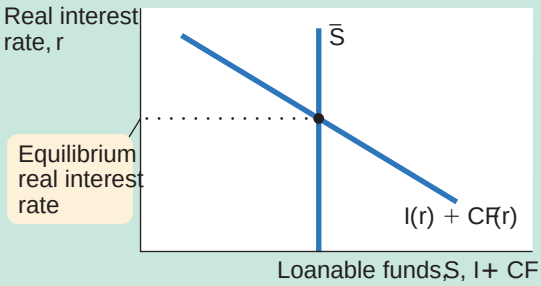
To understand how the large open economy works, we need to consider two key markets: the market for loanable funds (where the interest rate is determined) and the market for foreign exchange (where the exchange rate is determined). The interest rate and the exchange rate are two prices that guide the allocation of resources.

The Market for Loanable Funds An open economy's saving S is used in two ways: to finance domestic investment I and to finance the net capital outflow CF. We can write

$$S = I + CF.$$

Consider how these three variables are determined. National saving is fixed by the level of output, fiscal policy, and the consumption function. Investment

figure 5-17



The Market for Loanable Funds in the Large Open Economy At the equilibrium interest rate, the supply of loanable funds from saving, $\bar{S}$, balances the demand for loanable funds from domestic investment I and capital investments abroad, CF.

and net capital outflow both depend on the domestic real interest rate. We can write

$$\bar{S} = I(r) + CF(r).$$

Figure 5-17 shows the market for loanable funds. The supply of loanable funds is national saving. The demand for loanable funds is the sum of the demand for domestic investment and the demand for foreign investment (net capital outflow). The interest rate adjusts to equilibrate supply and demand.

The Market for Foreign Exchange Next, consider the relationship between the net capital outflow and the trade balance. The national income accounts identity tells us

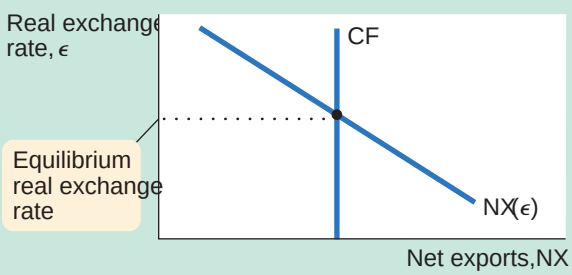
$$NX = S - I.$$

Because NX is a function of the real exchange rate, and because $CF = S - I$, we can write

$$NX(\epsilon) = CF.$$

Figure 5-18 shows the equilibrium in the market for foreign exchange. Once again, the real exchange rate is the price that equilibrates the trade balance and the net capital outflow.

figure 5-18



The Market for Foreign-Currency Exchange in the Large Open Economy At the equilibrium exchange rate, the supply of dollars from the net capital outflow, CF, balances the demand for dollars from our net exports of goods and services, NX.

The last variable we should consider is the nominal exchange rate. As before, the nominal exchange rate is the real exchange rate times the ratio of the price levels:

$$e = \epsilon \times (P^*/P).$$

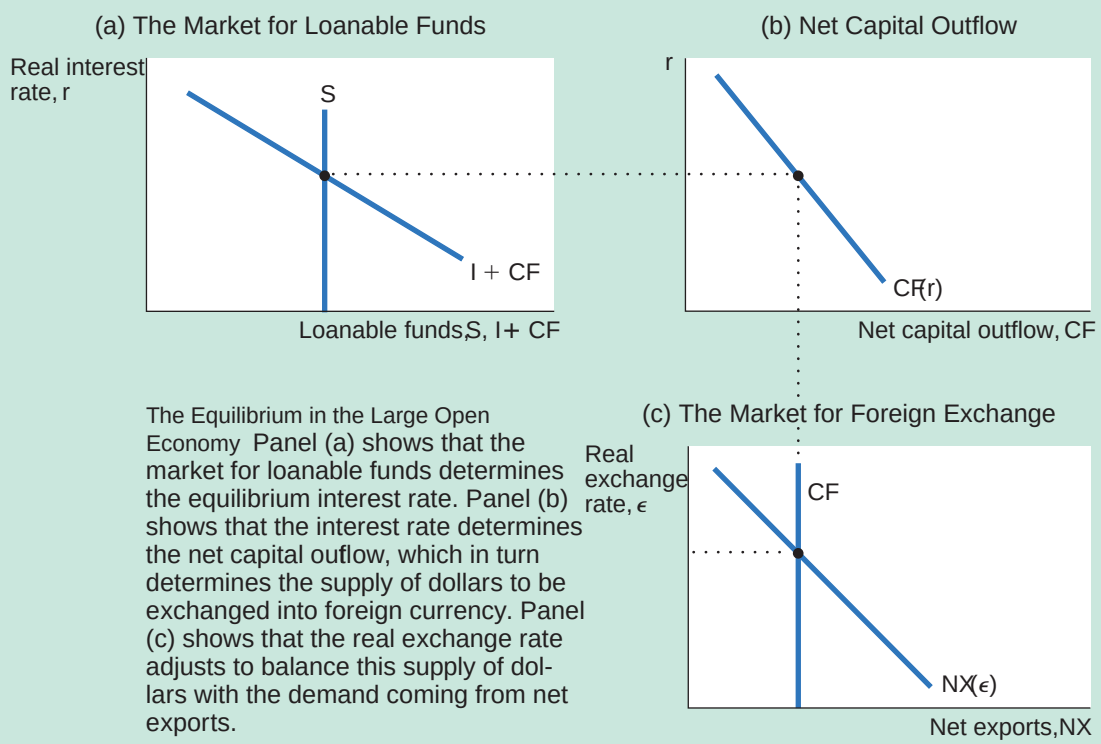
The real exchange rate is determined as in Figure 5-18, and the price levels are determined by monetary policies here and abroad, as we discussed in Chapter 4. Forces that move the real exchange rate or the price levels also move the nominal exchange rate.

Policies in the Large Open Economy

We can now consider how economic policies influence the large open economy. Figure 5-19 shows the three diagrams we need for the analysis. Panel (a) shows the equilibrium in the market for loanable funds; panel (b) shows the relationship between the equilibrium interest rate and the net capital outflow; and panel (c) shows the equilibrium in the market for foreign exchange.

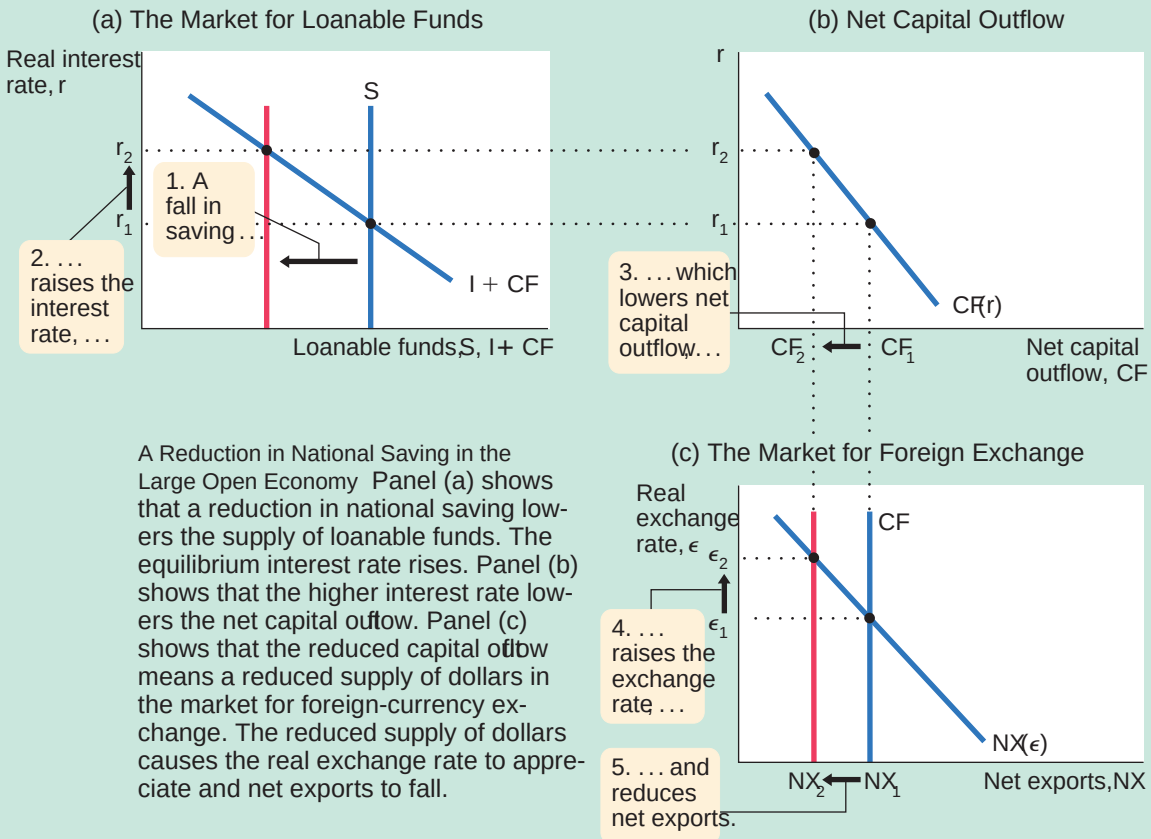
Fiscal Policy at Home Consider the effects of expansionary fiscal policy, such as an increase in government purchases or a decrease in taxes. Figure 5-20 shows what

figure 5-19



The Equilibrium in the Large Open Economy Panel (a) shows that the market for loanable funds determines the equilibrium interest rate. Panel (b) shows that the interest rate determines the net capital outflow, which in turn determines the supply of dollars to be exchanged into foreign currency. Panel (c) shows that the real exchange rate adjusts to balance this supply of dollars with the demand coming from net exports.

figure 5-20



A Reduction in National Saving in the Large Open Economy Panel (a) shows that a reduction in national saving lowers the supply of loanable funds. The equilibrium interest rate rises. Panel (b) shows that the higher interest rate lowers the net capital outflow. Panel (c) shows that the reduced capital outflow means a reduced supply of dollars in the market for foreign-currency exchange. The reduced supply of dollars causes the real exchange rate to appreciate and net exports to fall.

happens. The policy reduces national saving, thereby reducing the supply of loanable funds and raising the equilibrium interest rate. The higher interest rate reduces both domestic investment and the net capital outflow CF. The fall in the net capital outflow reduces the supply of dollars to be exchanged into foreign currency. The exchange rate appreciates, and net exports fall.

Note that the impact of fiscal policy in this model combines its impact in the closed economy and its impact in the small open economy. As in the closed economy, a fiscal expansion in a large open economy raises the interest rate and crowds out investment. As in the small open economy, a fiscal expansion causes a trade deficit and an appreciation in the exchange rate.

One way to see how the three types of economy are related is to consider the identity

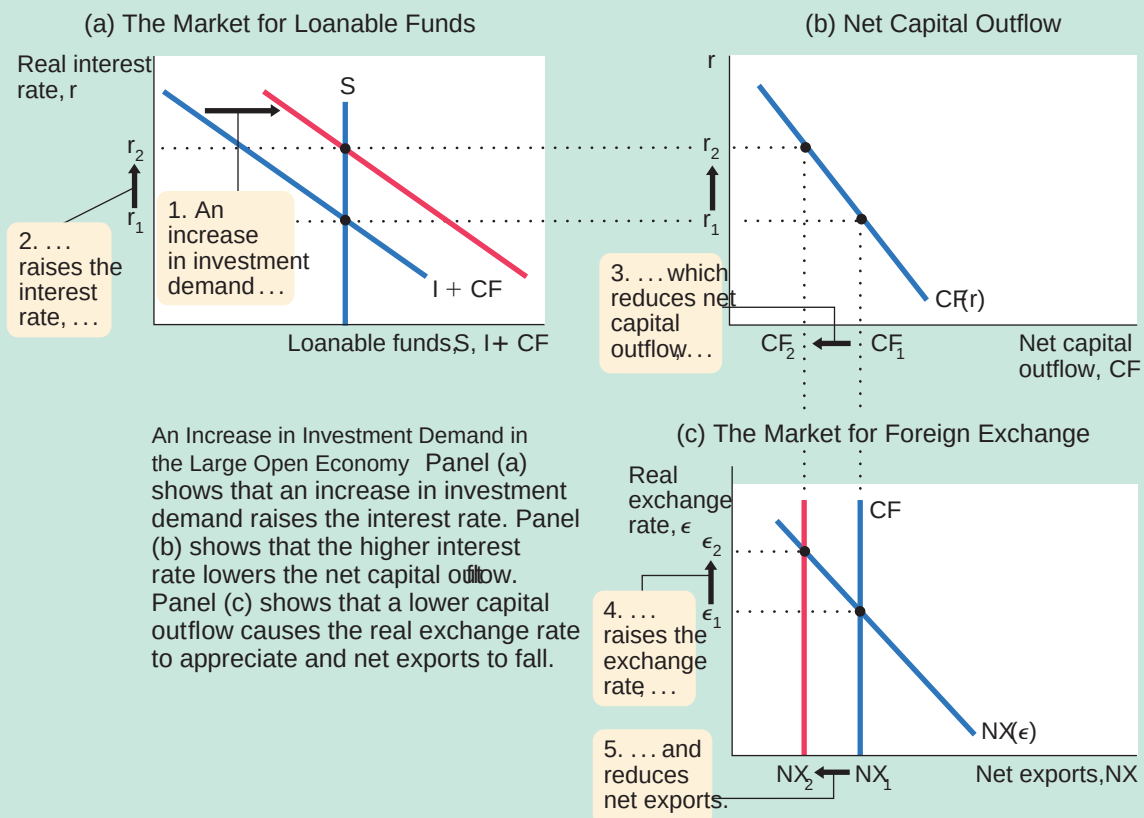
$$S = I + NX.$$

In all three cases, expansionary fiscal policy reduces national saving. In the closed economy, the fall in S coincides with an equal fall in I, and NX stays constant at zero. In the small open economy, the fall in S coincides with an equal fall

in NX, and I remains constant at the level fixed by the world interest rate. The large open economy is the intermediate case: both I and NX fall, each by less than the fall in S.

Shifts in Investment Demand Suppose that the investment demand schedule shifts outward, perhaps because Congress passes an investment tax credit. Figure 5-21 shows the effect. The demand for loanable funds rises, raising the equilibrium interest rate. The higher interest rate reduces the net capital outflow: Saving Americans make fewer loans abroad, and foreigners make more loans to Americans. The fall in the net capital outflow reduces the supply of dollars in the market for foreign exchange. The exchange rate appreciates, and net exports fall.

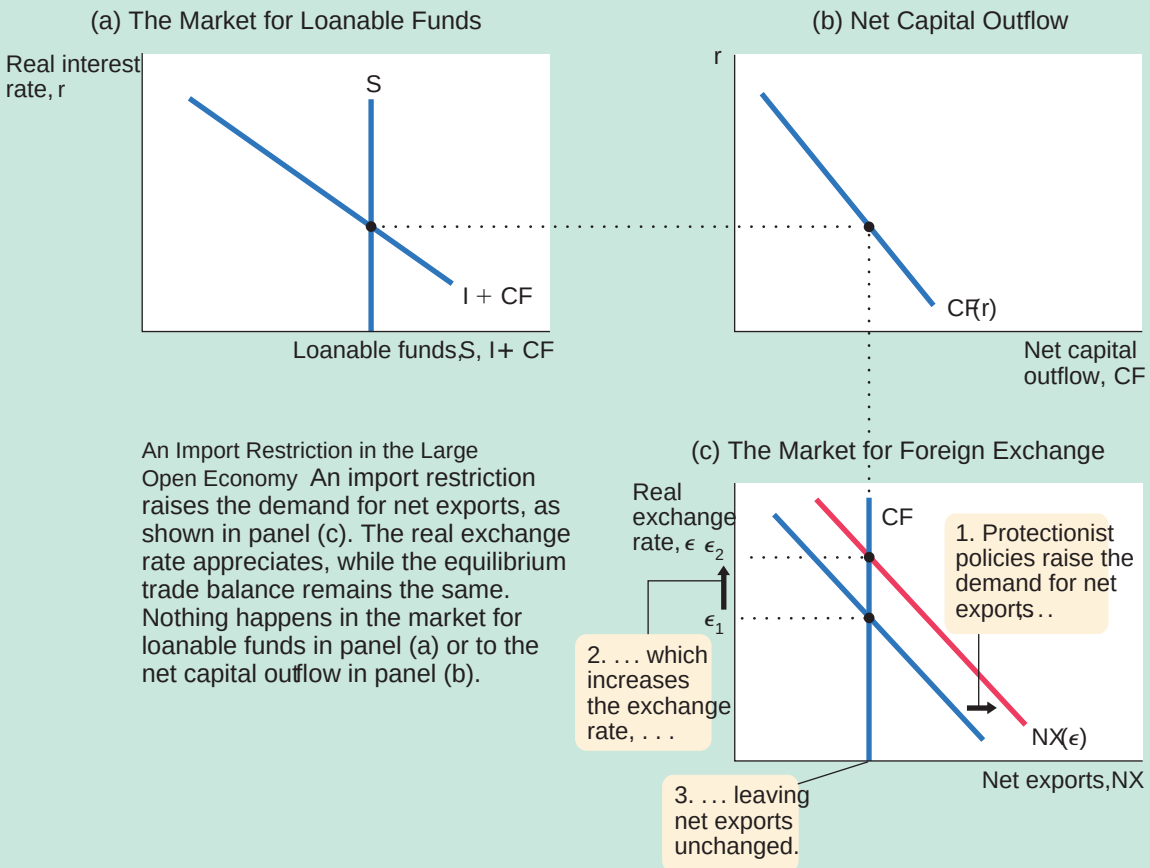
figure 5-21



An Increase in Investment Demand in the Large Open Economy Panel (a) shows that an increase in investment demand raises the interest rate. Panel (b) shows that the higher interest rate lowers the net capital outflow. Panel (c) shows that a lower capital outflow causes the real exchange rate to appreciate and net exports to fall.

Trade Policies Figure 5-22 shows the effect of a trade restriction, such as an import quota. The reduced demand for imports shifts the net-exports schedule outward in panel (c). Because nothing has changed in the market for loanable funds, the interest rate remains the same, which in turn implies that the net capital outflow remains the same. The shift in the net-exports schedule causes the exchange rate to appreciate. The rise in the exchange rate makes U.S. goods

figure 5-22



An Import Restriction in the Large Open Economy An import restriction raises the demand for net exports, as shown in panel (c). The real exchange rate appreciates, while the equilibrium trade balance remains the same. Nothing happens in the market for loanable funds in panel (a) or to the net capital outflow in panel (b).

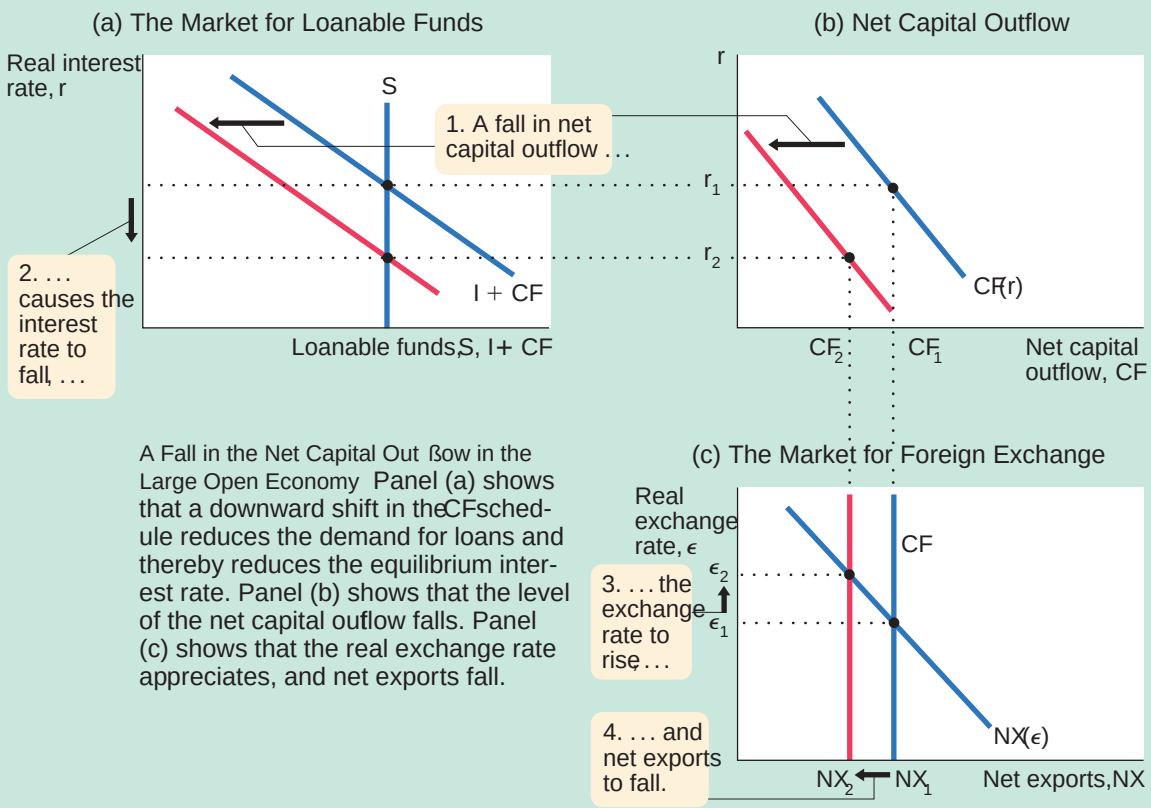
expensive relative to foreign goods, which depresses exports and stimulates imports. In the end, the trade restriction does not affect the trade balance.

Shifts in Net Capital Outflow There are various reasons that the CF schedule might shift. One reason is fiscal policy abroad. For example, suppose that Germany pursues a fiscal policy that raises German saving. This policy reduces the German interest rate. The lower German interest rate discourages American investors from lending in Germany and encourages German investors to lend in the United States. For any given U.S. interest rate, the U.S. net capital outflow falls.

Another reason the CF schedule might shift is political instability abroad. Suppose that a war or revolution breaks out in another country. Investors around the world will try to withdraw their assets from that country and seek a safe haven in a stable country such as the United States. The result is a reduction in the U.S. net capital outflow.

Figure 5-23 shows the impact of a shift in the CF schedule. The reduced demand for loanable funds lowers the equilibrium interest rate. The lower interest

figure 5-23



A Fall in the Net Capital Outflow in the Large Open Economy Panel (a) shows that a downward shift in the CF schedule reduces the demand for loans and thereby reduces the equilibrium interest rate. Panel (b) shows that the level of the net capital outflow falls. Panel (c) shows that the real exchange rate appreciates, and net exports fall.

rate tends to raise net capital outflow, but because this only partly mitigates the shift in the CF schedule, CF still falls. The reduced level of net capital outflow reduces the supply of dollars in the market for foreign exchange. The exchange rate appreciates, and net exports fall.

Conclusion

How different are large and small open economies? Certainly, policies affect the interest rate in a large open economy, unlike in a small open economy. But, in other ways, the two models yield similar conclusions. In both large and small open economies, policies that raise saving or lower investment lead to trade surpluses. Similarly, policies that lower saving or raise investment lead to trade deficits. In both economies, protectionist trade policies cause the exchange rate to appreciate and do not influence the trade balance. Because the results are so similar, for most questions one can use the simpler model of the small open economy, even if the economy being examined is not really small.

MORE PROBLEMS AND APPLICATIONS

1. If a war broke out abroad, it would affect the U.S. economy in many ways. Use the model of the large open economy to examine each of the following effects of such a war. What happens in the United States to saving, investment, the trade balance, the interest rate, and the exchange rate? (To keep things simple, consider each of the following effects separately.)
   a. The U.S. government, fearing it may need to enter the war, increases its purchases of military equipment.
   b. Other countries raise their demand for high-tech weapons, a major export of the United States.
   c. The war makes U.S. firms uncertain about the future, and the firms delay some investment projects.
   d. The war makes U.S. consumers uncertain about the future, and the consumers save more in response.
   e. Americans become apprehensive about traveling abroad, so more of them spend their vacations in the United States.
   f. Foreign investors seek a safe haven for their portfolios in the United States.
2. On September 21, 1995, House Speaker Newt Gingrich threatened to send the United States into default on its debt for the first time in the nation's history, to force the Clinton Administration to balance the budget on Republican terms (*New York Times*, September 22, 1995, A1). That same day, the interest rate on 30-year U.S. government bonds rose from 6.46 to 6.55 percent, and the dollar fell in value from 102.7 to 99.0 yen. Use the model of the large open economy to explain this event.

CHAPTER 6 SIX

Unemployment

> A man willing to work, and unable to find work, is perhaps the saddest sight that fortune's inequality exhibits under the sun.
>
> —Thomas Carlyle

Unemployment is the macroeconomic problem that affects people most directly and severely. For most people, the loss of a job means a reduced living standard and psychological distress. It is no surprise that unemployment is a frequent topic of political debate and that politicians often claim that their proposed policies would help create jobs.

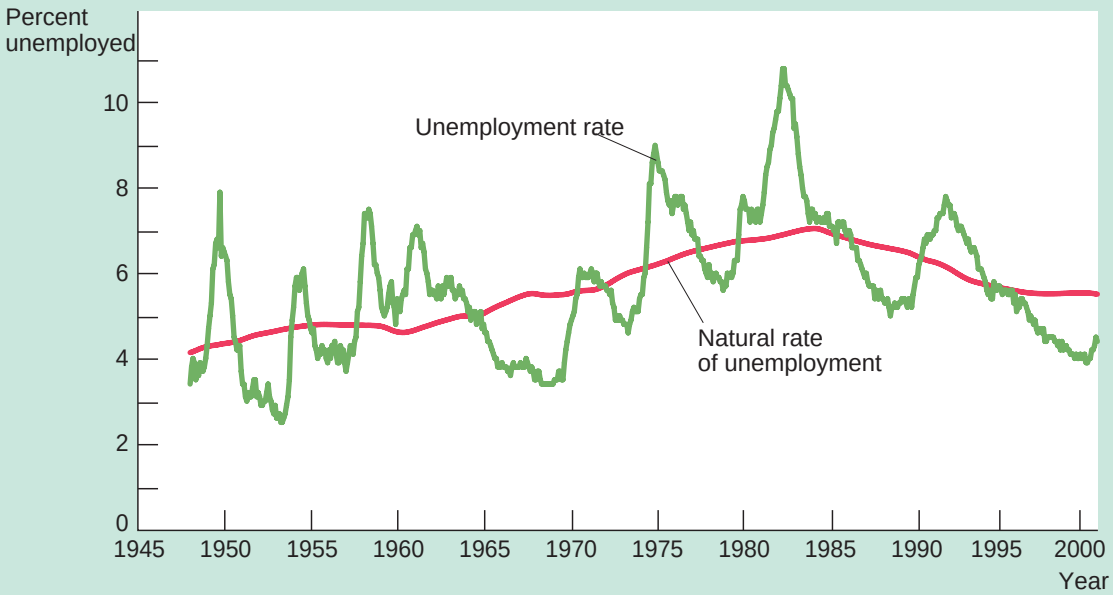
Economists study unemployment to identify its causes and to help improve the public policies that affect the unemployed. Some of these policies, such as job-training programs, assist people in finding employment. Others, such as unemployment insurance, alleviate some of the hardships that the unemployed face. Still other policies affect the prevalence of unemployment inadvertently. Laws mandating a high minimum wage, for instance, are widely thought to raise unemployment among the least skilled and experienced members of the labor force. By showing the effects of various policies, economists help policymakers evaluate their options.

Our discussions of the labor market so far have ignored unemployment. In particular, the model of national income in Chapter 3 was built with the assumption that the economy was always at full employment. In reality, of course, not everyone in the labor force has a job all the time: all free-market economies experience some unemployment.

Figure 6-1 shows the rate of unemployment—the percentage of the labor force unemployed—in the United States since 1948. Although the rate of unemployment fluctuates from year to year, it never gets even close to zero. The average is between 5 and 6 percent, meaning that about 1 out of every 18 people wanting a job does not have one.

In this chapter we begin our study of unemployment by discussing why there is always some unemployment and what determines its level. We do not study what determines the year-to-year fluctuations in the rate of unemployment until Part IV of this book, where we examine short-run economic fluctuations. Here we examine the determinants of the natural rate of unemployment—the average rate of unemployment around which the economy fluctuates. The natural

figure 6-1



The Unemployment Rate and the Natural Rate of Unemployment in the United States There is always some unemployment. The natural rate of unemployment is the average level around which the unemployment rate fluctuates. (The natural rate of unemployment for any particular year is estimated here by averaging all the unemployment rates from ten years earlier to ten years later. Future unemployment rates are set at 5.5 percent.)

rate is the rate of unemployment toward which the economy gravitates in the long run, given all the labor-market imperfections that impede workers from instantly finding jobs.

6-1 Job Loss, Job Finding, and the Natural Rate of Unemployment

Every day some workers lose or quit their jobs, and some unemployed workers are hired. This perpetual ebb and flow determines the fraction of the labor force that is unemployed. In this section we develop a model of labor-force dynamics that shows what determines the natural rate of unemployment.[1]

We start with some notation. Let L denote the labor force, E the number of employed workers, and U the number of unemployed workers. Because every

[1] Robert E. Hall, "A Theory of the Natural Rate of Unemployment and the Duration of Unemployment," Journal of Monetary Economics 5 (April 1979): 153–169.

worker is either employed or unemployed, the labor force is the sum of the employed and the unemployed:

$$L = E + U.$$

In this notation, the rate of unemployment is U/L.

To see what determines the unemployment rate, we assume that the labor force L is fixed and focus on the transition of individuals in the labor force between employment and unemployment. This is illustrated in Figure 6-2. Let s denote the rate of job separation, the fraction of employed individuals who lose their job each month. Let f denote the rate of job finding, the fraction of unemployed individuals who find a job each month. Together, the rate of job separation s and the rate of job finding f determine the rate of unemployment.

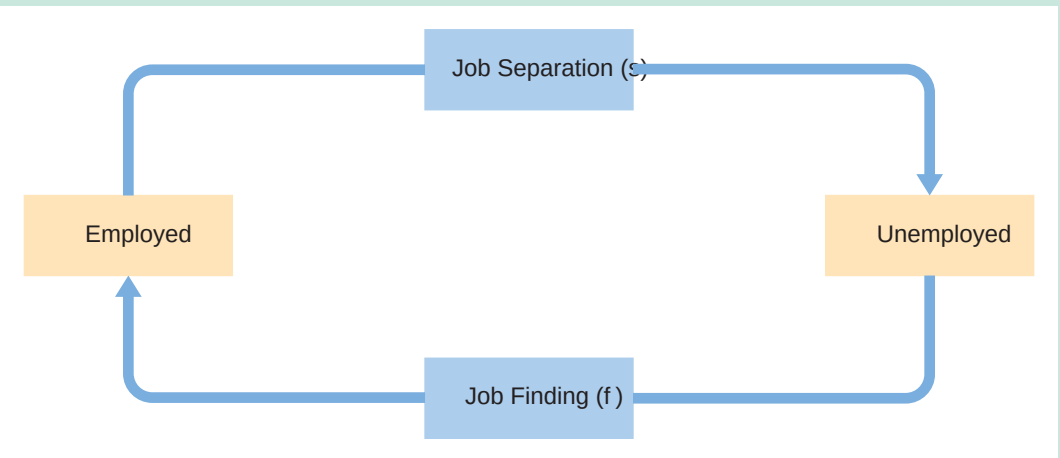
If the unemployment rate is neither rising nor falling—that is, if the labor market is in a steady state—then the number of people finding jobs must equal the number of people losing jobs. The number of people finding jobs is fU and the number of people losing jobs is sE, so we can write the steady-state condition as

$$fU = sE.$$

We can use this equation to find the steady-state unemployment rate. From an earlier equation, we know that $E = L - U$; that is, the number of employed equals the labor force minus the number of unemployed. If we substitute $(L - U)$ for E in the steady-state condition, we find

$$fU = s(L - U).$$

figure 6-2



The Transitions Between Employment and Unemployment In every period, a fraction s of the employed lose their jobs, and a fraction f of the unemployed find jobs. The rates of job separation and job finding determine the rate of unemployment.

To get closer to solving for the unemployment rate, divide both sides of this equation by L to obtain

$$f \frac{U}{L} = s(1 - \frac{U}{L}).$$

Now we can solve for U/L to find

$$\frac{U}{L} = \frac{s}{s + f}.$$

This equation shows that the steady-state rate of unemployment U/L depends on the rates of job separation s and job finding f. The higher the rate of job separation, the higher the unemployment rate. The higher the rate of job finding, the lower the unemployment rate.

Here's a numerical example. Suppose that 1 percent of the employed lose their jobs each month ($s = 0.01$). This means that on average jobs last 100 months, or about 8 years. Suppose further that about 20 percent of the unemployed find a job each month ($f = 0.20$), so that spells of unemployment last 5 months on average. Then the steady-state rate of unemployment is

$$\frac{U}{L} = \frac{0.01}{0.01 + 0.20}$$
$$= 0.0476.$$

The rate of unemployment in this example is about 5 percent.

This model of the natural rate of unemployment has an obvious but important implication for public policy. Any policy aimed at lowering the natural rate of unemployment must either reduce the rate of job separation or increase the rate of job finding. Similarly, any policy that affects the rate of job separation or job finding also changes the natural rate of unemployment.

Although this model is useful in relating the unemployment rate to job separation and job finding, it fails to answer a central question: Why is there unemployment in the first place? If a person could always find a job quickly, then the rate of job finding would be very high and the rate of unemployment would be near zero. This model of the unemployment rate assumes that job finding is not instantaneous, but it fails to explain why. In the next two sections, we examine two underlying reasons for unemployment: job search and wage rigidity.

6-2 Job Search and Frictional Unemployment

One reason for unemployment is that it takes time to match workers and jobs. The equilibrium model of the aggregate labor market discussed in Chapter 3 assumes that all workers and all jobs are identical, and therefore that all workers are

equally well suited for all jobs. If this were true and the labor market were in equilibrium, then a job loss would not cause unemployment: a laid-off worker would immediately find a new job at the market wage.

In fact, workers have different preferences and abilities, and jobs have different attributes. Furthermore, the flow of information about job candidates and job vacancies is imperfect, and the geographic mobility of workers is not instantaneous. For all these reasons, searching for an appropriate job takes time and effort, and this tends to reduce the rate of job finding. Indeed, because different jobs require different skills and pay different wages, unemployed workers may not accept the first job offer they receive. The unemployment caused by the time it takes workers to search for a job is called *frictional unemployment*.

Some frictional unemployment is inevitable in a changing economy. For many reasons, the types of goods that firms and households demand vary over time. As the demand for goods shifts, so does the demand for the labor that produces those goods. The invention of the personal computer, for example, reduced the demand for typewriters and, as a result, for labor by typewriter manufacturers. At the same time, it increased the demand for labor in the electronics industry. Similarly, because different regions produce different goods, the demand for labor may be rising in one part of the country and falling in another. A decline in the price of oil may cause the demand for labor to fall in oil-producing states such as Texas, but because cheap oil makes driving less expensive, it increases the demand for labor in auto-producing states such as Michigan. Economists call a change in the composition of demand among industries or regions a *sectoral shift*. Because sectoral shifts are always occurring, and because it takes time for workers to change sectors, there is always frictional unemployment.

Sectoral shifts are not the only cause of job separation and frictional unemployment. In addition, workers find themselves unexpectedly out of work when their firms fail, when their job performance is deemed unacceptable, or when their particular skills are no longer needed. Workers also may quit their jobs to change careers or to move to different parts of the country. As long as the supply and demand for labor among firms is changing, frictional unemployment is unavoidable.

Public Policy and Frictional Unemployment

Many public policies seek to decrease the natural rate of unemployment by reducing frictional unemployment. Government employment agencies disseminate information about job vacancies in order to match jobs and workers more efficiently. Publicly funded retraining programs are designed to ease the transition of workers from declining to growing industries. If these programs succeed at increasing the rate of job finding, they decrease the natural rate of unemployment.

Other government programs inadvertently increase the amount of frictional unemployment. One of these is *unemployment insurance*. Under

this program, unemployed workers can collect a fraction of their wages for a certain period after losing their jobs. Although the precise terms of the program differ from year to year and from state to state, a typical worker covered by unemployment insurance in the United States receives 50 percent of his or her former wages for 26 weeks. In many European countries, unemployment-insurance programs are even more generous.

By softening the economic hardship of unemployment, unemployment insurance increases the amount of frictional unemployment and raises the natural rate. The unemployed who receive unemployment-insurance benefits are less pressed to search for new employment and are more likely to turn down unattractive job offers. Both of these changes in behavior reduce the rate of job finding. In addition, because workers know that their incomes are partially protected by unemployment insurance, they are less likely to seek jobs with stable employment prospects and are less likely to bargain for guarantees of job security. These behavioral changes raise the rate of job separation.

That unemployment insurance raises the natural rate of unemployment does not necessarily imply that the policy is ill advised. The program has the benefit of reducing workers' uncertainty about their incomes. Moreover, inducing workers to reject unattractive job offers may lead to a better matching between workers and jobs. Evaluating the costs and benefits of different systems of unemployment insurance is a difficult task that continues to be a topic of much research.

Economists who study unemployment insurance often propose reforms that would reduce the amount of unemployment. One common proposal is to require a firm that lays off a worker to bear the full cost of that worker's unemployment benefits. Such a system is called 100 percent experience rated because the rate that each firm pays into the unemployment-insurance system fully reflects the unemployment experience of its own workers. Most current programs are only partially experience rated. Under this system, when a firm lays off a worker, it is charged for only part of the worker's unemployment benefits; the remainder comes from the program's general revenue. Because a firm pays only a fraction of the cost of the unemployment it causes, it has an incentive to lay off workers when its demand for labor is temporarily low. By reducing that incentive, the proposed reform may reduce the prevalence of temporary layoffs.

CASE STUDY

Unemployment Insurance and the Rate of Job Finding

Many studies have examined the effect of unemployment insurance on job search. The most persuasive studies use data on the experiences of unemployed individuals, rather than economy-wide rates of unemployment. Individual data often yield sharp results that are open to few alternative explanations.

One study followed the experience of individual workers as they used up their eligibility for unemployment-insurance benefits. It found that when

unemployed workers become ineligible for benefits, they are more likely to find new jobs. In particular, the probability of a person finding a new job more than doubles when his or her benefits run out. One possible explanation is that an absence of benefits increases the search effort of unemployed workers. Another possibility is that workers without benefits are more likely to accept job offers that would otherwise be declined because of low wages or poor working conditions.[2]

Additional evidence on how economic incentives affect job search comes from an experiment that the state of Illinois ran in 1985. Randomly selected new claimants for unemployment insurance were each offered a $500 bonus if they found employment within 11 weeks. The subsequent experience of this group was compared to that of a control group not offered the incentive. The average duration of unemployment for the group offered the $500 bonus was 17.0 weeks, compared to 18.3 weeks for the control group. Thus, the bonus reduced the average spell of unemployment by 7 percent, suggesting that more effort was devoted to job search. This experiment shows clearly that the incentives provided by the unemployment-insurance system affect the rate of job finding.[3]

6-3 Real-Wage Rigidity and Structural Unemployment

A second reason for unemployment is **wage rigidity**—the failure of wages to adjust until labor supply equals labor demand. In the equilibrium model of the labor market, as outlined in Chapter 3, the real wage adjusts to equilibrate supply and demand. Yet wages are not always flexible. Sometimes the real wage is stuck above the market-clearing level.

Figure 6-3 shows why wage rigidity leads to unemployment. When the real wage is above the level that equilibrates supply and demand, the quantity of labor supplied exceeds the quantity demanded. Firms must in some way ration the scarce jobs among workers. Real-wage rigidity reduces the rate of job finding and raises the level of unemployment.

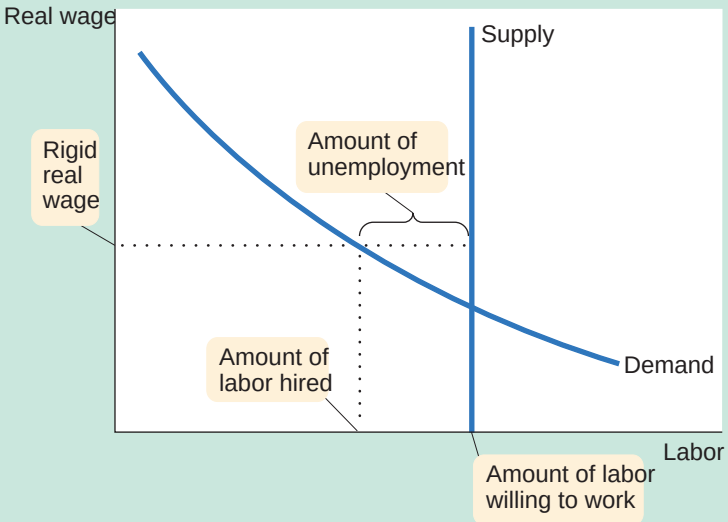
The unemployment resulting from wage rigidity and job rationing is called **structural unemployment**. Workers are unemployed not because they are actively searching for the jobs that best suit their individual skills but because, at the going wage, the supply of labor exceeds the demand. These workers are simply waiting for jobs to become available.

---

[2] Lawrence F. Katz and Bruce D. Meyer, "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes," *Quarterly Journal of Economics* 105 (November 1990): 973–1002.

[3] Stephen A. Woodbury and Robert G. Spiegelman, "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois," *American Economic Review* 77 (September 1987): 513–530.

figure 6-3



Real-Wage Rigidity Leads to Job Rationing If the real wage is stuck above the equilibrium level, then the supply of labor exceeds the demand. The result is unemployment.

To understand wage rigidity and structural unemployment, we must examine why the labor market does not clear. When the real wage exceeds the equilibrium level and the supply of workers exceeds the demand, we might expect firms to lower the wages they pay. Structural unemployment arises because firms fail to reduce wages despite an excess supply of labor. We now turn to three causes of this wage rigidity: minimum-wage laws, the monopoly power of unions, and efficiency wages.

Minimum-Wage Laws

The government causes wage rigidity when it prevents wages from falling to equilibrium levels. Minimum-wage laws set a legal minimum on the wages that firms pay their employees. Since the passage of the Fair Labor Standards Act of 1938, the U.S. federal government has enforced a minimum wage that usually has been between 30 and 50 percent of the average wage in manufacturing. For most workers, this minimum wage is not binding, because they earn well above the minimum. Yet for some workers, especially the unskilled and inexperienced, the minimum wage raises their wage above its equilibrium level. It therefore reduces the quantity of their labor that firms demand.

Economists believe that the minimum wage has its greatest impact on teenage unemployment. The equilibrium wages of teenagers tend to be low for two reasons. First, because teenagers are among the least skilled and least experienced members of the labor force, they tend to have low marginal productivity. Second, teenagers often take some of their compensation in the

form of on-the-job training rather than direct pay. An apprenticeship is a classic example of training offered in place of wages. For both these reasons, the wage at which the supply of teenage workers equals the demand is low. The minimum wage is therefore more often binding for teenagers than for others in the labor force.

Many economists have studied the impact of the minimum wage on teenage employment. These researchers compare the variation in the minimum wage over time with the variation in the number of teenagers with jobs. These studies find that a 10-percent increase in the minimum wage reduces teenage employment by 1 to 3 percent.[4]

The minimum wage is a perennial source of political debate. Advocates of a higher minimum wage view it as a means of raising the income of the working poor. Certainly, the minimum wage provides only a meager standard of living: in the United States, two adults working full time at minimum-wage jobs would just exceed the official poverty level for a family of four. Although minimum-wage advocates often admit that the policy causes unemployment for some workers, they argue that this cost is worth bearing to raise others out of poverty.

Opponents of a higher minimum wage claim that it is not the best way to help the working poor. They contend not only that the increased labor costs would raise unemployment but also that the minimum wage is poorly targeted. Many minimum-wage earners are teenagers from middle-class homes working for discretionary spending money. Of the approximately 3 million workers who earn the minimum wage, more than one-third are teenagers.

To mitigate the effects on teenage unemployment, some economists and policymakers have long advocated exempting young workers from the regular minimum wage. This would permit a lower wage for teenagers, thereby reducing their unemployment and enabling them to get training and job experience. Opponents of this exemption argue that it gives firms an incentive to substitute teenagers for unskilled adults, thereby raising unemployment among that group. A limited exemption of this kind was tried from 1991 to 1993. Because of many restrictions on its use, however, it had only limited effect and, therefore, was not renewed by Congress.

Many economists and policymakers believe that tax credits are a better way to increase the incomes of the working poor. The earned income tax credit is an amount that poor working families are allowed to subtract from the taxes they owe. For a family with a very low income, the credit exceeds its taxes, and the family receives a payment from the government. Unlike the minimum wage, the earned income tax credit does not raise labor costs to firms and, therefore, does not reduce the quantity of labor that firms demand. It has the disadvantage, however, of reducing the government's tax revenue.

---

[4] Charles Brown, "Minimum Wage Laws: Are They Overrated?" *Journal of Economic Perspectives* 2 (Summer 1988): 133–146.

CASE STUDY

A Revisionist View of the Minimum Wage

Although most economists believe that increases in the minimum wage reduce employment among workers with little skill and experience, some recent studies question this conclusion. Three respected labor economists—David Card, Lawrence Katz, and Alan Krueger—examined several instances of minimum-wage changes in order to determine the magnitude of the employment response. What they found was startling.

One study examined hiring by fast-food restaurants in New Jersey when New Jersey raised the state minimum wage. Fast-food restaurants are a natural type of firm to examine because they employ many low-wage workers. To control for other effects, such as overall economic conditions, the New Jersey restaurants were compared to similar restaurants across the river in Pennsylvania. Pennsylvania did not raise its minimum wage at the same time. According to standard theory, employment in New Jersey restaurants should have fallen relative to employment in Pennsylvania restaurants. In contrast to this hypothesis, the data showed that employment rose in the New Jersey restaurants.

How is this seemingly perverse result possible? One explanation is that firms have some market power in the labor market. As you may have learned in courses in microeconomics, a monopsony firm buys less labor at a lower wage than a competitive firm would. In essence, the firm reduces employment in order to depress the wage it has to pay. A minimum wage prevents the monopsony firm from following this strategy and so (up to a point) can increase employment.

This new view of the minimum wage is controversial. Critics have questioned the reliability of the data used in the New Jersey study. Some studies using other data sources have reached the traditional conclusion that the minimum wage depresses employment. Moreover, most economists are skeptical of the monopsony explanation, because most firms compete with many other firms for workers. Yet this new view has directly affected the policy debate. Lawrence Katz was first the chief economist in the Department of Labor during the Clinton administration. He was followed in this job by Alan Krueger. It is therefore not surprising that President Clinton supported increases in the national minimum wage.[5]

Unions and Collective Bargaining

A second cause of wage rigidity is the monopoly power of unions. Table 6-1 shows the importance of unions in 12 major countries. In the United States, only 16 percent of workers belong to unions. In most European countries, unions play a much larger role.

---

[5] To read more about this new view of the minimum wage, see David Card and Alan Krueger, Myth and Measurement: The New Economics of the Minimum Wage (Princeton, N.J.: Princeton University Press, 1995); and Lawrence Katz and Alan Krueger, "The Effect of the Minimum Wage on the Fast-Food Industry," Industrial and Labor Relations Review 46 (October 1992): 6–21.

table 6-1

Union Membership as a Percentage of Employment

Country	Percentage Union Workers	Country	Percentage Union Workers
Sweden	84	Germany	33
Denmark	75	Netherlands	28
Italy	47	Switzerland	28
United Kingdom	41	Japan	26
Australia	34	United States	16
Canada	33	France	11

Source: Clara Chang and Constance Sorrentino, "Union Membership Statistics in 12 Countries," Monthly Labor Review (December 1991): 46–53.

The wages of unionized workers are determined not by the equilibrium of supply and demand but by collective bargaining between union leaders and management. Often, the final agreement raises the wage above the equilibrium level and allows the firm to decide how many workers to employ. The result is a reduction in the number of workers hired, a lower rate of job-finding, and an increase in structural unemployment.

Unions can also influence the wages paid by firms whose workforces are not unionized because the threat of unionization can keep wages above the equilibrium level. Most firms dislike unions. Unions not only raise wages but also increase the bargaining power of labor on many other issues, such as hours of employment and working conditions. A firm may choose to pay its workers high wages to keep them happy in order to discourage them from forming a union.

The unemployment caused by unions and by the threat of unionization is an instance of conflict between different groups of workers—insiders and outsiders. Those workers already employed by a firm, the insiders, typically try to keep their firm's wages high. The unemployed, the outsiders, bear part of the cost of higher wages because at a lower wage they might be hired. These two groups inevitably have conflicting interests. The effect of any bargaining process on wages and employment depends crucially on the relative influence of each group.

The conflict between insiders and outsiders is resolved differently in different countries. In some countries, such as the United States, wage bargaining takes place at the level of the firm or plant. In other countries, such as Sweden, wage bargaining takes place at the national level, with the government often playing a key role. Despite a highly unionized labor force, Sweden has not experienced extraordinarily high unemployment throughout its history. One possible explanation is that the centralization of wage bargaining and the role of the government in the bargaining process give more influence to the outsiders, which keeps wages closer to the equilibrium level.

Efficiency Wages

Efficiency-wage theories propose a third cause of wage rigidity in addition to minimum-wage laws and unionization. These theories hold that high wages make workers more productive. The influence of wages on worker efficiency may explain the failure of firms to cut wages despite an excess supply of labor. Even though a wage reduction would lower a firm's wage bill, it would also—if these theories are correct—lower worker productivity and the firm's profits.

Economists have proposed various theories to explain how wages affect worker productivity. One efficiency-wage theory, which is applied mostly to poorer countries, holds that wages influence nutrition. Better-paid workers can afford a more nutritious diet, and healthier workers are more productive. A firm may decide to pay a wage above the equilibrium level to maintain a healthy workforce. Obviously, this consideration is not important for employers in wealthy countries, such as the United States and most of Europe, because the equilibrium wage is well above the level necessary to maintain good health.

A second efficiency-wage theory, which is more relevant for developed countries, holds that high wages reduce labor turnover. Workers quit jobs for many reasons—to accept better positions at other firms, to change careers, or to move to other parts of the country. The more a firm pays its workers, the greater their incentive to stay with the firm. By paying a high wage, a firm reduces the frequency of quits, thereby decreasing the time spent hiring and training new workers.

A third efficiency-wage theory holds that the average quality of a firm's workforce depends on the wage it pays its employees. If a firm reduces its wage, the best employees may take jobs elsewhere, leaving the firm with inferior employees who have fewer alternative opportunities. Economists recognize this unfavorable sorting as an example of **adverse selection**—the tendency of people with more information (in this case, the workers, who know their own outside opportunities) to self-select in a way that disadvantages people with less information (the firm). By paying a wage above the equilibrium level, the firm may reduce adverse selection, improve the average quality of its workforce, and thereby increase productivity.

A fourth efficiency-wage theory holds that a high wage improves worker effort. This theory posits that firms cannot perfectly monitor their employees' work effort, and that employees must themselves decide how hard to work. Workers can choose to work hard, or they can choose to shirk and risk getting caught and fired. Economists recognize this possibility as an example of **moral hazard**—the tendency of people to behave inappropriately when their behavior is imperfectly monitored. The firm can reduce the problem of moral hazard by paying a high wage. The higher the wage, the greater the cost to the worker of getting fired. By paying a higher wage, a firm induces more of its employees not to shirk and thus increases their productivity.

Although these four efficiency-wage theories differ in detail, they share a common theme: because a firm operates more efficiently if it pays its workers a high wage, the firm may find it profitable to keep wages above the level that balances supply and demand. The result of this higher-than-equilibrium wage is a lower rate of job finding and greater unemployment.

CASE STUDY

Henry Ford's $5 Workday

In 1914 the Ford Motor Company started paying its workers $5 per day. The prevailing wage at the time was between $2 and $3 per day, so Ford's wage was well above the equilibrium level. Not surprisingly, long lines of job seekers waited outside the Ford plant gates hoping for a chance to earn this high wage.

What was Ford's motive? Henry Ford later wrote, "We wanted to pay these wages so that the business would be on a lasting foundation. We were building for the future. A low wage business is always insecure. . . . The payment of five dollars a day for an eight hour day was one of the finest cost cutting moves we ever made."

From the standpoint of traditional economic theory, Ford's explanation seems peculiar. He was suggesting that high wages imply low costs. But perhaps Ford had discovered efficiency-wage theory. Perhaps he was using the high wage to increase worker productivity.

Evidence suggests that paying such a high wage did benefit the company. According to an engineering report written at the time, "The Ford high wage does away with all the inertia and living force resistance. The workingmen are absolutely docile, and it is safe to say that since the last day of 1913, every single day has seen major reductions in Ford shop labor costs." Absenteeism fell by 75 percent, suggesting a large increase in worker effort. Alan Nevins, a historian who studied the early Ford Motor Company, wrote, "Ford and his associates freely declared on many occasions that the high wage policy had turned out to be good business. By this they meant that it had improved the discipline of the workers, given them a more loyal interest in the institution, and raised their personal efficiency."[7]

6-4 Patterns of Unemployment

So far we have developed the theory behind the natural rate of unemployment. We began by showing that the economy's steady-state unemployment rate depends on the rates of job separation and job finding. Then we discussed two reasons why job finding is not instantaneous: the process of job search (which leads to frictional unemployment) and wage rigidity (which leads to structural unemployment). Wage rigidity, in turn, arises from minimum-wage laws, unionization, and efficiency wages.

With these theories as background, we now examine some additional facts about unemployment. These facts will help us evaluate our theories and assess public policies aimed at reducing unemployment.

---

[6] For more extended discussions of efficiency wages, see Janet Yellen, "Efficiency Wage Models of Unemployment," *American Economic Review Papers and Proceedings* (May 1984): 200–205; and Lawrence Katz, "Efficiency Wages: A Partial Evaluation," *NBER Macroeconomics Annual* (1986): 235–276.

[7] Jeremy I. Bulow and Lawrence H. Summers, "A Theory of Dual Labor Markets With Application to Industrial Policy, Discrimination, and Keynesian Unemployment," *Journal of Labor Economics* (July 1986): 376–414; and Daniel M. G. Raff and Lawrence H. Summers, "Did Henry Ford Pay Efficiency Wages?" *Journal of Labor Economics* (October 1987, Part 2): S57–S86.

The Duration of Unemployment

When a person becomes unemployed, is the spell of unemployment likely to be short or long? The answer to this question is important because it indicates the reasons for the unemployment and what policy response is appropriate. On the one hand, if most unemployment is short term, one might argue that it is frictional and perhaps unavoidable. Unemployed workers may need some time to search for the job that is best suited to their skills and tastes. On the other hand, long-term unemployment cannot easily be attributed to the time it takes to match jobs and workers: we would not expect this matching process to take many months. Long-term unemployment is more likely to be structural unemployment. Thus, data on the duration of unemployment can affect our view about the reasons for unemployment.

The answer to our question turns out to be subtle. The data show that most spells of unemployment are short but that most weeks of unemployment are attributable to the long-term unemployed. Consider the data for a typical year, 1974, during which the unemployment rate was 5.6 percent. In that year, 60 percent of the spells of unemployment ended within one month, yet 69 percent of the weeks of unemployment occurred in spells that lasted two or more months.[8]

To see how both these facts can be true, consider the following example. Suppose that 10 people are unemployed for part of a given year. Of these 10 people, 8 are unemployed for 1 month, and 2 are unemployed for 12 months, totaling 32 months of unemployment. In this example, most spells of unemployment are short: 8 of the 10 unemployment spells, or 80 percent, end in 1 month. Yet most months of unemployment are attributable to the long-term unemployed: 24 of the 32 months of unemployment, or 75 percent, are experienced by the 2 workers who are unemployed for 12 months. Depending on whether we look at spells of unemployment or months of unemployment, most unemployment can appear to be short term or long term.

This evidence on the duration of unemployment has an important implication for public policy. If the goal is to lower substantially the natural rate of unemployment, policies must aim at the long-term unemployed, because these individuals account for a large amount of unemployment. Yet policies must be carefully targeted, because the long-term unemployed constitute a small minority of those who become unemployed. Most people who become unemployed find work within a short time.

Variation in the Unemployment Rate Across Demographic Groups

The rate of unemployment varies substantially across different groups within the population. Table 6-2 presents the U.S. unemployment rates for different demographic groups in 2000, when the overall unemployment rate was 4.0 percent.

This table shows that younger workers have much higher unemployment rates than older ones. To explain this difference, recall our model of the natural rate of

[8] Kim B. Clark and Lawrence H. Summers, "Labor Market Dynamics and Unemployment: A Reconsideration," *Brookings Papers on Economic Activity* (1979:1): 13–72.

table 6-2

Unemployment Rate by Demographic Group: 2000

Age	White Male	White Female	Black Male	Black Female
16–19	12.3	10.4	26.4	23.0
20 and over	2.8	3.1	7.0	6.3

Source: U.S. Department of Labor.

unemployment. The model isolates two possible causes for a high rate of unemployment: a low rate of job finding and a high rate of job separation. When economists study data on the transition of individuals between employment and unemployment, they find that those groups with high unemployment tend to have high rates of job separation. They find less variation across groups in the rate of job finding. For example, employed white males are four times more likely to become unemployed if they are teenagers than if they are middle-aged; once someone is unemployed, the rate of finding a job is not closely related to age.

These findings help explain the higher unemployment rates for younger workers. Younger workers have only recently entered the labor market, and they are often uncertain about their career plans. It may be best for them to try different types of jobs before making a long-term commitment to a specific occupation. If so, we should expect a higher rate of job separation and a higher rate of frictional unemployment for this group.

Another fact that stands out from Table 6-2 is that unemployment rates are much higher for blacks than for whites. This phenomenon is not well understood. Data on transitions between employment and unemployment show that the higher unemployment rates for blacks, and especially for black teenagers, arise because of both higher rates of job separation and lower rates of finding jobs. Possible reasons for the lower rates of finding jobs include less access to informal job-finding networks and discrimination by employers.

Trends in U.S. Unemployment

Over the past half century, the natural rate of unemployment in the United States has not been stable. If you look back at Figure 6-1, you will see that unemployment averaged below 5 percent in the 1950s and 1960s, rose to over 6 percent in the 1970s and 1980s, and then drifted back below 5 percent in the 1990s. Although economists do not have a conclusive explanation for these changes, they have proposed several hypotheses.

Demographics One explanation stresses the changing composition of the U.S. labor force. After World War II, birthrates rose dramatically: the number of births rose from 2.9 million in 1945 to a peak of 4.3 million in 1957, before falling back to 3.1 million in 1973. This rise in births in the 1950s led to a rise in the number of young workers in the 1970s. Younger workers have higher unemployment rates, however, so when the baby-boom generation entered the labor force,

they increased the average level of unemployment. Then as the baby-boom workers aged, the average age of the labor force increased, lowering the average unemployment rate in the 1990s.

This demographic change, however, cannot fully explain the trends in unemployment because similar trends are apparent for fixed demographic groups. For example, for men between the ages of 25 and 54, the average unemployment rate rose from 3.0 percent in the 1960s to 6.1 percent in the 1980s. Thus, although demographic changes may be part of the story of rising unemployment over this period, there must be other explanations of the long-term trend as well.

Sectoral Shifts A second explanation is based on changes in the prevalence of sectoral shifts. The greater the amount of sectoral reallocation, the greater the rate of job separation and the higher the level of frictional unemployment. One source of sectoral shifts during the 1970s and early 1980s was the great volatility in oil prices caused by OPEC, the international oil cartel. These large changes in oil prices may have required reallocating labor between more-energy-intensive and less-energy-intensive sectors. If so, oil-price volatility may have increased unemployment during this period. Although this explanation is hard to evaluate, it is consistent with recent developments: the fall in unemployment during the 1990s coincided with increased stability in oil prices.

Productivity A third explanation for the trends in unemployment emphasizes the link between unemployment and productivity. As Chapter 8 discusses more fully, the 1970s experienced a slowdown in productivity growth, and the 1990s experienced a pickup in productivity growth. These productivity changes roughly coincide with changes in unemployment. Perhaps slowing productivity during the 1970s raised the natural rate of unemployment, and accelerating productivity during the 1990s growth lowered it.

Why such an effect would occur, however, is not obvious. In standard theories of the labor market, higher productivity means greater labor demand and thus higher real wages, but unemployment is unchanged. This prediction is consistent with the long-term data, which show consistent upward trends in productivity and real wages but no trend in unemployment. Yet suppose that workers are slow to catch on to news about productivity. When productivity changes, workers may only gradually alter the real wages they ask from their employers, making real wages sluggish in response to labor demand. An acceleration in productivity growth, such as that experienced during the 1990s, will increase labor demand and, with a sluggish real wage, reduce the amount of unemployment.

In the end, the trends in the unemployment rate remain a mystery. The proposed explanations are plausible, but none seems conclusive on its own. Perhaps there is no single answer. The upward drift in the unemployment rate in the 1970s and 1980s and the downward drift in the 1990s may be the result of several unrelated developments.[9]

---

[9] On the role of demographics, see Robert Shimer, "Why Is the U.S. Unemployment Rate So Much Lower?" NBER Macroeconomics Annual 13 (1998). On the role of sectoral shifts, see David M. Lilien, "Sectoral Shifts and Cyclical Unemployment," Journal of Political Economy 90 (August 1982): 777–793. On the role of productivity, see Laurence Ball and Robert Moffitt, "Productivity Growth and the Phillips Curve," NBER Working Paper No. 8421, August 2001.

Transitions Into and Out of the Labor Force

So far we have ignored an important aspect of labor-market dynamics: the movement of individuals into and out of the labor force. Our model of the natural rate of unemployment assumes that the size of the labor force is fixed. In this case, the sole reason for unemployment is job separation, and the sole reason for leaving unemployment is job finding.

In fact, changes in the labor force are important. About one-third of the unemployed have only recently entered the labor force. Some of these entrants are young workers still looking for their first jobs; others have worked before but temporarily left the labor force. In addition, not all unemployment ends with job finding: almost half of all spells of unemployment end in the unemployed person's withdrawal from the labor market.

Individuals entering and leaving the labor force make unemployment statistics more difficult to interpret. On the one hand, some individuals calling themselves unemployed may not be seriously looking for jobs and perhaps should best be viewed as out of the labor force. Their unemployment may not represent a social problem. On the other hand, some individuals may want jobs but, after unsuccessful searches, have given up looking. These discouraged workers are counted as being out of the labor force and do not show up in unemployment statistics. Even though their joblessness is unmeasured, it may nonetheless be a social problem.

Because of these and many other issues that complicate the interpretation of the unemployment data, the Bureau of Labor Statistics calculates several measures of labor underutilization. Table 6-3 gives the definitions and their values as

table 6-3

Alternative Measures of Labor Underutilization

Definition		Percentage in March 2001
U-1	Persons unemployed 15 weeks or longer, as a percentage of the civilian labor force (includes only very long term unemployed)	1.2 %
U-2	Job losers and persons who have completed temporary jobs, as a percentage of the civilian labor force (excludes job leavers)	2.4
U-3	Total unemployed, as a percentage of the civilian labor force (official unemployment rate)	4.6
U-4	Total unemployed, plus discouraged workers, as a percentage of the civilian labor force plus discouraged workers	4.8
U-5	Total unemployed plus all marginally attached workers, as a percentage of the civilian labor force plus all marginally attached workers	5.3
U-6	Total unemployed, plus all marginally attached workers, plus total employed part time for economic reasons, as a percentage of the civilian labor force plus all marginally attached workers	7.6

Note: Marginally attached workers are persons who currently are neither working nor looking for work but indicate that they want and are available for a job and have looked for work sometime in the recent past. Discouraged workers, a subset of the marginally attached, have given a job-market related reason for not currently looking for a job. Persons employed part time for economic reasons are those who want and are available for full-time work but have had to settle for a part-time schedule.
Source: U.S. Department of Labor.

of March 2001. The measures range from 1.2 to 7.6 percent, depending on the characteristics one uses to classify a worker as not fully employed.

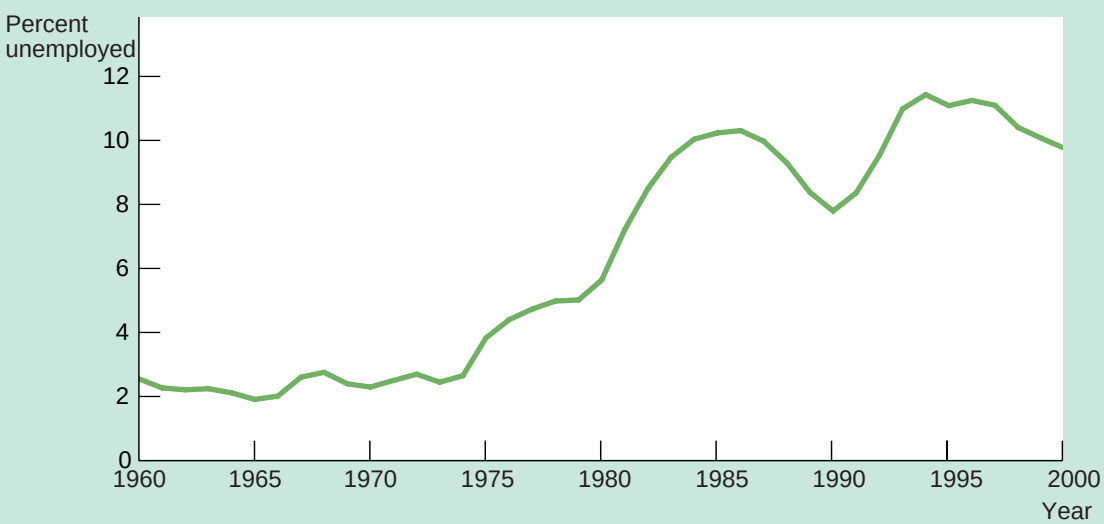
The Rise in European Unemployment

Although our discussion has focused largely on the United States, one puzzling question about unemployment concerns recent developments in Europe. Figure 6-4 shows the rate of unemployment in the countries that make up the European Community—Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and the United Kingdom. As you can see, the rate of unemployment in these countries has risen substantially: it averaged less than 3 percent in the 1960s and more than 10 percent in recent years.

What is the cause of rising European unemployment? No one knows for sure, but there is a leading theory. Many economists believe that the problem can be traced to generous benefits for unemployed workers, coupled with a technologically driven fall in the demand for unskilled workers relative to skilled workers.

There is no question that most European countries have generous programs for those without jobs. These programs go by various names: social insurance, the welfare state, or simply "the dole". Many countries allow the unemployed to

figure 6-4



Unemployment in the European Community This figure shows the unemployment rate in the 15 countries that make up the European Community. The figure shows that the European unemployment rate has risen substantially since 1980.

Source: OECD.

collect benefits indefinitely, rather than for only a short period of time as in the United States. Studies have shown that countries with more generous benefits tend to have higher rates of unemployment. In some sense, those living on the dole are really out of the labor force: given the employment opportunities available, taking a job is less attractive than remaining without work. Yet these people are often counted as unemployed in government statistics.

There is also no question that the demand for unskilled workers has fallen relative to the demand for skilled workers. This change in demand is probably attributable to changes in technology: computers, for example, increase the demand for workers who can use them and reduce the demand for those who cannot. In the United States, this change in demand has been reflected in wages rather than unemployment: over the past two decades, the wages of unskilled workers have fallen substantially relative to the wages of skilled workers. In Europe, however, the welfare state provides unskilled workers with an alternative to working for low wages. As the wages of unskilled workers fall, more workers view the dole as their best available option. The result is higher unemployment.

This diagnosis of high European unemployment does not suggest an easy remedy. Reducing the magnitude of government benefits for the unemployed would encourage workers to get off the dole and accept low-wage jobs. But it would also exacerbate economic inequality—the very problem that welfare-state policies were designed to address.[10]

CASE STUDY

The Secrets to Happiness

Why are some people more satisfied with their lives than others? This is a deep and difficult question, most often left to philosophers and psychologists. But part of the answer is macroeconomic. Recent research has shown that people are happier when they are living in a country with low inflation and low unemployment.

From 1975 to 1991, a survey called the Euro-Barometer Survey Series asked 264,710 people living in 12 European countries about their happiness and overall satisfaction with life. One question asked, "On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?" To see what determines happiness, the answers to this question were correlated with individual and macroeconomic variables. Other things being equal, people are more satisfied with their lives if they are rich, educated, married, in school, self-employed, retired, female, and young or old (as opposed to middle-aged).

[10] For more discussion of these issues, see Paul Krugman, "Past and Prospective Causes of High Unemployment," in Reducing Unemployment: Current Issues and Policy Options, Federal Reserve Bank of Kansas City, August 1994.

They are less satisfied if they are unemployed, divorced, or living with adolescent children. (Some of these correlations may reflect the effects, rather than causes, of happiness: for example, a happy person may find it easier than an unhappy one to keep a job and a spouse.)

Beyond these individual characteristics, the economy's overall rates of unemployment and inflation also play a significant role in explaining reported happiness. An increase in the unemployment rate of 4 percentage points is large enough to move 11 percent of the population down from one life-satisfaction category to another. The overall unemployment rate reduces satisfaction even after controlling for an individual's employment status. That is, the employed in a high-unemployment nation are less happy than their counterparts in a low-unemployment nation, perhaps because they are more worried about job loss or perhaps out of sympathy with their fellow citizens.

High inflation is also associated with lower life satisfaction, although the effect is not as large. A 1.7-percentage-point increase in inflation reduces happiness by about as much as a 1-percentage-point increase in unemployment. The commonly cited "misery index," which is the sum of the inflation and unemployment rates, apparently gives too much weight to inflation relative to unemployment.[11]

6-5 Conclusion

Unemployment represents wasted resources. Unemployed workers have the potential to contribute to national income but are not doing so. Those searching for jobs to suit their skills are happy when the search is over, and those waiting for jobs in firms that pay above-equilibrium wages are happy when positions open up.

Unfortunately, neither frictional unemployment nor structural unemployment can be easily reduced. The government cannot make job search instantaneous, and it cannot easily bring wages closer to equilibrium levels. Zero unemployment is not a plausible goal for free-market economies.

Yet public policy is not powerless in the fight to reduce unemployment. Job-training programs, the unemployment-insurance system, the minimum wage, and the laws governing collective bargaining are often topics of political debate. The policies we choose are likely to have important effects on the economy's natural rate of unemployment.

---

[11] Rafael Di Tella, Robert J. MacCulloch, and Andrew J. Oswald, "Preferences Over Inflation and Unemployment: Evidence From Surveys of Happiness," *American Economic Review* 91 (March 2001): 335–341.

Summary

1. The natural rate of unemployment is the steady-state rate of unemployment. It depends on the rate of job separation and the rate of job finding.

2. Because it takes time for workers to search for the job that best suits their individual skills and tastes, some frictional unemployment is inevitable. Various government policies, such as unemployment insurance, alter the amount of frictional unemployment.

3. Structural unemployment results when the real wage remains above the level that equilibrates labor supply and labor demand. Minimum-wage legislation is one cause of wage rigidity. Unions and the threat of unionization are another. Finally, efficiency-wage theories suggest that, for various reasons, firms may find it profitable to keep wages high despite an excess supply of labor.

4. Whether we conclude that most unemployment is short term or long term depends on how we look at the data. Most spells of unemployment are short. Yet most weeks of unemployment are attributable to the small number of long-term unemployed.

5. The unemployment rates among demographic groups differ substantially. In particular, the unemployment rates for younger workers are much higher than for older workers. This results from a difference in the rate of job separation rather than from a difference in the rate of job finding.

6. The natural rate of unemployment in the United States has exhibited long-term trends. In particular, it rose from the 1950s to the 1970s and then started drifting downward again in the 1990s. Various explanations have been proposed, including the changing demographic composition of the labor force, changes in the prevalence of sectoral shifts, and changes in the rate of productivity growth.

7. Individuals who have recently entered the labor force, including both new entrants and reentrants, make up about one-third of the unemployed. Transitions into and out of the labor force make unemployment statistics more difficult to interpret.

KEY CONCEPTS

Natural rate of unemployment
Frictional unemployment
Sectoral shift
Unemployment insurance
Wage rigidity
Structural unemployment
Insiders versus outsiders
Efficiency wages
Discouraged workers

QUESTIONS FOR REVIEW

1. What determines the natural rate of unemployment?
2. Describe the difference between frictional unemployment and structual unemployment.
3. Give three explanations why the real wage may remain above the level that equilibrates labor supply and labor demand.
4. Is most unemployment long term or short term? Explain your answer.
5. How do economists explain the high natural rate of unemployment in the 1970s and 1980s? How do they explain the fall in the natural rate in the 1990s?

PROBLEMS AND APPLICATIONS

1. Answer the following questions about your own experience in the labor force:
   a. When you or one of your friends is looking for a part-time job, how many weeks does it typically take? After you find a job, how many weeks does it typically last?
   b. From your estimates, calculate (in a rate per week) your rate of job finding f and your rate of job separation s. (Hint: If f is the rate of job finding, then the average spell of unemployment is $1/f$.)
   c. What is the natural rate of unemployment for the population you represent?
2. In this chapter we saw that the steady-state rate of unemployment is $U/L = s/(s + f)$. Suppose that the unemployment rate does not begin at this level. Show that unemployment will evolve over time and reach this steady state. (Hint: Express the change in the number of unemployed as a function of s, f, and U. Then show that if unemployment is above the natural rate, unemployment falls, and if unemployment is below the natural rate, unemployment rises.)
3. The residents of a certain dormitory have collected the following data: People who live in the dorm can be classified as either involved in a relationship or uninvolved. Among involved people, 10 percent experience a breakup of their relationship every month. Among uninvolved people, 5 percent will enter into a relationship every month. What is the steady-state fraction of residents who are uninvolved?
4. Suppose that Congress passes legislation making it more difficult for firms to fire workers. (An example is a law requiring severance pay for fired workers.) If this legislation reduces the rate of job separation without affecting the rate of job finding, how would the natural rate of unemployment change? Do you think that it is plausible that the legislation would not affect the rate of job finding? Why or why not?
5. Consider an economy with the following Cobb-Douglas production function:

$$Y = K^{1/3}L^{2/3}.$$

   The economy has 1,000 units of capital and a labor force of 1,000 workers.
   a. Derive the equation describing labor demand in this economy as a function of the real wage and the capital stock. (Hint: Review the appendix to Chapter 3.)
   b. If the real wage can adjust to equilibrate labor supply and labor demand, what is the real wage? In this equilibrium, what are employment, output, and the total amount earned by workers?
   c. Now suppose that Congress, concerned about the welfare of the working class, passes a law requiring firms to pay workers a real wage of 1 unit of output. How does this wage compare to the equilibrium wage?
   d. Congress cannot dictate how many workers firms hire at the mandated wage. Given this fact, what are the effects of this law? Specifi

cally, what happens to employment, output, and the total amount earned by workers?

e. Will Congress succeed in its goal of helping the working class? Explain.

f. Do you think that this analysis provides a good way of thinking about a minimum-wage law? Why or why not?

6. Suppose that a country experiences a reduction in productivity—that is, an adverse shock to the production function.

a. What happens to the labor demand curve?

b. How would this change in productivity affect the labor market—that is, employment, unemployment, and real wages—if the labor market were always in equilibrium?

c. How would this change in productivity affect the labor market if unions prevented real wages from falling?

7. In any city at any time, some of the stock of usable office space is vacant. This vacant office space is unemployed capital. How would you explain this phenomenon? Is it a social problem?

part III

Growth Theory: The Economy in the Very Long Run

CHAPTER 7 SEVEN

Economic Growth I

> The question of growth is nothing new but a new disguise for an age-old issue, one which has always intrigued and preoccupied economics: the present versus the future.
>
> — James Tobin

If you have ever spoken with your grandparents about what their lives were like when they were young, most likely you learned an important lesson about economics: material standards of living have improved substantially over time for most families in most countries. This advance comes from rising incomes, which have allowed people to consume greater quantities of goods and services.

To measure economic growth, economists use data on gross domestic product, which measures the total income of everyone in the economy. The real GDP of the United States today is more than three times its 1950 level, and real GDP per person is more than twice its 1950 level. In any given year, we can also observe large differences in the standard of living among countries. Table 7-1 shows income per person in 1999 of the world's 12 most populous countries. The United States tops the list with an income of $31,910 per person. Nigeria has an income per person of only $770, less than 3 percent of this figure for the United States.

Our goal in this part of the book is to understand what causes these differences in income over time and across countries. In Chapter 3 we identified the factors of production—capital and labor—and the production technology as the sources of the economy's output and, thus, of its total income. Differences in income, then, must come from differences in capital, labor, and technology.

Our primary task is to develop a theory of economic growth called the Solow growth model. Our analysis in Chapter 3 enabled us to describe how the economy produces and uses its output at one point in time. The analysis was static—a snapshot of the economy. To explain why our national income grows, and why some economies grow faster than others, we must broaden our analysis so that it describes changes in the economy over time. By developing such a model, we make our analysis dynamic—more like a movie than a photograph. The Solow growth model shows how saving, population growth, and

table 7-1

International Differences in the Standard of Living: 1999

Country	Income per Person (in U.S. dollars)	Country	Income per Person (in U.S. dollars)
United States	$31,910	China	3,550
Japan	25,170	Indonesia	2,660
Germany	23,510	India	2,230
Mexico	8,070	Pakistan	1,860
Russia	6,990	Bangladesh	1,530
Brazil	6,840	Nigeria	770

Source: World Bank.

technological progress affect the level of an economy's output and its growth over time. In this chapter we analyze the roles of saving and population growth. In the next chapter we introduce technological progress.[1]

7-1 The Accumulation of Capital

The Solow growth model is designed to show how growth in the capital stock, growth in the labor force, and advances in technology interact in an economy, and how they affect a nation's total output of goods and services. We build this model in steps. Our first step is to examine how the supply and demand for goods determine the accumulation of capital. In this first step, we assume that the labor force and technology are fixed. We then relax these assumptions by introducing changes in the labor force later in this chapter and by introducing changes in technology in the next.

The Supply and Demand for Goods

The supply and demand for goods played a central role in our static model of the closed economy in Chapter 3. The same is true for the Solow model. By considering the supply and demand for goods, we can see what determines how much output is produced at any given time and how this output is allocated among alternative uses.

[1] The Solow growth model is named after economist Robert Solow and was developed in the 1950s and 1960s. In 1987 Solow won the Nobel Prize in economics for his work in economic growth. The model was introduced in Robert M. Solow, "A Contribution to the Theory of Economic Growth," Quarterly Journal of Economics (February 1956): 65–94.

The Supply of Goods and the Production Function The supply of goods in the Solow model is based on the now-familiar production function, which states that output depends on the capital stock and the labor force:

$$Y = F(K, L).$$

The Solow growth model assumes that the production function has constant returns to scale. This assumption is often considered realistic, and as we will see shortly, it helps simplify the analysis. Recall that a production function has constant returns to scale if

$$zY = F(zK, zL)$$

for any positive number z. That is, if we multiply both capital and labor by z, we also multiply the amount of output by z.

Production functions with constant returns to scale allow us to analyze all quantities in the economy relative to the size of the labor force. To see that this is true, set $z = 1/L$ in the preceding equation to obtain

$$Y/L = F(K/L, 1).$$

This equation shows that the amount of output per worker Y/L is a function of the amount of capital per worker K/L. (The number "1" is, of course, constant and thus can be ignored.) The assumption of constant returns to scale implies that the size of the economy—as measured by the number of workers—does not affect the relationship between output per worker and capital per worker.

Because the size of the economy does not matter, it will prove convenient to denote all quantities in per-worker terms. We designate these with lowercase letters, so $y = Y/L$ is output per worker, and $k = K/L$ is capital per worker. We can then write the production function as

$$y = f(k),$$

where we define $f(k) = F(k,1)$. Figure 7-1 illustrates this production function.

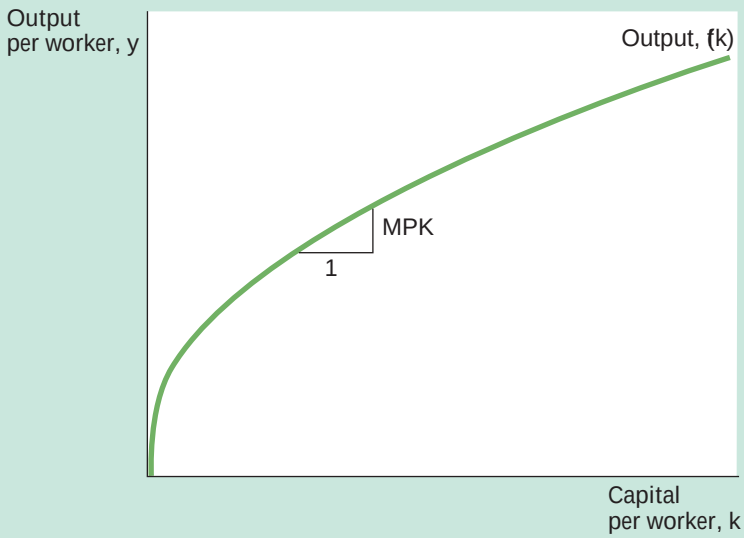
The slope of this production function shows how much extra output a worker produces when given an extra unit of capital. This amount is the marginal product of capital, MPK. Mathematically, we write

$$MPK = f(k + 1) - f(k).$$

Note that in Figure 7-1, as the amount of capital increases, the production function becomes flatter, indicating that the production function exhibits diminishing marginal product of capital. When k is low, the average worker has only a little capital to work with, so an extra unit of capital is very useful and produces a lot of additional output. When k is high, the average worker has a lot of capital, so an extra unit increases production only slightly.

The Demand for Goods and the Consumption Function The demand for goods in the Solow model comes from consumption and investment. In other

figure 7-1



The Production Function The production function shows how the amount of capital per worker k determines the amount of output per worker $y = f(k)$. The slope of the production function is the marginal product of capital: if k increases by 1 unit, y increases by MPK units. The production function becomes flatter as k increases, indicating diminishing marginal product of capital.

words, output per worker y is divided between consumption per worker c and investment per worker i:

$$y = c + i.$$

This equation is the per-worker version of the national income accounts identity for an economy. Notice that it omits government purchases (which for present purposes we can ignore) and net exports (because we are assuming a closed economy).

The Solow model assumes that each year people save a fraction s of their income and consume a fraction $(1 - s)$. We can express this idea with a consumption function with the simple form

$$c = (1 - s)y,$$

where s, the saving rate, is a number between zero and one. Keep in mind that various government policies can potentially influence a nation's saving rate, so one of our goals is to find what saving rate is desirable. For now, however, we just take the saving rate as given.

To see what this consumption function implies for investment, substitute $(1 - s)y$ for c in the national income accounts identity:

$$y = (1 - s)y + i.$$

Rearrange the terms to obtain

$$i = sy.$$

This equation shows that investment equals saving, as it was in Chapter 3. Thus, the rate of saving s is also the fraction of output devoted to investment.

We have now introduced the two main ingredients of the Solow model—the production function and the consumption function—which describe the economy at any moment in time. For any given capital stock, the production function $y = f(k)$ determines how much output the economy produces, and the saving rate s determines the allocation of that output between consumption and investment.

Growth in the Capital Stock and the Steady State

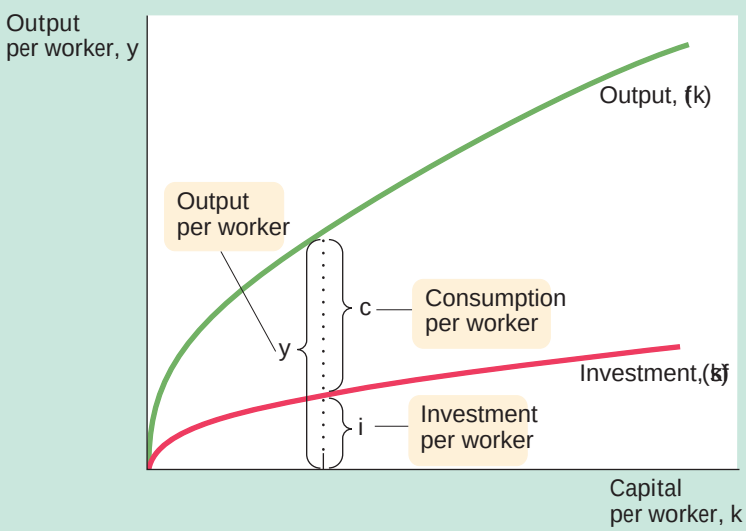
At any moment, the capital stock is a key determinant of the economy's output, but the capital stock can change over time, and those changes can lead to economic growth. In particular, two forces influence the capital stock: investment and depreciation. Investment refers to the expenditure on new plant and equipment, and it causes the capital stock to rise. Depreciation refers to the wearing out of old capital, and it causes the capital stock to fall. Let's consider each of these in turn.

As we have already noted, investment per worker equals $i = sy$. By substituting the production function for y, we can express investment per worker as a function of the capital stock per worker:

$$i = sf(k).$$

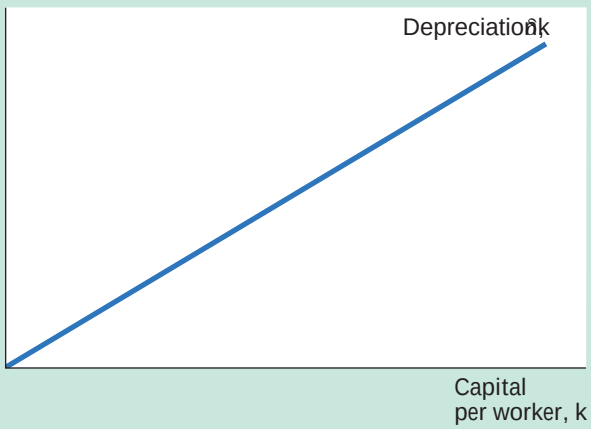
This equation relates the existing stock of capital to the accumulation of new capital i. Figure 7-2 shows this relationship. The figure illustrates how, for any value of k, the amount of output is determined by the production function $f(k)$, and the allocation of that output between consumption and saving is determined by the saving rate s.

figure 7-2



Output, Consumption, and Investment The saving rate s determines the allocation of output between consumption and investment. For any level of capital k, output is $f(k)$, investment is $sf(k)$, and consumption is $f(k) - sf(k)$.

figure 7-3



Depreciation A constant fraction δ of the capital stock wears out every year. Depreciation is therefore proportional to the capital stock.

To incorporate depreciation into the model, we assume that a certain fraction δ of the capital stock wears out each year. Here δ (the lowercase Greek letter delta) is called the *depreciation rate*. For example, if capital lasts an average of 25 years, then the depreciation rate is 4 percent per year ($\delta = 0.04$). The amount of capital that depreciates each year is δk. Figure 7-3 shows how the amount of depreciation depends on the capital stock.

We can express the impact of investment and depreciation on the capital stock with this equation:

$$\begin{array}{ccccc} \text{Change in Capital Stock} & = & \text{Investment} & - & \text{Depreciation} \\ \Delta k & = & i & - & \delta k, \end{array}$$

where Δk is the change in the capital stock between one year and the next. Because investment equals $sf(k)$, we can write this as

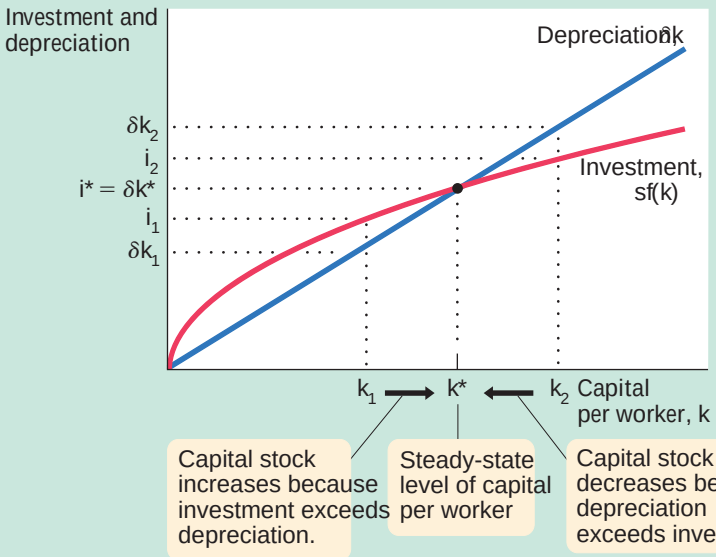
$$\Delta k = sf(k) - \delta k.$$

Figure 7-4 graphs the terms of this equation—investment and depreciation—for different levels of the capital stock k. The higher the capital stock, the greater the amounts of output and investment. Yet the higher the capital stock, the greater also the amount of depreciation.

As Figure 7-4 shows, there is a single capital stock at which the amount of investment equals the amount of depreciation. If the economy finds itself at this level of the capital stock, the capital stock will not change because the two forces acting on it—investment and depreciation—just balance. That is, at k^*, $\Delta k = 0$, so the capital stock k and output $f(k)$ are steady over time (rather than growing or shrinking). We therefore call k^* the *steady-state* level of capital.

The steady state is significant for two reasons. As we have just seen, an economy at the steady state will stay there. In addition, and just as important, an economy not at the steady state will go there. That is, regardless of the level of capital

figure 7-4



Investment, Depreciation, and the Steady State The steady-state level of capital k^* is the level at which investment equals depreciation, indicating that the amount of capital will not change over time. Below k^*, investment exceeds depreciation, so the capital stock grows. Above k^*, investment is less than depreciation, so the capital stock shrinks.

with which the economy begins, it ends up with the steady-state level of capital. In this sense, the steady state represents the long-run equilibrium of the economy.

To see why an economy always ends up at the steady state, suppose that the economy starts with less than the steady-state level of capital, such as level k_1 in Figure 7-4. In this case, the level of investment exceeds the amount of depreciation. Over time, the capital stock will rise and will continue to rise along with output $f(k)$—until it approaches the steady state k^*.

Similarly, suppose that the economy starts with more than the steady-state level of capital, such as level k_2. In this case, investment is less than depreciation: capital is wearing out faster than it is being replaced. The capital stock will fall, again approaching the steady-state level. Once the capital stock reaches the steady state, investment equals depreciation, and there is no pressure for the capital stock to either increase or decrease.

Approaching the Steady State: A Numerical Example

Let's use a numerical example to see how the Solow model works and how the economy approaches the steady state. For this example, we assume that the production function is[2]

$$Y = K^{1/2}L^{1/2}.$$

[2] If you read the appendix to Chapter 3, you will recognize this as the Cobb–Douglas production function with the parameter α equal to 1/2.

To derive the per-worker production function $f(k)$, divide both sides of the production function by the labor force L:

$$\frac{Y}{L} = \frac{K^{1/2}L^{1/2}}{L}.$$

Rearrange to obtain

$$\frac{Y}{L} = \left(\frac{K}{L}\right)^{1/2}.$$

Because $y = Y/L$ and $k = K/L$, this becomes

$$y = k^{1/2}.$$

This equation can also be written as

$$y = \sqrt{k}.$$

This form of the production function states that output per worker is equal to the square root of the amount of capital per worker.

To complete the example, let's assume that 30 percent of output is saved ($s = 0.3$), that 10 percent of the capital stock depreciates every year ($\delta = 0.1$), and that the economy starts off with 4 units of capital per worker ($k = 4$). Given these numbers, we can now examine what happens to this economy over time.

We begin by looking at the production and allocation of output in the first year. According to the production function, the 4 units of capital per worker produce 2 units of output per worker. Because 30 percent of output is saved and invested and 70 percent is consumed, $i = 0.6$ and $c = 1.4$. Also, because 10 percent of the capital stock depreciates, $\delta k = 0.4$. With investment of 0.6 and depreciation of 0.4, the change in the capital stock is $\Delta k = 0.2$. The second year begins with 4.2 units of capital per worker.

Table 7-2 shows how the economy progresses year by year. Every year, new capital is added and output grows. Over many years, the economy approaches a steady state with 9 units of capital per worker. In this steady state, investment of 0.9 exactly offsets depreciation of 0.9, so that the capital stock and output are no longer growing.

Following the progress of the economy for many years is one way to find the steady-state capital stock, but there is another way that requires fewer calculations. Recall that

$$\Delta k = sf(k) - \delta k.$$

This equation shows how k evolves over time. Because the steady state is (by definition) the value of k at which $\Delta k = 0$, we know that

$$0 = sf(k^*) - \delta k^*,$$

or, equivalently,

$$\frac{k^*}{f(k^*)} = \frac{s}{\delta}.$$

table 7-2

Approaching the Steady State: A Numerical Example

Assumptions: $y = \sqrt{k}$; $s = 0.3$; $\delta = 0.1$; initial $k = 4.0$

Year	k	y	c	i	δk	Δk
1	4.000	2.000	1.400	0.600	0.400	0.200
2	4.200	2.049	1.435	0.615	0.420	0.195
3	4.395	2.096	1.467	0.629	0.440	0.189
4	4.584	2.141	1.499	0.642	0.458	0.184
5	4.768	2.184	1.529	0.655	0.477	0.178
.						
.						
.						
10	5.602	2.367	1.657	0.710	0.560	0.150
.						
.						
.						
25	7.321	2.706	1.894	0.812	0.732	0.080
.						
.						
.						
100	8.962	2.994	2.096	0.898	0.896	0.002
.						
.						
.						
∞	9.000	3.000	2.100	0.900	0.900	0.000

This equation provides a way of finding the steady-state level of capital per worker, k^*. Substituting in the numbers and production function from our example, we obtain

$$\frac{k^*}{\sqrt{k^*}} = \frac{0.3}{0.1}.$$

Now square both sides of this equation to find

$$k^* = 9.$$

The steady-state capital stock is 9 units per worker. This result confirms the calculation of the steady state in Table 7-2.

CASE STUDY

The Miracle of Japanese and German Growth

Japan and Germany are two success stories of economic growth. Although today they are economic superpowers, in 1945 the economies of both countries were in shambles. World War II had destroyed much of their capital stocks. In the

decades after the war, however, these two countries experienced some of the most rapid growth rates on record. Between 1948 and 1972, output per person grew at 8.2 percent per year in Japan and 5.7 percent per year in Germany, compared to only 2.2 percent per year in the United States.

Are the postwar experiences of Japan and Germany so surprising from the standpoint of the Solow growth model? Consider an economy in steady state. Now suppose that a war destroys some of the capital stock. (That is, suppose the capital stock drops from k^* to k_1 in Figure 7-4.) Not surprisingly, the level of output immediately falls. But if the saving rate—the fraction of output devoted to saving and investment—is unchanged, the economy will then experience a period of high growth. Output grows because, at the lower capital stock, more capital is added by investment than is removed by depreciation. This high growth continues until the economy approaches its former steady state. Hence, although destroying part of the capital stock immediately reduces output, it is followed by higher-than-normal growth. The "miracle" of rapid growth in Japan and Germany, as it is often described in the business press, is what the Solow model predicts for countries in which war has greatly reduced the capital stock.

How Saving Affects Growth

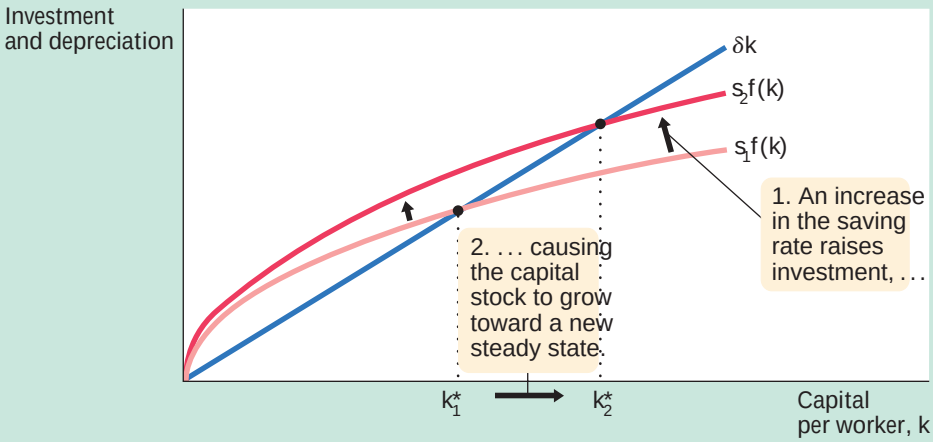
The explanation of Japanese and German growth after World War II is not quite as simple as suggested in the preceding case study. Another relevant fact is that both Japan and Germany save and invest a higher fraction of their output than the United States. To understand more fully the international differences in economic performance, we must consider the effects of different saving rates.

Consider what happens to an economy when its saving rate increases. Figure 7-5 shows such a change. The economy is assumed to begin in a steady state with saving rate s_1 and capital stock k_1^*. When the saving rate increases from s_1 to s_2, the $sf(k)$ curve shifts upward. At the initial saving rate, s_1, and the initial capital stock k_1^*, the amount of investment just offsets the amount of depreciation. Immediately after the saving rate rises, investment is higher, but the capital stock and depreciation are unchanged. Therefore, investment exceeds depreciation. The capital stock will gradually rise until the economy reaches the new steady state k_2^*, which has a higher capital stock and a higher level of output than the old steady state.

The Solow model shows that the saving rate is a key determinant of the steady-state capital stock. If the saving rate is high, the economy will have a large capital stock and a high level of output. If the saving rate is low, the economy will have a small capital stock and a low level of output. This conclusion sheds light on many discussions of fiscal policy. As we saw in Chapter 3, a government budget deficit can reduce national saving and crowd out investment. Now we can see that the long-run consequences of a reduced saving rate are a lower capital stock and lower national income. This is why many economists are critical of persistent budget deficits.

What does the Solow model say about the relationship between saving and economic growth? Higher saving leads to faster growth in the Solow model, but

figure 7-5



An Increase in the Saving Rate An increase in the saving rate implies that the amount of investment for any given capital stock is higher. It therefore shifts the saving function upward. At the initial steady state k_1^*, investment now exceeds depreciation. The capital stock rises until the economy reaches a new steady state k_2^* with more capital and output.

only temporarily. An increase in the rate of saving raises growth only until the economy reaches the new steady state. If the economy maintains a high saving rate, it will maintain a large capital stock and a high level of output, but it will not maintain a high rate of growth forever.

Now that we understand how saving affects growth, we can more fully explain the impressive economic performances of Germany and Japan after World War II. Not only were their initial capital stocks low because of the war, but their steady-state capital stocks were high because of their high saving rates. Both of these facts help explain the rapid growth of these two countries in the 1950s and 1960s.

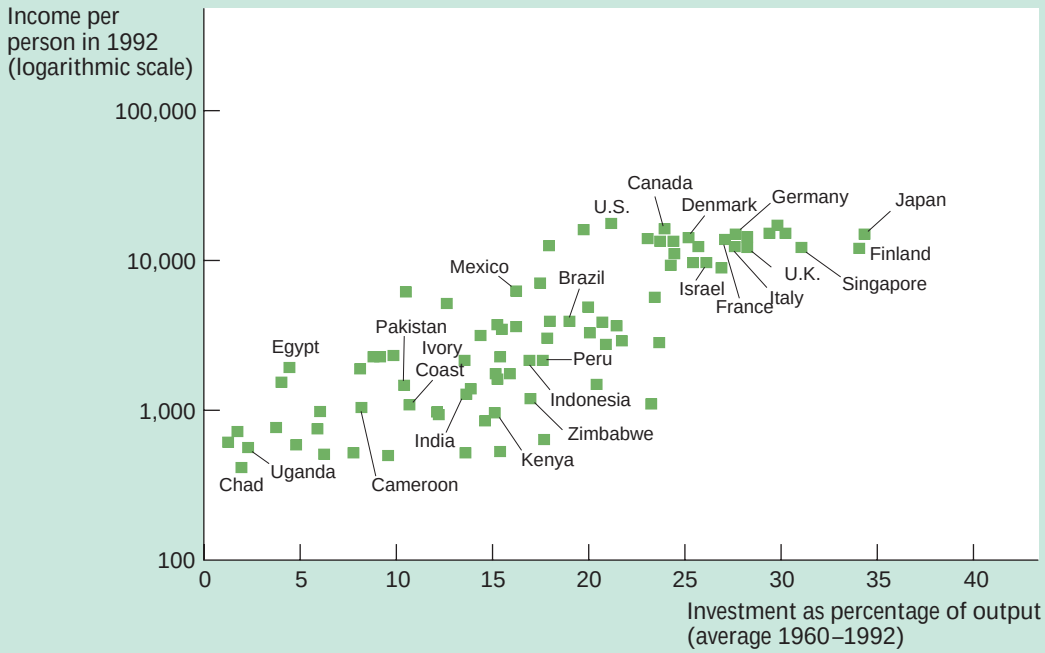
CASE STUDY

Saving and Investment Around the World

We started this chapter with an important question: Why are some countries so rich while others are mired in poverty? Our analysis has taken us a step closer to the answer. According to the Solow model, if a nation devotes a large fraction of its income to saving and investment, it will have a high steady-state capital stock and a high level of income. If a nation saves and invests only a small fraction of its income, its steady-state capital and income will be low.

Let's now look at some data to see if this theoretical result in fact helps explain the large international variation in standards of living. Figure 7-6 is a scatterplot of data from 84 countries. The figure includes most of the world's economies. It excludes major oil-producing countries and countries that were communist during much of this period, because their experiences are explained by their special

figure 7-6



International Evidence on Investment Rates and Income per Person This scatterplot shows the experience of 84 countries, each represented by a single point. The horizontal axis shows the country's rate of investment, and the vertical axis shows the country's income per person. High investment is associated with high income per person, as the Solow model predicts.

Source: Robert Summers and Alan Heston, Supplement (Mark 5.6) to "The Penn World Table (Mark 5): An Expanded Set of International Comparisons 1950–1988," Quarterly Journal of Economics (May 1991): 327–368.

circumstances.) The data show a positive relationship between the fraction of output devoted to investment and the level of income per person. That is, countries with high rates of investment, such as the United States and Japan, usually have high incomes, whereas countries with low rates of investment, such as Uganda and Chad, have low incomes. Thus, the data are consistent with the Solow model's prediction that the investment rate is a key determinant of whether a country is rich or poor.

The strong correlation shown in this figure is an important fact, but it raises as many questions as it resolves. One might naturally ask, why do rates of saving and investment vary so much from country to country? There are many potential answers, such as tax policy, retirement patterns, the development of financial markets, and cultural differences. In addition, political stability may play a role: not surprisingly, rates of saving and investment tend to be low in countries with frequent wars, revolutions, and coups. Saving and investment also tend to be low in countries with poor political institutions, as measured by estimates of corruption. A final interpretation of the evidence in Figure 7-6 is reverse causation:

perhaps high levels of income somehow foster high rates of saving and investment. Unfortunately, there is no consensus among economists about which of the many possible explanations is most important.

The association between investment rates and income per person is strong, and it is an important clue as to why some countries are rich and others poor, but it is not the whole story. The correlation between these two variables is far from perfect. Mexico and Zimbabwe, for instance, have had similar investment rates, but income per person is more than three times higher in Mexico. There must be other determinants of living standards beyond saving and investment. We therefore return to the international differences later in the chapter to see what other variables enter the picture.

7-2 The Golden Rule Level of Capital

So far, we have used the Solow model to examine how an economy's rate of saving and investment determines its steady-state levels of capital and income. This analysis might lead you to think that higher saving is always a good thing, for it always leads to greater income. Yet suppose a nation had a saving rate of 100 percent. That would lead to the largest possible capital stock and the largest possible income. But if all of this income is saved and none is ever consumed, what good is it?

This section uses the Solow model to discuss what amount of capital accumulation is optimal from the standpoint of economic well-being. In the next chapter, we discuss how government policies influence a nation's saving rate. But first, in this section, we present the theory behind these policy decisions.

Comparing Steady States

To keep our analysis simple, let's assume that a policymaker can set the economy's saving rate at any level. By setting the saving rate, the policymaker determines the economy's steady state. What steady state should the policymaker choose?

When choosing a steady state, the policymaker's goal is to maximize the well-being of the individuals who make up the society. Individuals themselves do not care about the amount of capital in the economy, or even the amount of output. They care about the amount of goods and services they can consume. Thus, a benevolent policymaker would want to choose the steady state with the highest level of consumption. The steady-state value of k that maximizes consumption is called the **Golden Rule level of capital** and is denoted k^*_{gold}.[3]

How can we tell whether an economy is at the Golden Rule level? To answer this question, we must first determine steady-state consumption per worker. Then we can see which steady state provides the most consumption.

[3] Edmund Phelps, "The Golden Rule of Accumulation: A Fable for Growthmen," *American Economic Review* 51 (September 1961): 638–643.

To find steady-state consumption per worker, we begin with the national income accounts identity

$$y = c + i$$

and rearrange it as

$$c = y - i.$$

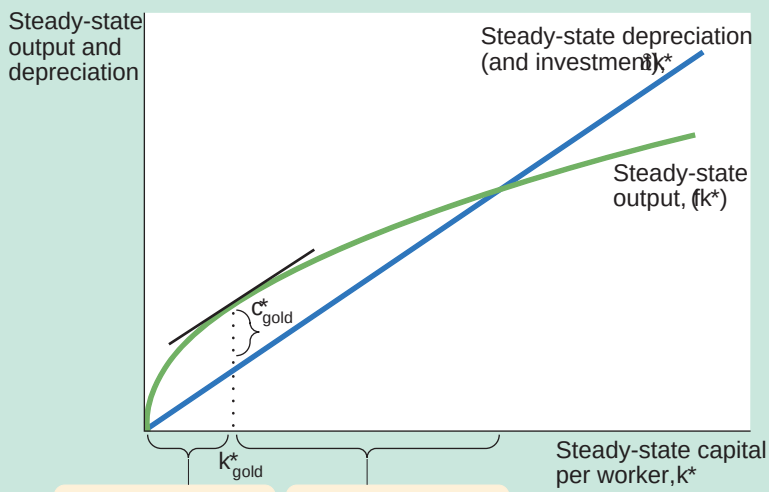
Consumption is simply output minus investment. Because we want to find steady-state consumption, we substitute steady-state values for output and investment. Steady-state output per worker is $f(k^*)$, where k^* is the steady-state capital stock per worker. Furthermore, because the capital stock is not changing in the steady state, investment is equal to depreciation. That is, δk^*. Substituting $f(k^*)$ for y and δk^* for i, we can write steady-state consumption per worker as

$$c^* = f(k^*) - \delta k^*.$$

According to this equation, steady-state consumption is what is left of steady-state output after paying for steady-state depreciation. This equation shows that an increase in steady-state capital has two opposing effects on steady-state consumption. On the one hand, more capital means more output. On the other hand, more capital also means that more output must be used to replace capital that is wearing out.

Figure 7-7 graphs steady-state output and steady-state depreciation as a function of the steady-state capital stock. Steady-state consumption is the gap between

figure 7-7



Steady-State Consumption
The economy's output is used for consumption or investment. In the steady state, investment equals depreciation. Therefore, steady-state consumption is the difference between output $f(k^*)$ and depreciation δk^*. Steady-state consumption is maximized at the Golden Rule steady state. The Golden Rule capital stock is denoted k^*_{gold}, and the Golden Rule consumption is denoted c^*_{gold}.

output and depreciation. This figure shows that there is one level of the capital stock—the Golden Rule level k^*_{gold}—that maximizes consumption.

When comparing steady states, we must keep in mind that higher levels of capital affect both output and depreciation. If the capital stock is below the Golden Rule level, an increase in the capital stock raises output more than depreciation, so that consumption rises. In this case, the production function is steeper than the δk^* line, so the gap between these two curves—which equals consumption—grows as k^* rises. By contrast, if the capital stock is above the Golden Rule level, an increase in the capital stock reduces consumption, since the increase in output is smaller than the increase in depreciation. In this case, the production function is flatter than the δk^* line, so the gap between the curves—consumption—shrinks as k^* rises. At the Golden Rule level of capital, the production function and the δk^* line have the same slope, and consumption is at its greatest level.

We can now derive a simple condition that characterizes the Golden Rule level of capital. Recall that the slope of the production function is the marginal product of capital, MPK. The slope of the δk^* line is δ. Because these two slopes are equal at k^*_{gold}, the Golden Rule is described by the equation

$$MPK = \delta.$$

At the Golden Rule level of capital, the marginal product of capital equals the depreciation rate.

To make the point somewhat differently, suppose that the economy starts at some steady-state capital stock k^* and that the policymaker is considering increasing the capital stock to $k^* + 1$. The amount of extra output from this increase in capital would be $f(k^* + 1) - f(k^*)$, which is the marginal product of capital, MPK. The amount of extra depreciation from having 1 more unit of capital is the depreciation rate, δ. Thus, the net effect of this extra unit of capital on consumption is $MPK - \delta$. If $MPK - \delta > 0$, then increases in capital increase consumption, so k^* must be below the Golden Rule level. If $MPK - \delta < 0$, then increases in capital decrease consumption, so k^* must be above the Golden Rule level. Therefore, the following condition describes the Golden Rule:

$$MPK - \delta = 0.$$

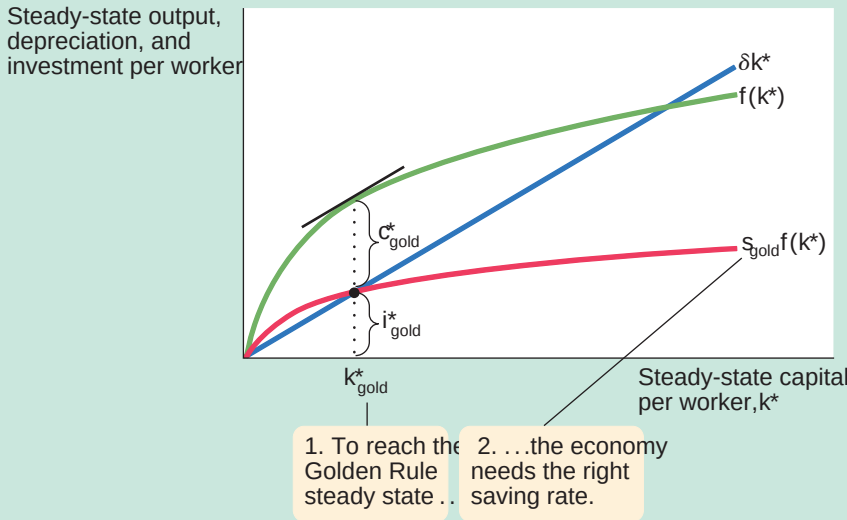
At the Golden Rule level of capital, the marginal product of capital net of depreciation ($MPK - \delta$) equals zero. As we will see, a policymaker can use this condition to find the Golden Rule capital stock for an economy.[4]

Keep in mind that the economy does not automatically gravitate toward the Golden Rule steady state. If we want any particular steady-state capital stock, such as the Golden Rule, we need a particular saving rate to support it. Figure 7-8

---

[4] Mathematical note: Another way to derive the condition for the Golden Rule uses a bit of calculus. Recall that $c^* = f(k^*) - \delta k^*$. To find the k^* that maximizes c^*, differentiate to find $dc^*/dk^* = f'(k^*) - \delta$ and set this derivative equal to zero. Noting that $f'(k^*)$ is the marginal product of capital, we obtain the Golden Rule condition in the text.

figure 7-8



The Saving Rate and the Golden Rule There is only one saving rate that produces the Golden Rule level of capital k^*_{gold}. Any change in the saving rate would shift the $sf(k)$ curve and would move the economy to a steady state with a lower level of consumption.

shows the steady state if the saving rate is set to produce the Golden Rule level of capital. If the saving rate is higher than the one used in this figure, the steady-state capital stock will be too high. If the saving rate is lower, the steady-state capital stock will be too low. In either case, steady-state consumption will be lower than it is at the Golden Rule steady state.

Finding the Golden Rule Steady State: A Numerical Example

Consider the decision of a policymaker choosing a steady state in the following economy. The production function is the same as in our earlier example:

$$y = \sqrt{k}.$$

Output per worker is the square root of capital per worker. Depreciation is again 10 percent of capital. This time, the policymaker chooses the saving rate and thus the economy's steady state.

To see the outcomes available to the policymaker, recall that the following equation holds in the steady state:

$$\frac{k^*}{f(k^*)} = \frac{s}{\delta}.$$

In this economy, this equation becomes

$$\frac{k^*}{\sqrt{k^*}} = \frac{s}{0.1}.$$

Squaring both sides of this equation yields a solution for the steady-state capital stock. We find

$$k^* = 100s^2.$$

Using this result, we can compute the steady-state capital stock for any saving rate.

Table 7-3 presents calculations showing the steady states that result from various saving rates in this economy. We see that higher saving leads to a higher capital stock, which in turn leads to higher output and higher depreciation. Steady-state consumption, the difference between output and depreciation, first rises with higher saving rates and then declines. Consumption is highest when the saving rate is 0.5. Hence, a saving rate of 0.5 produces the Golden Rule steady state.

table 7-3

Finding the Golden Rule Steady State: A Numerical Example

Assumptions: $y = \sqrt{k}$; $\delta = 0.1$

s	k^*	y^*	δk^*	c^*	MPK	$MPK - \delta$
0.0	0.0	0.0	0.0	0.0	∞	∞
0.1	1.0	1.0	0.1	0.9	0.500	0.400
0.2	4.0	2.0	0.4	1.6	0.250	0.150
0.3	9.0	3.0	0.9	2.1	0.167	0.067
0.4	16.0	4.0	1.6	2.4	0.125	0.025
0.5	25.0	5.0	2.5	2.5	0.100	0.000
0.6	36.0	6.0	3.6	2.4	0.083	–0.017
0.7	49.0	7.0	4.9	2.1	0.071	–0.029
0.8	64.0	8.0	6.4	1.6	0.062	–0.038
0.9	81.0	9.0	8.1	0.9	0.056	–0.044
1.0	100.0	10.0	10.0	0.0	0.050	–0.050

Recall that another way to identify the Golden Rule steady state is to find the capital stock at which the net marginal product of capital, $MPK - \delta$, equals zero. For this production function, the marginal product is[5]

$$MPK = \frac{1}{2\sqrt{k}}.$$

[5] Mathematical note: To derive this formula, note that the marginal product of capital is the derivative of the production function with respect to k.

Using this formula, the last two columns of Table 7-3 present the values of MPK and $MPK - \delta$ in the different steady states. Note that the net marginal product of capital is exactly zero when the saving rate is at its Golden Rule value of 0.5. Because of diminishing marginal product, the net marginal product of capital is greater than zero whenever the economy saves less than this amount, and it is less than zero whenever the economy saves more.

This numerical example confirms that the two ways of finding the Golden Rule steady state—looking at steady-state consumption or looking at the marginal product of capital—give the same answer. If we want to know whether an actual economy is currently at, above, or below its Golden Rule capital stock, the second method is usually more convenient, because estimates of the marginal product of capital are easy to come by. By contrast, evaluating an economy with the first method requires estimates of steady-state consumption at many different saving rates; such information is hard to obtain. Thus, when we apply this kind of analysis to the U.S. economy in the next chapter, we will find it useful to examine estimates of the marginal product of capital.

The Transition to the Golden Rule Steady State

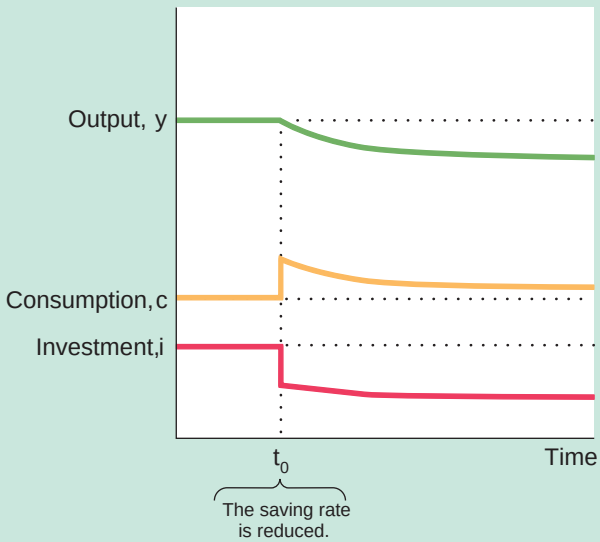
Let's now make our policymaker's problem more realistic. So far, we have been assuming that the policymaker can simply choose the economy's steady state and jump there immediately. In this case, the policymaker would choose the steady state with highest consumption—the Golden Rule steady state. But now suppose that the economy has reached a steady state other than the Golden Rule. What happens to consumption, investment, and capital when the economy makes the transition between steady states? Might the impact of the transition deter the policymaker from trying to achieve the Golden Rule?

We must consider two cases: the economy might begin with more capital than in the Golden Rule steady state, or with less. It turns out that the two cases offer very different problems for policymakers. (As we will see in the next chapter, the second case—too little capital—describes most actual economies, including that of the United States.)

Starting With Too Much Capital We first consider the case in which the economy begins at a steady state with more capital than it would have in the Golden Rule steady state. In this case, the policymaker should pursue policies aimed at reducing the rate of saving in order to reduce the capital stock. Suppose that these policies succeed and that at some point in time t_0—the saving rate falls to the level that will eventually lead to the Golden Rule steady state.

Figure 7-9 shows what happens to output, consumption, and investment when the saving rate falls. The reduction in the saving rate causes an immediate increase in consumption and a decrease in investment. Because investment and depreciation were equal in the initial steady state, investment will now be less than depreciation, which means the economy is no longer in a steady state. Gradually, the capital stock falls, leading to reductions in output, consumption, and investment. These variables continue to fall until the economy reaches the new steady state. Because we are assuming that the new steady state is the Golden

figure 7-9



Reducing Saving When Starting With More Capital Than in the Golden Rule Steady State This figure shows what happens over time to output, consumption, and investment when the economy begins with more capital than the Golden Rule level and the saving rate is reduced. The reduction in the saving rate (at time t_0) causes an immediate increase in consumption and an equal decrease in investment. Over time, as the capital stock falls, output, consumption, and investment fall together. Because the economy began with too much capital, the new steady state has a higher level of consumption than the initial steady state.

Rule steady state, consumption must be higher than it was before the change in the saving rate, even though output and investment are lower.

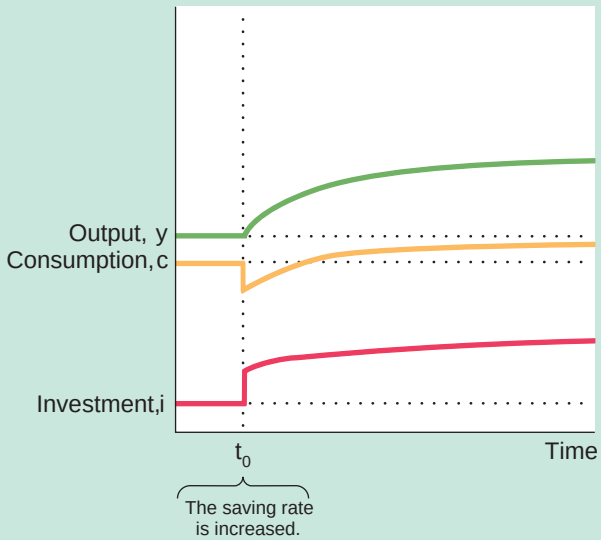
Note that, compared to the old steady state, consumption is higher not only in the new steady state but also along the entire path to it. When the capital stock exceeds the Golden Rule level, reducing saving is clearly a good policy, for it increases consumption at every point in time.

Starting With Too Little Capital When the economy begins with less capital than in the Golden Rule steady state, the policymaker must raise the saving rate to reach the Golden Rule. Figure 7-10 shows what happens. The increase in the saving rate at time t_0 causes an immediate fall in consumption and a rise in investment. Over time, higher investment causes the capital stock to rise. As capital accumulates, output, consumption, and investment gradually increase, eventually approaching the new steady-state levels. Because the initial steady state was below the Golden Rule, the increase in saving eventually leads to a higher level of consumption than that which prevailed initially.

Does the increase in saving that leads to the Golden Rule steady state raise economic welfare? Eventually it does, because the steady-state level of consumption is higher. But achieving that new steady state requires an initial period of reduced consumption. Note the contrast to the case in which the economy begins above the Golden Rule. When the economy begins above the Golden Rule, reaching the Golden Rule produces higher consumption at all points in time. When the economy beg below the Golden Rule, reaching the Golden Rule requires initially reducing consumpti to increase consumption in the future

When deciding whether to try to reach the Golden Rule steady state, policymakers have to take into account that current consumers and future consumers are not always the same people. Reaching the Golden Rule achieves the highest

figure 7-10



Increasing Saving When Starting With Less Capital Than in the Golden Rule Steady State This figure shows what happens over time to output, consumption, and investment when the economy begins with less capital than the Golden Rule, and the saving rate is increased. The increase in the saving rate (at time t_0) causes an immediate drop in consumption and an equal jump in investment. Over time, as the capital stock grows, output, consumption, and investment increase together. Because the economy began with less capital than the Golden Rule, the new steady state has a higher level of consumption than the initial steady state.

steady-state level of consumption and thus benefits future generations. But when the economy is initially below the Golden Rule, reaching the Golden Rule requires raising investment and thus lowering the consumption of current generations. Thus, when choosing whether to increase capital accumulation, the policymaker faces a tradeoff among the welfare of different generations. A policymaker who cares more about current generations than about future generations may decide not to pursue policies to reach the Golden Rule steady state. By contrast, a policymaker who cares about all generations equally will choose to reach the Golden Rule. Even though current generations will consume less, future generations will be better off in number by moving to the Golden Rule.

Thus, optimal capital accumulation depends crucially on how we weigh the interests of current and future generations. The biblical Golden Rule tells us, "Do unto others as you would have them do unto you." If we heed this advice, we give all generations equal weight. In this case, it is optimal to reach the Golden Rule level of capital—which is why it is called the "Golden Rule."

7-3 Population Growth

The basic Solow model shows that capital accumulation, by itself, cannot explain sustained economic growth: high rates of saving lead to high growth temporarily, but the economy eventually approaches a steady state in which capital and output are constant. To explain the sustained economic growth that we observe in most parts of the world, we must expand the Solow model to incorporate the other two sources of economic growth—population growth and technological progress. In this section we add population growth to the model.

Instead of assuming that the population is fixed, as we did in Sections 7-1 and 7-2, we now suppose that the population and the labor force grow at a constant rate n. For example, the U.S. population grows about 1 percent per year, so $n = 0.01$. This means that if 150 million people are working one year, then 151.5 million (1.01×150) are working the next year, and 153.015 million (1.01×151.5) the year after that, and so on.

The Steady State With Population Growth

How does population growth affect the steady state? To answer this question, we must discuss how population growth, along with investment and depreciation, influences the accumulation of capital per worker. As we noted before, investment raises the capital stock, and depreciation reduces it. But now there is a third force acting to change the amount of capital per worker: the growth in the number of workers causes capital per worker to fall.

We continue to let lowercase letters stand for quantities per worker. Thus, $k = K/L$ is capital per worker, and $y = Y/L$ is output per worker. Keep in mind, however, that the number of workers is growing over time.

The change in the capital stock per worker is

$$\Delta k = i - (\delta + n)k.$$

This equation shows how investment, depreciation, and population growth influence the per-worker capital stock. Investment increases k, whereas depreciation and population growth decrease k. We saw this equation earlier in this chapter for the special case of a constant population ($n = 0$).

We can think of the term $(\delta + n)k$ as defining the break-even investment—the amount of investment necessary to keep the capital stock per worker constant. Break-even investment includes the depreciation of existing capital, which equals δk. It also includes the amount of investment necessary to provide new workers with capital. The amount of investment necessary for this purpose is nk, because there are n new workers for each existing worker, and because k is the amount of capital for each worker. The equation shows that population growth reduces the accumulation of capital per worker much the way depreciation does. Depreciation reduces k by wearing out the capital stock, whereas population growth reduces k by spreading the capital stock more thinly among a larger population of workers.[6]

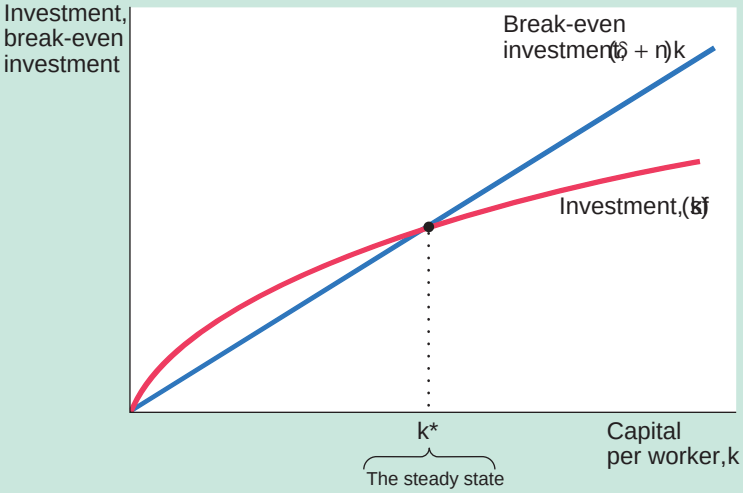
Our analysis with population growth now proceeds much as it did previously. First, we substitute $sf(k)$ for i. The equation can then be written as

$$\Delta k = sf(k) - (\delta + n)k.$$

---

[6] Mathematical note: Formally deriving the equation for the change in k requires a bit of calculus. Note that the change in k per unit of time is $dk/dt = d(K/L)/dt$. After applying the chain rule, we can write this as $dk/dt = (1/L)(dK/dt) - (K/L^2)(dL/dt)$. Now use the following facts to substitute in this equation: $dK/dt = I - \delta K$ and $(dL/dt)/L = n$. After a bit of manipulation, this produces the equation in the text.

figure 7-11



Population Growth in the Solow Model Like depreciation, population growth is one reason why the capital stock per worker shrinks. If n is the rate of population growth and δ is the rate of depreciation, then $(\delta + n)k$ is break-even investment—the amount of investment necessary to keep constant the capital stock per worker k. For the economy to be in a steady state, investment $sf(k)$ must offset the effects of depreciation and population growth $(\delta + n)k$. This is represented by the crossing of the two curves.

To see what determines the steady-state level of capital per worker, we use Figure 7-11, which extends the analysis of Figure 7-4 to include the effects of population growth. An economy is in a steady state if capital per worker k is unchanging. As before, we designate the steady-state value of k as k^*. If k is less than k^*, investment is greater than break-even investment, and k rises. If k is greater than k^*, investment is less than break-even investment, and k falls.

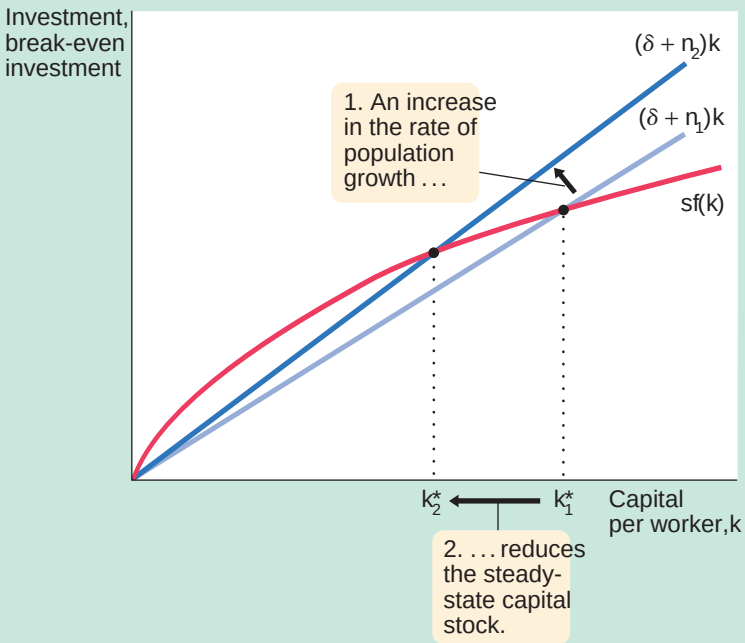
In the steady state, the positive effect of investment on the capital stock per worker exactly balances the negative effects of depreciation and population growth. That is, at k^*, $\Delta k = 0$ and $i^* = \delta k^* + nk^*$. Once the economy is in the steady state, investment has two purposes. Some of it (δk^*) replaces the depreciated capital, and the rest (nk^*) provides the new workers with the steady-state amount of capital.

The Effects of Population Growth

Population growth alters the basic Solow model in three ways. First, it brings us closer to explaining sustained economic growth. In the steady state with population growth, capital per worker and output per worker are constant. Because the number of workers is growing at rate n, however, total capital and total output must also be growing at rate n. Hence, although population growth cannot explain sustained growth in the standard of living (because output per worker is constant in the steady state), it can help explain sustained growth in total output.

Second, population growth gives us another explanation for why some countries are rich and others are poor. Consider the effects of an increase in population growth. Figure 7-12 shows that an increase in the rate of population growth from n_1 to n_2 reduces the steady-state level of capital per worker from k_1^* to k_2^*.

figure 7-12



The Impact of Population Growth An increase in the rate of population growth from n_1 to n_2 shifts the line representing population growth and depreciation upward. The new steady state k_2^* has a lower level of capital per worker than the initial steady state k_1^*. Thus, the Solow model predicts that economies with higher rates of population growth will have lower levels of capital per worker and therefore lower incomes.

Because k^* is lower, and because $y^* = f(k^*)$, the level of output per worker y^* is also lower. Thus, the Solow model predicts that countries with higher population growth will have lower levels of GDP per person.

Finally, population growth affects our criterion for determining the Golden Rule (consumption-maximizing) level of capital. To see how this criterion changes, note that consumption per worker is

$$c = y - i.$$

Because steady-state output is $f(k^*)$ and steady-state investment is $(\delta + n)k^*$, we can express steady-state consumption as

$$c^* = f(k^*) - (\delta + n)k^*.$$

Using an argument largely the same as before, we conclude that the level of k^* that maximizes consumption is the one at which

$$MPK = \delta + n,$$

or equivalently,

$$MPK - \delta = n.$$

In the Golden Rule steady state, the marginal product of capital net of depreciation equals the rate of population growth.

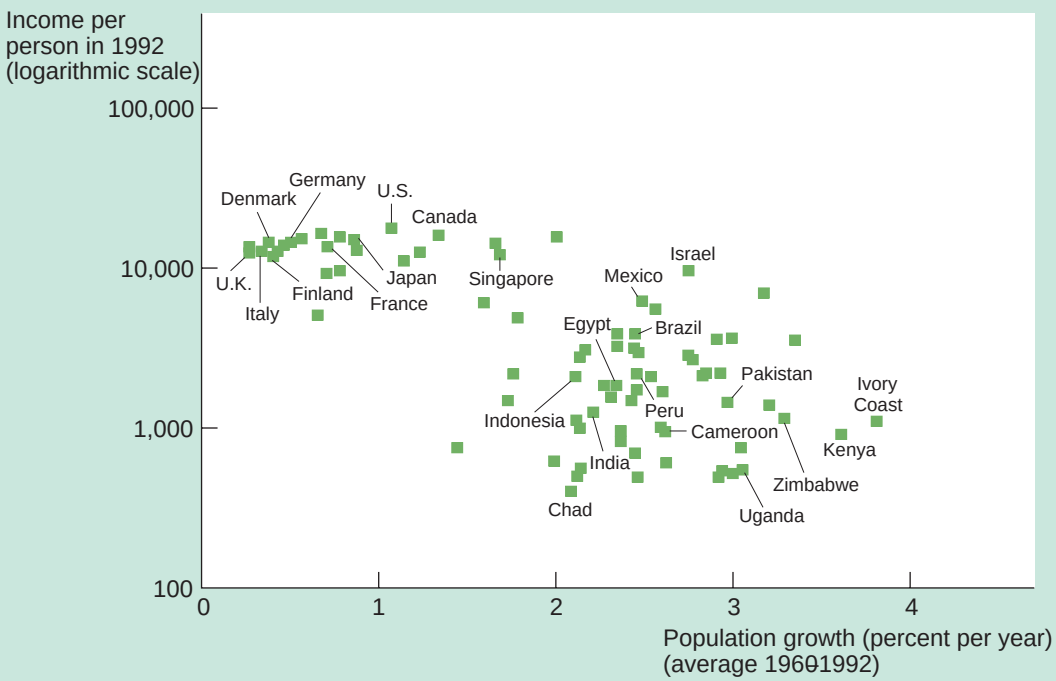
CASE STUDY

Population Growth Around the World

Let's return now to the question of why standards of living vary so much around the world. The analysis we have just completed suggests that population growth may be one of the answers. According to the Solow model, a nation with a high rate of population growth will have a low steady-state capital stock per worker and thus also a low level of income per worker. In other words, high population growth tends to impoverish a country because it is hard to maintain a high level of capital per worker when the number of workers is growing quickly. To see whether the evidence supports this conclusion, we again look at cross-country data.

Figure 7-13 is a scatterplot of data for the same 84 countries examined in the previous case study (and in Figure 7-6). The figure shows that countries with high rates of population growth tend to have low levels of income per person. The international evidence is consistent with our model's prediction that the rate of population growth is one determinant of a country's standard of living.

figure 7-13



International Evidence on Population Growth and Income per Person This figure is a scatterplot of data from 84 countries. It shows that countries with high rates of population growth tend to have low levels of income per person, as the Solow model predicts.

Source: Robert Summers and Alan Heston, Supplement (Mark 5.6) to "The Penn World Table (Mark 5): An Expanded Set of International Comparisons 1950–1988," Quarterly Journal of Economics (May 1991): 327–368.

This conclusion is not lost on policymakers. Those trying to pull the world's poorest nations out of poverty, such as the advisers sent to developing nations by the World Bank, often advocate reducing fertility by increasing education about birth-control methods and expanding women's job opportunities. Toward the same end, China has followed the totalitarian policy of allowing only one child per couple. These policies to reduce population growth should, if the Solow model is right, raise income per person in the long run.

In interpreting the cross-country data, however, it is important to keep in mind that correlation does not imply causation. The data show that low population growth is typically associated with high levels of income per person, and the Solow model offers one possible explanation for this fact, but other explanations are also possible. It is conceivable that high income encourages low population growth, perhaps because birth-control techniques are more readily available in richer countries. The international data can help us evaluate a theory of growth, such as the Solow model, because they show us whether the theory's predictions are borne out in the world. But often more than one theory can explain the same facts.

7-4 Conclusion

This chapter has started the process of building the Solow growth model. The model as developed so far shows how saving and population growth determine the economy's steady-state capital stock and its steady-state level of income per person. As we have seen, it sheds light on many features of actual growth experiences—why Germany and Japan grew so rapidly after being devastated by World War II, why countries that save and invest a high fraction of their output are richer than countries that save and invest a smaller fraction, and why countries with high rates of population growth are poorer than countries with low rates of population growth.

What the model cannot do, however, is explain the persistent growth in living standards we observe in most countries. In the model we now have, when the economy reaches its steady state, output per worker stops growing. To explain persistent growth, we need to introduce technological progress into the model. That is our first job in the next chapter.

Summary

1. The Solow growth model shows that in the long run, an economy's rate of saving determines the size of its capital stock and thus its level of production. The higher the rate of saving, the higher the stock of capital and the higher the level of output.
2. In the Solow model, an increase in the rate of saving causes a period of rapid growth, but eventually that growth slows as the new steady state is reached.

Thus, although a high saving rate yields a high steady-state level of output, saving by itself cannot generate persistent economic growth.

3. The level of capital that maximizes steady-state consumption is called the Golden Rule level. If an economy has more capital than in the Golden Rule steady state, then reducing saving will increase consumption at all points in time. By contrast, if the economy has less capital in the Golden Rule steady state, then reaching the Golden Rule requires increased investment and thus lower consumption for current generations.
4. The Solow model shows that an economy's rate of population growth is another long-run determinant of the standard of living. The higher the rate of population growth, the lower the level of output per worker.

KEY CONCEPTS

Solow growth model Steady state Golden Rule level of capital

QUESTIONS FOR REVIEW

1. In the Solow model, how does the saving rate affect the steady-state level of income? How does it affect the steady-state rate of growth?
2. Why might an economic policymaker choose the Golden Rule level of capital?
3. Might a policymaker choose a steady state with more capital than in the Golden Rule steady state? With less capital than in the Golden Rule steady state? Explain your answers.
4. In the Solow model, how does the rate of population growth affect the steady-state level of income? How does it affect the steady-state rate of growth?

PROBLEMS AND APPLICATIONS

1. Country A and country B both have the production function

$$Y = F(K, L) = K^{1/2}L^{1/2}.$$

   a. Does this production function have constant returns to scale? Explain.

   b. What is the per-worker production function, $y = f(k)$?

   c. Assume that neither country experiences population growth or technological progress and that 5 percent of capital depreciates each year. Assume further that country A saves 10 percent of output each year and country B saves 20 percent of output each year. Using your answer from part (b) and the steady-state condition that investment equals depreciation, find the steady-state level of capital per worker for each country. Then find the steady-state levels of income per worker and consumption per worker.

   d. Suppose that both countries start off with a capital stock per worker of 2. What are the levels of income per worker and consumption per worker? Remembering that the change in the capital stock is investment less depreciation, use a calculator to show how the capital stock per worker will evolve over time in both countries. For each year, calculate income per worker and consumption per worker. How many years will it be before the consumption in country B is higher than the consumption in country A?

2. In the discussion of German and Japanese postwar growth, the text describes what happens

when part of the capital stock is destroyed in a war. By contrast, suppose that a war does not directly affect the capital stock, but that casualties reduce the labor force.

a. What is the immediate impact on total output and on output per person?

b. Assuming that the saving rate is unchanged and that the economy was in a steady state before the war, what happens subsequently to output per worker in the postwar economy? Is the growth rate of output per worker after the war smaller or greater than normal?

3. Consider an economy described by the production function $Y = F(K, L) = K^{0.3}L^{0.7}$.

a. What is the per-worker production function?

b. Assuming no population growth or technological progress, find the steady-state capital stock per worker, output per worker, and consumption per worker as functions of the saving rate and the depreciation rate.

c. Assume that the depreciation rate is 10 percent per year. Make a table showing steady-state capital per worker, output per worker, and consumption per worker for saving rates of 0 percent, 10 percent, 20 percent, 30 percent, and so on. (You will need a calculator with an exponent key for this.) What saving rate maximizes output per worker? What saving rate maximizes consumption per worker?

d. (Harder) Use calculus to find the marginal product of capital. Add to your table the marginal product of capital net of depreciation for each of the saving rates. What does your table show?

4. The 1983 Economic Report of the President contained the following statement: "Devoting a larger share of national output to investment would help restore rapid productivity growth and rising living standards." Do you agree with this claim? Explain.

5. One view of the consumption function is that workers have high propensities to consume and capitalists have low propensities to consume. To explore the implications of this view, suppose that an economy consumes all wage income and saves all capital income. Show that if the factors of production earn their marginal product, this economy reaches the Golden Rule level of capital. (Hint: Begin with the identity that saving equals investment. Then use the steady-state condition that investment is just enough to keep up with depreciation and population growth, and the fact that saving equals capital income in this economy.)

6. Many demographers predict that the United States will have zero population growth in the twenty-first century, in contrast to average population growth of about 1 percent per year in the twentieth century. Use the Solow model to forecast the effect of this slowdown in population growth on the growth of total output and the growth of output per person. Consider the effects both in the steady state and in the transition between steady states.

7. In the Solow model, population growth leads to steady-state growth in total output, but not in output per worker. Do you think this would still be true if the production function exhibited increasing or decreasing returns to scale? Explain. (For the definitions of increasing and decreasing returns to scale, see Chapter 3, "Problems and Applications," Problem 2.)

8. Consider how unemployment would affect the Solow growth model. Suppose that output is produced according to the production function $Y = K^{\alpha}[(1 - u)L]^{1-\alpha}$, where K is capital, L is the labor force, and u is the natural rate of unemployment. The national saving rate, s, is the labor force grows at rate n, and capital depreciates at rate δ.

a. Express output per worker $y = (Y/L)$ as a function of capital per worker $k (= K/L)$ and the natural rate of unemployment. Describe the steady state of this economy.

b. Suppose that some change in government policy reduces the natural rate of unemployment. Describe how this change affects output both immediately and over time. Is the steady-state effect on output larger or smaller than the immediate effect? Explain.

9. Choose two countries that interest you—one rich and one poor. What is the income per person in each country? Find some data on country characteristics that might help explain the difference in income: investment rates, population growth rates, educational attainment, and so on. (Hint: The Web site of the World Bank, www.worldbank.org, is one place to find such data.) How might you figure out which of these factors is most responsible for the observed income difference?

CHAPTER 8 EIGHT

Economic Growth II

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

This chapter continues our analysis of the forces governing long-run economic growth. With the basic version of the Solow growth model as our starting point, we take on four new tasks.

Our first task is to make the Solow model more general and more realistic. In Chapter 3 we saw that capital, labor, and technology are the key determinants of a nation's production of goods and services. In Chapter 7 we developed the Solow model to show how changes in capital (saving and investment) and changes in the labor force (population growth) affect the economy's output. We are now ready to add the third source of growth—changes in technology—into the mix.

Our second task is to examine how a nation's public policies can influence the level and growth of its standard of living. In particular, we address four questions: Should our society save more or save less? How can policy influence the rate of saving? Are there some types of investment that policy should especially encourage? How can policy increase the rate of technological progress? The Solow growth model provides the theoretical framework within which we consider each of these policy issues.

Our third task is to move from theory to empirics. That is, we consider how well the Solow model fits the facts. During the 1990s, a large literature examined the predictions of the Solow model and other models of economic growth. It turns out that the glass is both half full and half empty. The Solow model can shed much light on international growth experiences, but it is far from the last word on the subject.

Our fourth and final task is to consider what the Solow model leaves out. As we have discussed previously, models help us understand the world by simplifying it. After completing an analysis of a model, therefore, it is important to consider

whether we have oversimplified matters. In the last section, we examine a new set of theories, called endogenous growth theories, that hope to explain the technological progress that the Solow model takes as exogenous.

8-1 Technological Progress in the Solow Model

So far, our presentation of the Solow model has assumed an unchanging relationship between the inputs of capital and labor and the output of goods and services. Yet the model can be modified to include exogenous technological progress, which over time expands society's ability to produce.

The Efficiency of Labor

To incorporate technological progress, we must return to the production function that relates total capital K and total labor L to total output Y. Thus far, the production function has been

$$Y = F(K, L).$$

We now write the production function as

$$Y = F(K, L \times E),$$

where E is a new (and somewhat abstract) variable called the efficiency of labor. The efficiency of labor is meant to reflect society's knowledge about production methods: as the available technology improves, the efficiency of labor rises. For instance, the efficiency of labor rose when assembly-line production transformed manufacturing in the early twentieth century, and it rose again when computerization was introduced in the the late twentieth century. The efficiency of labor also rises when there are improvements in the health, education, or skills of the labor force.

The term $L \times E$ measures the number of effective workers in the economy. It takes into account the number of workers L and the efficiency of each worker E. This new production function states that total output Y depends on the number of units of capital K and on the number of effective workers $L \times E$. Increases in the efficiency of labor E are, in effect, like increases in the labor force L.

The simplest assumption about technological progress is that it causes the efficiency of labor E to grow at some constant rate g. For example, if $g = 0.02$, then each unit of labor becomes 2 percent more efficient each year: output increases as if the labor force had increased by an additional 2 percent. This form of technological progress is called labor augmenting, and g is called the rate of labor-augmenting technological progress. Because the labor force L is growing at rate n, and the efficiency of each unit of labor E is growing at rate g, the number of effective workers $L \times E$ is growing at rate $n + g$.

The Steady State With Technological Progress

Expressing technological progress as labor augmenting makes it analogous to population growth. In Chapter 7 we analyzed the economy in terms of quantities per worker and allowed the number of workers to rise over time. Now we analyze the economy in terms of quantities per effective worker and allow the number of effective workers to rise.

To do this, we need to reconsider our notation. We now let $k = K/(L \times E)$ stand for capital per effective worker and $y = Y/(L \times E)$ stand for output per effective worker. With these definitions, we can again write $y = f(k)$.

This notation is not really as new as it seems. If we hold the efficiency of labor E constant at the arbitrary value of 1, as we have done implicitly up to now, then these new definitions of k and y reduce to our old ones. When the efficiency of labor is growing, however, we must keep in mind that k and y now refer to quantities per effective worker (not per actual worker).

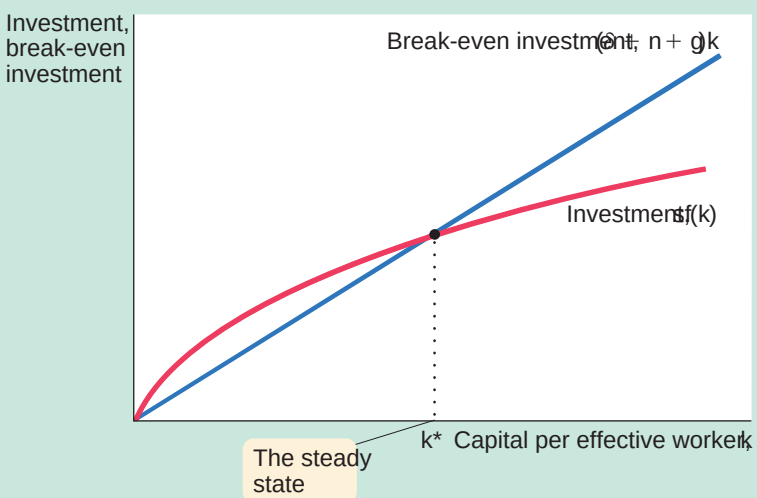
Our analysis of the economy proceeds just as it did when we examined population growth. The equation showing the evolution of k over time now changes to

$$\Delta k = sf(k) - (\delta + n + g)k.$$

As before, the change in the capital stock, Δk, equals investment, $sf(k)$, minus break-even investment, $(\delta + n + g)k$. Now, however, because $k = K/EL$, break-even investment includes three terms: to keep k constant, δk is needed to replace depreciating capital, nk is needed to provide capital for new workers, and gk is needed to provide capital for the new "effective workers" created by technological progress.

As shown in Figure 8-1, the inclusion of technological progress does not substantially alter our analysis of the steady state. There is one level of k, denoted

figure 8-1



Technological Progress and the Solow Growth Model Labor-augmenting technological progress at rate g affects the Solow growth model in much the same way as did population growth at rate n. Now that k is defined as the amount of capital per effective worker, increases in the number of effective workers because of technological progress tend to decrease k. In the steady state, investment $sf(k)$ exactly offsets the reductions in k attributable to depreciation, population growth, and technological progress.

k^*, at which capital per effective worker and output per effective worker are constant. As before, this steady state represents the long-run equilibrium of the economy.

The Effects of Technological Progress

Table 8-1 shows how four key variables behave in the steady state with technological progress. As we have just seen, capital per effective worker k is constant in the steady state. Because $y = f(k)$, output per effective worker is also constant. Remember, though, that the efficiency of each actual worker is growing at rate g. Hence, output per worker ($Y/L = y \times E$) also grows at rate g. Total output [$Y = y \times (E \times L)$] grows at rate $n + g$.

With the addition of technological progress, our model can finally explain the sustained increases in standards of living that we observe. That is, we have shown that technological progress can lead to sustained growth in output per worker. By contrast, a high rate of saving leads to a high rate of growth only until the steady state is reached. Once the economy is in steady state, the rate of growth of output per worker depends only on the rate of technological progress. According to the Solow model, only technological progress can explain persistently rising living standards.

The introduction of technological progress also modifies the criterion for the Golden Rule. The Golden Rule level of capital is now defined as the steady state that maximizes consumption per effective worker. Following the same arguments that we have used before, we can show that steady-state consumption per effective worker is

$$c^* = f(k^*) - (\delta + n + g)k^*.$$

Steady-state consumption is maximized if

$$MPK = \delta + n + g,$$

or

$$MPK - \delta = n + g.$$

That is, at the Golden Rule level of capital, the net marginal product of capital, $MPK - \delta$, equals the rate of growth of total output, $n + g$. Because actual

table 8-1

Steady-State Growth Rates in the Solow Model With Technological Progress

Variable	Symbol	Steady-State Growth Rate
Capital per effective worker	$k = K/(E \times L)$	0
Output per effective worker	$y = Y/(E \times L) = f(k)$	0
Output per worker	$Y/L = y \times E$	g
Total output	$Y = y \times (E \times L)$	$n + g$

economies experience both population growth and technological progress, we must use this criterion to evaluate whether they have more or less capital than at the Golden Rule steady state.

8-2 Policies to Promote Growth

Having used the Solow model to uncover the relationships among the different sources of economic growth, we can now use the theory to help guide our thinking about economic policy.

Evaluating the Rate of Saving

According to the Solow growth model, how much a nation saves and invests is a key determinant of its citizens' standard of living. So let us begin our policy discussion with a natural question: Is the rate of saving in the U.S. economy too low, too high, or about right?

As we have seen, the saving rate determines the steady-state levels of capital and output. One particular saving rate produces the Golden Rule steady state, which maximizes consumption per worker and thus economic well-being. The Golden Rule provides the benchmark against which we can compare the U.S. economy.

To decide whether the U.S. economy is at, above, or below the Golden Rule steady state, we need to compare the marginal product of capital net of depreciation ($MPK - \delta$) with the growth rate of total output ($n + g$). As we just established, at the Golden Rule steady state, $MPK - \delta = n + g$. If the economy is operating with less capital than in the Golden Rule steady state, then diminishing marginal product tells us that $MPK - \delta > n + g$. In this case, increasing the rate of saving will eventually lead to a steady state with higher consumption. However, if the economy is operating with too much capital, then $MPK - \delta < n + g$, and the rate of saving should be reduced.

To make this comparison for a real economy, such as the U.S. economy, we need an estimate of the growth rate ($n + g$) and an estimate of the net marginal product of capital ($MPK - \delta$). Real GDP in the United States grows an average of 3 percent per year, so $n + g = 0.03$. We can estimate the net marginal product of capital from the following three facts:

1. The capital stock is about 2.5 times one year's GDP.
2. Depreciation of capital is about 10 percent of GDP.
3. Capital income is about 30 percent of GDP.

Using the notation of our model (and the result from Chapter 3 that capital owners earn income of MPK for each unit of capital), we can write these facts as

1. $k = 2.5y$.
2. $\delta k = 0.1y$.
3. $MPK \times k = 0.3y$

We solve for the rate of depreciation δ by dividing equation 2 by equation 1:

$$\delta k / k = (0.1y)/(2.5y)$$
$$\delta = 0.04.$$

And we solve for the marginal product of capital MPK by dividing equation 3 by equation 1:

$$(MPK \times k)/k = (0.3y)/(2.5y)$$
$$MPK = 0.12$$

Thus, about 4 percent of the capital stock depreciates each year, and the marginal product of capital is about 12 percent per year. The net marginal product of capital, $MPK - \delta$, is about 8 percent per year.

We can now see that the return to capital ($MPK - \delta = 8$ percent per year) is well in excess of the economy's average growth rate ($n + g = 3$ percent per year). This fact, together with our previous analysis, indicates that the capital stock in the U.S. economy is well below the Golden Rule level. In other words, if the United States saved and invested a higher fraction of its income, it would grow more rapidly and eventually reach a steady state with higher consumption. This finding suggests that policymakers should want to increase the rate of saving and investment. In fact, for many years, increasing capital formation has been a high priority of economic policy.

Changing the Rate of Saving

The preceding calculations show that to move the U.S. economy toward the Golden Rule steady state, policymakers should increase national saving. But how can they do that? We saw in Chapter 3 that, as a matter of sheer accounting, higher national saving means higher public saving, higher private saving, or some combination of the two. Much of the debate over policies to increase growth centers on which of these options is likely to be most effective.

The most direct way in which the government affects national saving is through public saving—the difference between what the government receives in tax revenue and what it spends. When the government's spending exceeds its revenue, the government is said to run a *budget deficit*, which represents negative public saving. As we saw in Chapter 3, a budget deficit raises interest rates and crowds out investment; the resulting reduction in the capital stock is part of the burden of the national debt on future generations. Conversely, if the government spends less than it raises in revenue, it is said to run a *budget surplus* and can then retire some of the national debt and stimulate investment.

The government also affects national saving by influencing private saving—the saving done by households and firms. In particular, how much people decide to save depends on the incentives they face, and these incentives are altered by a variety of public policies. Many economists argue that high tax rates on capital—including the corporate income tax, the federal income tax, the estate tax, and many state income and estate taxes—discourage private saving by reducing the

rate of return that savers earn. However, tax-exempt retirement accounts, such as IRAs, are designed to encourage private saving by giving preferential treatment to income saved in these accounts.

Many disagreements among economists over public policy are rooted in different views about how much private saving responds to incentives. For example, suppose that the government were to expand the amount that people could put into tax-exempt retirement accounts. Would people respond to the increased incentive to save by saving more? Or would people merely transfer saving done in other forms into these accounts, reducing tax revenue and thus public saving without any stimulus to private saving? Clearly, the desirability of the policy depends on the answers to these questions. Unfortunately, despite much research on this issue, no consensus has emerged.

CASE STUDY

Should the Social Security System Be Reformed?

Although many government policies are designed to encourage saving, such as the preferential tax treatment given to pension plans and other retirement accounts, one important policy is often thought to reduce saving: the Social Security system. Social Security is a transfer system designed to maintain individuals' income in their old age. These transfers to the elderly are financed with a payroll tax on the working-age population. This system is thought to reduce private saving because it reduces individuals' need to provide for their own retirement.

To counteract the reduction in national saving attributed to Social Security, many economists have proposed reforms of the Social Security system. The system is now largely pay-as-you-go—most of the current tax receipts are paid out to the current elderly population. One suggestion is that Social Security should be fully funded. Under this plan, the government would put aside in a trust fund the payments a generation makes when it is young and working; the government would then pay out the principal and accumulated interest to this same generation when it is older and retired. Under a fully funded Social Security system, an increase in public saving would offset the reduction in private saving.

A closely related proposal is *privatization*, which means turning this government program for the elderly into a system of mandatory private savings accounts, much like private pension plans. In principle, the issues of funding and privatization are distinct. A fully funded system could be either public (in which case the government holds the funds) or private (in which case private financial institutions hold the funds). In practice, however, the issues are often linked. Some economists have argued that a fully funded public system is problematic. They note that such a system would end up holding a large share of the nation's wealth, which would increase the role of the government in allocating capital. In addition, they fear that a large publicly controlled fund would tempt politicians to cut taxes or increase spending, which could deplete the fund and cause the system to revert to pay-as-you-go status. History gives some support to this fear: the initial architects of Social Security wanted the system to accumulate a much larger trust fund than ever materialized.

These issues rose to prominence in the late 1990s as policymakers became aware that the current Social Security system was not sustainable. That is, the amount of revenue being raised by the payroll tax appeared insufficient to pay all the benefits being promised. According to most projections, this problem was to become acute as the large baby-boom generation retired during the early decades of the twenty-first century. Various solutions were proposed. One possibility was to maintain the current system with some combination of smaller benefits and higher taxes. Other possibilities included movements toward a fully funded system, perhaps also including private accounts. This issue was prominent in the presidential campaign of 2000, with candidate George W. Bush advocating a reform including private accounts. As this book was going to press, it was still unclear whether this reform would come to pass.[1]

Allocating the Economy's Investment

The Solow model makes the simplifying assumption that there is only one type of capital. In the world, of course, there are many types. Private businesses invest in traditional types of capital, such as bulldozers and steel plants, and newer types of capital, such as computers and robots. The government invests in various forms of public capital, called infrastructure, such as roads, bridges, and sewer systems.

In addition, there is human capital—the knowledge and skills that workers acquire through education, from early childhood programs such as Head Start to on-the-job training for adults in the labor force. Although the basic Solow model includes only physical capital and does not try to explain the efficiency of labor, in many ways human capital is analogous to physical capital. Like physical capital, human capital raises our ability to produce goods and services. Raising the level of human capital requires investment in the form of teachers, libraries, and student time. Recent research on economic growth has emphasized that human capital is at least as important as physical capital in explaining international differences in standards of living.[2]

Policymakers trying to stimulate economic growth must confront the issue of what kinds of capital the economy needs most. In other words, what kinds of capital yield the highest marginal products? To a large extent, policymakers can rely on the marketplace to allocate the pool of saving to alternative types of investment. Those industries with the highest marginal products of capital will naturally be most willing to borrow at market interest rates to finance new investment. Many economists advocate that the government should merely create a "level playing field" for different types of capital—for example, by ensuring that the tax system treats all forms of capital equally. The government can then rely on the market to allocate capital efficiently.

---

[1] To learn more about the debate over Social Security, see *Social Security Reform: Links to Saving, Investment, and Growth*, Steven A. Sass and Robert K. Triest, eds., Conference Series No. 41, Federal Reserve Bank of Boston, June 1997.

[2] N. Gregory Mankiw, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* (May 1992): 407–437.

Other economists have suggested that the government should actively encourage particular forms of capital. Suppose, for instance, that technological advance occurs as a by-product of certain economic activities. This would happen if new and improved production processes are devised during the process of building capital (a phenomenon called *learning by doing*) and if these ideas become part of society's pool of knowledge. Such a by-product is called a *technological externality* (or a *knowledge spillover*). In the presence of such externalities, the social returns to capital exceed the private returns, and the benefits of increased capital accumulation to society are greater than the Solow model suggests.[3] Moreover, some types of capital accumulation may yield greater externalities than others. If, for example, installing robots yields greater technological externalities than building a new steel mill, then perhaps the government should use the tax laws to encourage investment in robots. The success of such *industrial policy*, as it is sometimes called, requires that the government be able to measure the externalities of different economic activities so it can give the correct incentive to each activity.

Most economists are skeptical about industrial policies, for two reasons. First, measuring the externalities from different sectors is difficult, so it is likely to be virtually impossible. If policy is based on poor measurements, its effects might be close to random and, thus, worse than no policy at all. Second, the political process is far from perfect. Once the government gets in the business of rewarding specific industries with subsidies and tax breaks, the rewards are as likely to be based on political clout as on the magnitude of externalties.

One type of capital that necessarily involves the government is public capital. Local, state, and federal governments are always deciding whether to borrow to finance new roads, bridges, and transit systems. During his first presidential campaign, Bill Clinton argued that the United States had been investing too little in infrastructure. He claimed that a higher level of infrastructure investment would make the economy substantially more productive. Among economists, this claim had both defenders and critics. Yet all of them agree that measuring the marginal product of public capital is difficult. Private capital generates an easily measured rate of profit for the firm owning the capital, whereas the benefits of public capital are more diffuse.

Encouraging Technological Progress

The Solow model shows that sustained growth in income per worker must come from technological progress. The Solow model, however, takes technological progress as exogenous; it does not explain it. Unfortunately, the determinants of technological progress are not well understood.

Despite this limited understanding, many public policies are designed to stimulate technological progress. Most of these policies encourage the private sector to devote resources to technological innovation. For example, the patent system

[3] Paul Romer, "Crazy Explanations for the Productivity Slowdown," *NBER Macroeconomics Annual* 2 (1987): 163–201.

gives a temporary monopoly to inventors of new products; the tax code offers tax breaks for firms engaging in research and development; and government agencies such as the National Science Foundation directly subsidize basic research in universities. In addition, as discussed above, proponents of industrial policy argue that the government should take a more active role in promoting specific industries that are key for rapid technological progress.

CASE STUDY

The Worldwide Slowdown in Economic Growth

Beginning in the early 1970s, world policymakers faced a perplexing problem: a global slowdown in economic growth. Table 8-2 presents data on the growth in real GDP per person for the seven major world economies. Growth in the United States fell from 2.2 percent to 1.5 percent, and other countries experienced similar or more severe declines. Accumulated over many years, even a small change in the rate of growth has a large effect on economic well-being. Real income in the United States today is about 20 percent lower than it would have been had growth remained at its previous level.

Why did this slowdown occur? Studies have shown that it was attributable to a fall in the rate at which the production function was improving over time. The appendix to this chapter explains how economists measure changes in the production function with a variable called total factor productivity, which is closely related to the efficiency of labor in the Solow model. There are, however, many hypotheses to explain this fall in productivity growth. Here are four of them.

Measurement Problems One possibility is that the productivity slowdown did not really occur and that it shows up in the data because the data are flawed. As you may recall from Chapter 2, one problem in measuring inflation is correcting for changes in the quality of goods and services. The same issue arises when measuring output and productivity. For instance, if technological advance leads to more computers being built, then the increase in output and productivity is easy to measure. But if technological advance leads to faster computers being built, then output and productivity have increased, but that increase is more subtle and harder to measure. Government statisticians try to correct for changes in quality, but despite their best efforts, the resulting data are far from perfect.

Unmeasured quality improvements mean that our standard of living is rising more rapidly than the official data indicate. This issue should make us suspicious of the data, but by itself it cannot explain the productivity slowdown. To explain a slowdown in growth, one must argue that the measurement problems got worse. There is some indication that this might be so. As history passes, fewer people work in industries with tangible and easily measured output, such as agriculture, and more work in industries with intangible and less easily measured output, such as medical services. Yet few economists believe that measurement problems were the full story.

Oil Prices When the productivity slowdown began around 1973, the obvious hypothesis to explain it was the large increase in oil prices caused by the actions of the OPEC oil cartel. The primary piece of evidence was the timing: productivity growth slowed at the same time that oil prices skyrocketed. Over time, however,

table 8-2

The Slowdown in Growth Around the World

Country	GROWTH IN OUTPUT PER PERSON (PERCENT PER YEAR) 1948–1972	1972–1995	1995–2000
Canada	2.9	1.8	2.7
France	4.3	1.6	2.2
West Germany	5.7	2.0	
Germany			1.7
Italy	4.9	2.3	4.7
Japan	8.2	2.6	1.1
United Kingdom	2.4	1.8	2.5
United States	2.2	1.5	2.9

Source: Angus Maddison, Phases of Capitalist Development (Oxford: Oxford University Press, 1982); and OECD National Accounts and International Financial Statistics.
Note: Data before 1995 for Germany refer to West Germany; after 1995, to the United Germany.

this explanation has appeared less likely. One reason is that the accumulated shortfall in productivity seems too large to be explained by an increase in oil prices that is not that large a fraction of the typical firm's costs. In addition, if this explanation were right, productivity should have sped up when political turmoil in OPEC caused oil prices to plummet in 1986. Unfortunately, that did not happen.

Worker Quality Some economists suggest that the productivity slowdown might have been caused by changes in the labor force. In the early 1970s, the large baby-boom generation started leaving school and taking jobs. At the same time, changing social norms encouraged many women to leave full-time housework and enter the labor force. Both of these developments lowered the average level of experience among workers, which in turn lowered average productivity.

Other economists point to changes in worker quality as gauged by human capital. Although the educational attainment of the labor force continued to rise throughout this period, it was not increasing as rapidly as it did in the past. Moreover, declining performance on some standardized tests suggests that the quality of education was declining. If so, this could explain slowing productivity growth.

The Depletion of Ideas Still other economists suggest that the world started to run out of new ideas about how to produce in the early 1970s, pushing the economy into an age of slower technological progress. These economists often argue that the anomaly is not the period since 1970 but the preceding two decades. In the late 1940s, the economy had a large backlog of ideas that had not been fully implemented because of the Great Depression of the 1930s and World War II in the first half of 1940s. After the economy used up this backlog, the argument goes, a slowdown in productivity growth was likely. Indeed, although the growth rates in the 1970s, 1980s, and early 1990s were disappointing compared to those of the 1950s and 1960s, they were not lower than average growth rates from 1870 to 1950.

Which of these suspects is the culprit? All of them are plausible, but it is difficult to prove beyond a reasonable doubt that any one of them is guilty. The worldwide slowdown in economic growth that began in the mid-1970s remains a mystery.[4]

CASE STUDY

Information Technology and the New Economy

As any good doctor will tell you, sometimes a patient's illness goes away on its own, even if the doctor has failed to come up with a convincing diagnosis and remedy. This seems to be the outcome with the productivity slowdown discussed in the previous case study. Economists have not figured it out, but beginning in the middle of the 1990s, the problem disappeared. Economic growth took off, as shown in the third column of Table 8-2. In the United States, output per person accelerated from 1.5 to 2.9 percent per year. Commentators proclaimed we were living in a "new economy."

As with the slowdown in economic growth in the 1970s, the acceleration in the 1990s is hard to explain definitively. But part of the credit goes to advances in computer and information technology, including the Internet.

Observers of the computer industry often cite Moore's law, which states that the price of computing power falls by half every 18 months. This is not an inevitable law of nature but an empirical regularity describing the rapid technological progress this industry has enjoyed. In the 1980s and early 1990s, economists were surprised that the rapid progress in computing did not have a larger effect on the overall economy. Economist Robert Solow once quipped that "you can see the computer age everywhere but in the productivity statistics."

There are two reasons why the macroeconomic effects of the computer revolution might not have showed up until the mid-1990s. One is that the computer industry was previously only a small part of the economy. In 1990, computer hardware and software represented 0.9 percent of real GDP; by 1999, this share had risen to 4.2 percent. As computers made up a larger part of the economy, technological advance in that sector had a greater overall effect.

The second reason why the productivity benefits of computers may have been delayed is that it took time for firms to figure out how best to use the technology. Whenever firms change their production systems and train workers to use a technology, they disrupt the existing means of production. Measured productivity can fall for a while before the economy reaps the full benefits. Indeed, some economists even suggest that the spread of computers can help explain the productivity slowdown that began in the 1970s.

Economic history provides some support for the idea that new technologies influence growth with a long lag. The electric light bulb was invented in 1879. But it took several decades before electricity had a big economic impact. For businesses to reap large productivity gains, they had to do more than simply re-

[4] For various views on the growth slowdown, see "Symposium: The Slowdown in Productivity Growth," The Journal of Economic Perspectives 2 (Fall 1988): 3–98.

place steam engines with electric motors; they had to rethink the entire organization of factories. Similarly, replacing the typewriters on desks with computers and word processing programs, as was common in the 1980s, may have had small productivity effects. Only later, when the Internet and other advanced applications were invented, did the computers yield large economic gains.

Eventually, advances in technology should show up in economic growth, as was the case in the second half of the 1990s. This extra growth occurs through three channels. First, because the computer industry is part of the economy, productivity growth in that industry directly affects overall productivity growth. Second, because computers are a type of capital good, falling computer prices allow firms to accumulate more computing capital for every dollar of investment spending; the resulting increase in capital accumulation raises growth in all sectors that use computers as a factor of production. Third, the innovations in the computer industry may induce other industries to reconsider their own production methods, which in turn leads to productivity growth in those industries.

The big, open question is whether the computer industry will remain an engine of growth. Will Moore's law describe the future as well as it has described the past? Will the technological advances of the next decade be as profound as the Internet was during the 1990s? Stay tuned.[5]

8-3 From Growth Theory to Growth Empirics

So far in this chapter we have introduced exogenous technological progress into the Solow model to explain sustained growth in standards of living. We then used the theoretical framework as a lens through which to view some key issues facing policymakers. Let us now discuss what happens when the theory is asked to confront the facts.

Balanced Growth

According to the Solow model, technological progress causes the values of many variables to rise together in the steady state. This property, called balanced growth, does a good job of describing the long-run data for the U.S. economy.

Consider first output per worker, Y/L, and the capital stock per worker, K/L. According to the Solow model, in the steady state, both of these variables grow at the rate of technological progress. United States data for the half century show that output per worker and the capital stock per worker have in fact grown at approximately the same rate—about 2 percent per year. To put it another way, the capital–output ratio has remained approximately constant over time.

[5] For more on this topic, see the symposium "Computers and Productivity" in the Fall 2000 issue of *The Journal of Economic Perspectives*. On the parallel between electricity and computers, see Paul A. David, "The Dynamo and the Computer: A Historical Perspective on the Modern Productivity Paradox," *American Economic Review* 80, no. 2 (May 1990): 355–361.

Technological progress also affects factor prices. Problem 3(d) at the end of the chapter asks you to show that, in the steady state, the real wage grows at the rate of technological progress. The real rental price of capital, however, is constant over time. Again, these predictions hold true for the United States. Over the past 50 years, the real wage has increased about 2 percent per year; it has increased about the same amount as real GDP per worker. Yet the real rental price of capital (measured as real capital income divided by the capital stock) has remained about the same.

The Solow model's prediction about factor prices—and the success of this prediction—is especially noteworthy when contrasted with Karl Marx's theory of the development of capitalist economies. Marx predicted that the return to capital would decline over time and that this would lead to economic and political crises. Economic history has not supported Marx's prediction, which partly explains why we now study Solow's theory of growth rather than Marx's.

Convergence

If you travel around the world, you will see tremendous variations in living standards. The world's poor countries have average levels of income per person that are less than one-tenth the average levels in the world's rich countries. These differences in income are reflected in almost every measure of the quality of life, from the number of televisions and telephones per household to the infant mortality rate and life expectancy.

Much research has been devoted to the question of whether economies converge over time to one another. In particular, do economies that start off poor subsequently grow faster than economies that start off rich? If they do, then the world's poor economies will tend to catch up with the world's rich economies. This property of catch-up is called **convergence**.

To understand the study of convergence, consider an analogy. Imagine that you were to collect data on college students. At the end of their first year, some students have A averages, whereas others have C averages. Would you expect the A and the C students to converge over the remaining three years of college? The answer depends on why their first-year grades differed. If the differences arose because some students came from better high schools than others, then you might expect those who were initially disadvantaged to start catching up to their better-prepared peers. But if the differences arose because some students study more than others, you might expect the differences in grades to persist.

The Solow model predicts that much the same is true with nations: whether economies converge depends on why they differed in the first place. On the one hand, if two economies with the same steady state happened by historical accident to start off with different capital stocks, then we should expect them to converge. The economy with the smaller capital stock will naturally grow more quickly. (In a case study in Chapter 7, we applied this logic to explain rapid growth in Germany and Japan after World War II.) On the other hand, if two economies have different steady states, perhaps because the economies have different rates of saving, then we should not expect convergence. Instead, each economy will approach its own steady state.

Experience is consistent with this analysis. In samples of economies with similar cultures and policies, studies find that economies converge to one another at a rate of about 2 percent per year. That is, the gap between rich and poor economies closes by about 2 percent each year. An example is the economies of individual American states. For historical reasons, such as the Civil War of the 1860s, income levels varied greatly among states a century ago. Yet these differences have slowly disappeared over time.

In international data, a more complex picture emerges. When researchers examine only data on income per person, they find little evidence of convergence: countries that start off poor do not grow faster on average than countries that start off rich. This finding suggests that different countries have different steady states. If statistical techniques are used to control for some of the determinants of the steady state, such as saving rates, population growth rates, and educational attainment, then once again the data show convergence at a rate of about 2 percent per year. In other words, the economies of the world exhibit conditional convergence: they appear to be converging to their own steady states, which in turn are determined by saving, population growth, and education.[6]

Factor Accumulation Versus Production Efficiency

As a matter of accounting, international differences in income per person can be attributed to either (1) differences in the factors of production, such as the quantities of physical and human capital, or (2) differences in the efficiency with which economies use their factors of production. That is, a worker in a poor country may be poor because he lacks tools and skills or because the tools and skills he has are not being put to their best use. To describe this issue in terms of the Solow model, the question is whether the large gap between rich and poor is explained by differences in capital accumulation (including human capital) or differences in the production function.

Much research has attempted to estimate the relative importance of these two sources of income disparities. The exact answer varies from study to study, but both factor accumulation and production efficiency appear important. Moreover, a common finding is that they are positively correlated: nations with high levels of physical and human capital also tend to use those factors efficiently.[7]

There are several ways to interpret this positive correlation. One hypothesis is that an efficient economy may encourage capital accumulation. For example, a person in a well-functioning economy may have greater resources and incentive

---

[6] Robert Barro and Xavier Sala-i-Martin, "Convergence Across States and Regions," *Brookings Papers on Economic Activity* 1 (1991): 107–182; and N. Gregory Mankiw, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* (May 1992): 407–437.

[7] Robert E. Hall and Charles I. Jones, "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics* 114 (February 1999): 83–116; and Peter J. Klenow and Andres Rodriguez-Clare, "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual* (1997): 73–103.

to stay in school and accumulate human capital. Another hypothesis is that capital accumulation may induce greater efficiency. If there are positive externalities to physical and human capital, a possibility mentioned earlier in the chapter, then countries that save and invest more will appear to have better production functions (unless the research study accounts for these externalities, which is hard to do). Thus, greater production efficiency may cause greater factor accumulation, or the other way around.

A final hypothesis is that both factor accumulation and production efficiency are driven by a common third variable. Perhaps the common third variable is the quality of the nation's institutions, including the government's policymaking process. As one economist put it, when governments screw up, they screw up big time. Bad policies, such as high inflation, excessive budget deficits, widespread market interference, and rampant corruption, often go hand in hand. We should not be surprised that such economies both accumulate less capital and fail to use the capital they have efficiently as they might.

8-4 Beyond the Solow Model: Endogenous Growth Theory

> A chemist, a physicist, and an economist are all trapped on a desert island, trying to figure out how to open a can of food.
>
> "Let's heat the can over the fire until it explodes," says the chemist.
>
> "No, no," says the physicist. "Let's drop the can onto the rocks from the top of a high tree."
>
> "I have an idea," says the economist. "First, we assume a can opener." . . .

This old joke takes aim at how economists use assumptions to simplify—and sometimes oversimplify—the problems they face. It is particularly apt when evaluating the theory of economic growth. One goal of growth theory is to explain the persistent rise in living standards that we observe in most parts of the world. The Solow growth model shows that such persistent growth must come from technological progress. But where does technological progress come from? In the Solow model, it is simply assumed!

To understand fully the process of economic growth, we need to go beyond the Solow model and develop models that explain technological progress. Models that do this often go by the label **endogenous growth theory**, because they reject the Solow model's assumption of exogenous technological change. Although the field of endogenous growth theory is large and sometimes complex, here we get a quick sampling of this modern research.[8]

---

[8] This section provides a brief introduction to the large and fascinating literature on endogenous growth theory. Early and important contributions to this literature include Paul M. Romer, "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94 (October 1986): 1002–1037; and Robert E. Lucas, Jr., "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22 (1988): 3–42. The reader can learn more about this topic in the undergraduate textbook by Charles I. Jones, *Introduction to Economic Growth* (New York: Norton, 1998).

The Basic Model

To illustrate the idea behind endogenous growth theory, let's start with a particularly simple production function:

$$Y = AK,$$

where Y is output, K is the capital stock, and A is a constant measuring the amount of output produced for each unit of capital. Notice that this production function does not exhibit the property of diminishing returns to capital. One extra unit of capital produces A extra units of output, regardless of how much capital there is. This absence of diminishing returns to capital is the key difference between this endogenous growth model and the Solow model.

Now let's see how this production function relates to economic growth. As before, we assume a fraction s of income is saved and invested. We therefore describe capital accumulation with an equation similar to those we used previously:

$$\Delta K = sY - \delta K.$$

This equation states that the change in the capital stock (ΔK) equals investment (sY) minus depreciation (δK). Combining this equation with the $Y = AK$ production function, we obtain, after a bit of manipulation,

$$\Delta Y/Y = \Delta K/K = sA - \delta.$$

This equation shows what determines the growth rate of output, $\Delta Y/Y$. Notice that, as long as $sA > \delta$, the economy's income grows forever, even without the assumption of exogenous technological progress.

Thus, a simple change in the production function can alter dramatically the predictions about economic growth. In the Solow model, saving leads to growth temporarily, but diminishing returns to capital eventually force the economy to approach a steady state in which growth depends only on exogenous technological progress. By contrast, in this endogenous growth model, saving and investment can lead to persistent growth.

But is it reasonable to abandon the assumption of diminishing returns to capital? The answer depends on how we interpret the variable K in the production function $Y = AK$. If we take the traditional view that K includes only the economy's stock of plants and equipment, then it is natural to assume diminishing returns. Giving 10 computers to each worker does not make the worker 10 times as productive as he or she is with one computer.

Advocates of endogenous growth theory, however, argue that the assumption of constant (rather than diminishing) returns to capital is more palatable if K is interpreted more broadly. Perhaps the best case for the endogenous growth model is to view knowledge as a type of capital. Clearly, knowledge is an important input into the economy's production—both its production of goods and services and its production of new knowledge. Compared to other forms of capital, however, it is less natural to assume that knowledge exhibits the property of diminishing returns. (Indeed, the increasing pace of scientific and technological innovation over the past few centuries has led some economists to argue that there are increasing

returns to knowledge.) If we accept the view that knowledge is a type of capital, then this endogenous growth model with its assumption of constant returns to capital becomes a more plausible description of long-run economic growth.

A Two-Sector Model

Although the $Y = AK$ model is the simplest example of endogenous growth, the theory has gone well beyond this. One line of research has tried to develop models with more than one sector of production in order to offer a better description of the forces that govern technological progress. To see what we might learn from such models, let's sketch out an example.

The economy has two sectors, which we can call manufacturing firms and research universities. Firms produce goods and services, which are used for consumption and investment in physical capital. Universities produce a factor of production called "knowledge", which is then freely used in both sectors. The economy is described by the production function for firms, the production function for universities, and the capital-accumulation equation:

$$Y = F[K, (1 - u)EL] \quad \text{(production function in manufacturing firms)},$$
$$\Delta E = g(u)E \quad \text{(production function in research universities)},$$
$$\Delta K = sY - \delta K \quad \text{(capital accumulation)},$$

where u is the fraction of the labor force in universities (and $1 - u$ is the fraction in manufacturing), E is the stock of knowledge (which in turn determines the efficiency of labor), and g is a function that shows how the growth in knowledge depends on the fraction of the labor force in universities. The rest of the notation is standard. As usual, the production function for the manufacturing firms is assumed to have constant returns to scale: if we double both the amount of physical capital (K) and the number of effective workers in manufacturing [$(1 - u)EL$], we double the output of goods and services (Y).

This model is a cousin of the $Y = AK$ model. Most important, this economy exhibits constant (rather than diminishing) returns to capital, as long as capital is broadly defined to include knowledge. In particular, if we double both physical capital K and knowledge E, then we double the output of both sectors in the economy. As a result, like the $Y = AK$ model, this model can generate persistent growth without the assumption of exogenous shifts in the production function. Here persistent growth arises endogenously because the creation of knowledge in universities never slows down.

At the same time, however, this model is also a cousin of the Solow growth model. If u, the fraction of the labor force in universities, is held constant, then the efficiency of labor E grows at the constant rate $g(u)$. This result of constant growth in the efficiency of labor at rate g is precisely the assumption made in the Solow model with technological progress. Moreover, the rest of the model—the manufacturing production function and the capital-accumulation equation—also resembles the rest of the Solow model. As a result, for any given value of u, this endogenous growth model works just like the Solow model.

There are two key decision variables in this model. As in the Solow model, the fraction of output used for saving and investment, s, determines the steady-state stock of physical capital. In addition, the fraction of labor in universities, u, determines the growth in the stock of knowledge. Both s and u affect the level of income, although only u affects the steady-state growth rate of income. Thus, this model of endogenous growth takes a small step in the direction of showing which societal decisions determine the rate of technological change.

The Microeconomics of Research and Development

The two-sector endogenous growth model just presented takes us closer to understanding technological progress, but it still tells only a rudimentary story about the creation of knowledge. If one thinks about the process of research and development for even a moment, three facts become apparent. First, although knowledge is largely a public good (that is, a good freely available to everyone), much research is done by firms that are driven by the profit motive. Second, research is profitable because innovations give firms temporary monopolies, either because of the patent system or because there is an advantage to being the first firm on the market with a new product. Third, when one firm innovates, other firms build on that innovation to produce the next generation of innovations. These (essentially microeconomic) facts are not easily connected with the (essentially macroeconomic) growth models we have discussed so far.

Some endogenous growth models try to incorporate these facts about research and development. Doing this requires modeling the decisions that firms face as they engage in research and modeling the interactions among firms that have some degree of monopoly power over their innovations. Going into more detail about these models is beyond the scope of this book. But it should be clear already that one virtue of these endogenous growth models is that they offer a more complete description of the process of technological innovation.

One question these models are designed to address is whether, from the standpoint of society as a whole, private profit-maximizing firms tend to engage in too little or too much research. In other words, is the social return to research (which is what society cares about) greater or smaller than the private return (which is what motivates individual firms)? It turns out that, as a theoretical matter, there are effects in both directions. On the one hand, when a firm creates a new technology, it makes other firms better off by giving them a base of knowledge on which to build future research. As Isaac Newton famously remarked, "If I have seen farther than others, it is because I was standing on the shoulders of giants." On the other hand, when one firm invests in research, it can also make other firms worse off by merely being first to discover a technology that another firm would have invented. This duplication of research effort has been called the "stepping on toes" effect. Whether firms left to their own devices do too little or too much research depends on whether the positive "standing on shoulders" externality or the negative "stepping on toes" externality is more prevalent.

Although theory alone is ambiguous about the optimality of research effort, the empirical work in this area is usually less so. Many studies have suggested the "standing on shoulders" externality is important and, as a result, the social return to research is large—often in excess of 40 percent per year. This is an impressive rate of return, especially when compared to the return to physical capital, which we earlier estimated to be about 8 percent per year. In the judgment of some economists, this finding justifies substantial government subsidies to research.[9]

8-5 Conclusion

Long-run economic growth is the single most important determinant of the economic well-being of a nation's citizens. Everything else that macroeconomists study—unemployment, inflation, trade deficits, and so on—pales in comparison.

Fortunately, economists know quite a lot about the forces that govern economic growth. The Solow growth model and the more recent endogenous growth models show how saving, population growth, and technological progress interact in determining the level and growth of a nation's standard of living. Although these theories offer no magic pill to ensure an economy achieves rapid growth, they do offer much insight, and they provide the intellectual framework for much of the debate over public policy.

Summary

1. In the steady state of the Solow growth model, the growth rate of income per person is determined solely by the exogenous rate of technological progress.
2. In the Solow model with population growth and technological progress, the Golden Rule (consumption-maximizing) steady state is characterized by equality between the net marginal product of capital, $MPK - \delta$, and the steady-state growth rate of total income, $n + g$. By contrast, in the U.S. economy, the net marginal product of capital is well in excess of the growth rate, indicating that the U.S. economy has much less capital than in the Golden Rule steady state.
3. Policymakers in the United States and other countries often claim that their nations should devote a larger percentage of their output to saving and investment. Increased public saving and tax incentives for private saving are two ways to encourage capital accumulation.
4. In the early 1970s, the rate of growth fell substantially in most industrialized countries. The cause of this slowdown is not well understood. In the mid-1990s, the rate of growth increased, most likely because of advances in information technology.

---

[9] For an overview of the empirical literature on the effects of research, see Zvi Griliches, "The Search for R&D Spillovers," *Scandinavian Journal of Economics* 94 (1991): 29–47.

5. Many empirical studies have examined to what extent the Solow model can help explain long-run economic growth. The model can explain much of what we see in the data, such as balanced growth and conditional convergence. Recent studies have also found that international variation in standards of living is attributable to a combination of capital accumulation and the efficiency with which capital is used.
6. Modern theories of endogenous growth attempt to explain the rate of technological progress, which the Solow model takes as exogenous. These models try to explain the decisions that determine the creation of knowledge through research and development.

KEY CONCEPTS

Efficiency of labor

Labor-augmenting technological progress

Endogenous growth theory

QUESTIONS FOR REVIEW

1. In the Solow model, what determines the steady-state rate of growth of income per worker?
2. What data would you need to determine whether an economy has more or less capital than in the Golden Rule steady state?
3. How can policymakers influence a nation's saving rate?
4. What has happened to the rate of productivity growth over the past 40 years? How might you explain this phenomenon?
5. In the steady state of the Solow model, at what rate does output per person grow? At what rate does capital per person grow? How does this compare with U.S. experience?
6. How does endogenous growth theory explain persistent growth without the assumption of exogenous technological progress? How does this differ from the Solow model?

PROBLEMS AND APPLICATIONS

1. An economy described by the Solow growth model has the following production function:

$$y = \sqrt{k}.$$

   a. Solve for the steady-state value of y as a function of s, n, g, and δ.

   b. A developed country has a saving rate of 28 percent and a population growth rate of 1 percent per year. A less-developed country has a saving rate of 10 percent and a population growth rate of 4 percent per year. In both countries, $g = 0.02$ and $\delta = 0.04$. Find the steady-state value of y for each country.

   c. What policies might the less-developed country pursue to raise its level of income?

2. In the United States, the capital share of GDP is about 30 percent; the average growth in output is about 3 percent per year; the depreciation rate is about 4 percent per year; and the capital–output ratio is about 2.5. Suppose that the production function is Cobb–Douglas, so that the capital

share in output is constant, and that the United States has been in a steady state. (For a discussion of the Cobb–Douglas production function, see the appendix to Chapter 3.)

a. What must the saving rate be in the initial steady state? [Hint: Use the steady-state relationship $sy = (\delta + n + g)k$.]

b. What is the marginal product of capital in the initial steady state?

c. Suppose that public policy raises the saving rate so that the economy reaches the Golden Rule level of capital. What will the marginal product of capital be at the Golden Rule steady state? Compare the marginal product at the Golden Rule steady state to the marginal product in the initial steady state. Explain.

d. What will the capital–output ratio be at the Golden Rule steady state? [Hint: For the Cobb–Douglas production function, the capital–output ratio is related to the marginal product of capital.)

e. What must the saving rate be to reach the Golden Rule steady state?

3. Prove each of the following statements about the steady state with population growth and technological progress.

a. The capital–output ratio is constant.

b. Capital and labor each earn a constant share of an economy's income. [Hint: Recall the definition $MPK = f(k + 1) - f(k)$.]

c. Total capital income and total labor income both grow at the rate of population growth plus the rate of technological progress, $n + g$.

d. The real rental price of capital is constant, and the real wage grows at the rate of technological progress, g. (Hint: The real rental price of capital equals total capital income divided by the capital stock, and the real wage equals total labor income divided by the labor force.)

4. The amount of education the typical person receives varies substantially among countries. Suppose you were to compare a country with a highly educated labor force and a country with a less educated labor force. Assume that education affects only the level of the efficiency of labor. Also assume that the countries are otherwise the same: they have the same saving rate, the same depreciation rate, the same population growth rate, and the same rate of technological progress. Both countries are described by the Solow model and are in their steady states. What would you predict for the following variables?

a. The rate of growth of total income.

b. The level of income per worker.

c. The real rental price of capital.

d. The real wage.

5. This question asks you to analyze in more detail the two-sector endogenous growth model presented in the text.

a. Rewrite the production function for manufactured goods in terms of output per effective worker and capital per effective worker.

b. In this economy, what is break-even investment (the amount of investment needed to keep capital per effective worker constant)?

c. Write down the equation of motion for k, which shows Δk as saving minus break-even investment. Use this equation to draw a graph showing the determination of steady-state k. (Hint: This graph will look much like those we used to analyze the Solow model.)

d. In this economy, what is the steady-state growth rate of output per worker Y/L? How do the saving rate s and the fraction of the labor force in universities u affect this steady-state growth rate?

e. Using your graph, show the impact of an increase in u. (Hint: This change affects both curves.) Describe both the immediate and the steady-state effects.

f. Based on your analysis, is an increase in u an unambiguously good thing for the economy? Explain.

APPENDIX

Accounting for the Sources of Economic Growth

Real GDP in the United States has grown an average of 3 percent per year over the past 40 years. What explains this growth? In Chapter 3 we linked the output of the economy to the factors of production—capital and labor—and to the production technology. Here we develop a technique called *growth accounting* that divides the growth in output into three different sources: increases in capital, increases in labor, and advances in technology. This breakdown provides us with a measure of the rate of technological change.

Increases in the Factors of Production

We first examine how increases in the factors of production contribute to increases in output. To do this, we start by assuming there is no technological change, so the production function relating output Y to capital K and labor L is constant over time:

$$Y = F(K, L).$$

In this case, the amount of output changes only because the amount of capital or labor changes.

Increases in Capital First, consider changes in capital. If the amount of capital increases by ΔK units, by how much does the amount of output increase? To answer this question, we need to recall the definition of the marginal product of capital, MPK:

$$MPK = F(K + 1, L) - F(K, L).$$

The marginal product of capital tells us how much output increases when capital increases by 1 unit. Therefore, when capital increases by ΔK units, output increases by approximately $MPK \times \Delta K$.[10]

For example, suppose that the marginal product of capital is 1/5; that is, an additional unit of capital increases the amount of output produced by one-fifth of a

[10] Note the word "approximately" here. This answer is only an approximation because the marginal product of capital varies: it falls as the amount of capital increases. An exact answer would take into account the fact that each unit of capital has a different marginal product. If the change in K is not too large, however, the approximation of a constant marginal product is very accurate.

unit. If we increase the amount of capital by 10 units, we can compute the amount of additional output as follows:

$$\begin{aligned}\Delta Y &= \text{MPK} \times \Delta K \\ &= 1/5 \frac{\text{Units of Output}}{\text{Unit of Capital}} \times 10 \text{ Units of Capital} \\ &= 2 \text{ Units of Output.}\end{aligned}$$

By increasing capital 10 units, we obtain 2 more units of output. Thus, we use the marginal product of capital to convert changes in capital into changes in output.

Increases in Labor Next, consider changes in labor. If the amount of labor increases by ΔL units, by how much does output increase? We answer this question the same way we answered the question about capital. The marginal product of labor MPL tells us how much output changes when labor increases by 1 unit that is,

$$\text{MPL} = F(K, L + 1) - F(K, L).$$

Therefore, when the amount of labor increases by ΔL units, output increases by approximately $\text{MPL} \times \Delta L$.

For example, suppose that the marginal product of labor is 2; that is, an additional unit of labor increases the amount of output produced by 2 units. If we increase the amount of labor by 10 units, we can compute the amount of additional output as follows:

$$\begin{aligned}\Delta Y &= \text{MPL} \times \Delta L \\ &= 2 \frac{\text{Units of Ouput}}{\text{Unit of Labor}} \times 10 \text{ Units of Labor} \\ &= 20 \text{ Units of Output.}\end{aligned}$$

By increasing labor 10 units, we obtain 20 more units of output. Thus, we use the marginal product of labor to convert changes in labor into changes in output.

Increases in Capital and Labor Finally, let's consider the more realistic case in which both factors of production change. Suppose that the amount of capital increases by ΔK and the amount of labor increases by ΔL. The increase in output then comes from two sources: more capital and more labor. We can divide this increase into the two sources using the marginal products of the two inputs:

$$\Delta Y = (\text{MPK} \times \Delta K) + (\text{MPL} \times \Delta L).$$

The first term in parentheses is the increase in output resulting from the increase in capital, and the second term in parentheses is the increase in output resulting from the increase in labor. This equation shows us how to attribute growth to each factor of production.

We now want to convert this last equation into a form that is easier to interpret and apply to the available data. First, with some algebraic rearrangement, the equation becomes[11]

$$\frac{\Delta Y}{Y} = \left(\frac{MPK \times K}{Y}\right)\frac{\Delta K}{K} + \left(\frac{MPL \times L}{Y}\right)\frac{\Delta L}{L}.$$

This form of the equation relates the growth rate of output, $\Delta Y/Y$, to the growth rate of capital, $\Delta K/K$, and the growth rate of labor, $\Delta L/L$.

Next, we need to find some way to measure the terms in parentheses in the last equation. In Chapter 3 we showed that the marginal product of capital equals its real rental price. Therefore, $MPK \times K$ is the total return to capital, and $(MPK \times K)/Y$ is capital's share of output. Similarly, the marginal product of labor equals the real wage. Therefore, $MPL \times L$ is the total compensation that labor receives, and $(MPL \times L)/Y$ is labor's share of output. Under the assumption that the production function has constant returns to scale, Euler's theorem (which we discussed in Chapter 3) tells us that these two shares sum to 1. In this case, we can write

$$\frac{\Delta Y}{Y} = \alpha\frac{\Delta K}{K} + (1-\alpha)\frac{\Delta L}{L}.$$

where α is capital's share and $(1-\alpha)$ is labor's share.

This last equation gives us a simple formula for showing how changes in inputs lead to changes in output. In particular, we must weight the growth rates of the inputs by the factor shares. As we discussed in the appendix to Chapter 3, capital's share in the United States is about 30 percent, that is, $\alpha = 0.30$. Therefore, a 10-percent increase in the amount of capital $(\Delta K/K = 0.10)$ leads to a 3-percent increase in the amount of output $(\Delta Y/Y = 0.03)$. Similarly, a 10-percent increase in the amount of labor $(\Delta L/L = 0.10)$ leads to a 7-percent increase in the amount of output $(\Delta Y/Y = 0.07)$.

Technological Progress

So far in our analysis of the sources of growth, we have been assuming that the production function does not change over time. In practice, of course, technological progress improves the production function. For any given amount of inputs, we get more output today than we did in the past. We now extend the analysis to allow for technological progress.

[11] Mathematical note: To see that this is equivalent to the previous equation, note that we can multiply both sides of this equation by Y and thereby cancel Y from three places in which it appears. We can cancel the K in the top and bottom of the first term on the right-hand side and the L in the top and bottom of the second term on the right-hand side. These algebraic manipulations turn this equation into the previous one.

We include the effects of the changing technology by writing the production function as

$$Y = AF(K, L),$$

where A is a measure of the current level of technology called **total factor productivity**. Output now increases not only because of increases in capital and labor but also because of increases in total factor productivity. If total factor productivity increases by 1 percent and if the inputs are unchanged, then output increases by 1 percent.

Allowing for a changing technology adds another term to our equation accounting for economic growth:

$$\frac{\Delta Y}{Y} = \alpha\frac{\Delta K}{K} + (1-\alpha)\frac{\Delta L}{L} + \frac{\Delta A}{A}$$

$$\text{Growth in Output} = \text{Contribution of Capital} + \text{Contribution of Labor} + \text{Growth in Total Factor Productivity}.$$

This is the key equation of growth accounting. It identifies and allows us to measure the three sources of growth: changes in the amount of capital, changes in the amount of labor, and changes in total factor productivity.

Because total factor productivity is not observable directly, it is measured indirectly. We have data on the growth in output, capital, and labor; we also have data on capital's share of output. From these data and the growth-accounting equation, we can compute the growth in total factor productivity to make sure that everything adds up:

$$\frac{\Delta A}{A} = \frac{\Delta Y}{Y} - \alpha\frac{\Delta K}{K} - (1-\alpha)\frac{\Delta L}{L}.$$

$\Delta A/A$ is the change in output that cannot be explained by changes in inputs. Thus, the growth in total factor productivity is computed as a **residual**—that is, as the amount of output growth that remains after we have accounted for the determinants of growth that we can measure. Indeed, $\Delta A/A$ is sometimes called the **Solow residual**, after Robert Solow, who first showed how to compute it.[12]

Total factor productivity can change for many reasons. Changes most often arise because of increased knowledge about production methods, and the Solow residual is often used as a measure of technological progress. Yet other

[12] Robert M. Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39 (1957): 312–320. It is natural to ask how growth in labor efficiency E relates to growth in total factor productivity. One can show that $\Delta A/A = (1-\alpha)\Delta E/E$, where α is capital's share. Thus, technological change as measured by growth in the efficiency of labor is proportional to technological change as measured by the Solow residual.

factors, such as education and government regulation, can affect total factor productivity as well. For example, if higher public spending raises the quality of education, then workers may become more productive and output may rise, which implies higher total factor productivity. As another example, if government regulations require firms to purchase capital to reduce pollution or increase worker safety, then the capital stock may rise without any increase in measured output, which implies lower total factor productivity. Total factor productivity captures anything that changes the relation between measured inputs and measured output

The Sources of Growth in the United States

Having learned how to measure the sources of economic growth, we now look at the data. Table 8-3 uses U.S. data to measure the contributions of the three sources of growth between 1950 and 1999.

This table shows that real GDP has grown an average of 3.6 percent per year since 1950. Of this 3.6 percent, 1.2 percent is attributable to increases in the capital stock, 1.3 percent to increases in the labor input, and 1.1 percent to increases in total factor productivity. These data show that increases in capital, labor, and productivity have contributed almost equally to economic growth in the United States.

Table 8-3 also shows that the growth in total factor productivity slowed substantially around 1970. In a previous case study in this chapter, we discussed some hypotheses to explain this productivity slowdown.

table 8-3

Accounting for Economic Growth in the United States

			SOURCE OF GROWTH				
Years	Output Growth $\Delta Y/Y$	=	Capital $\alpha \Delta K/K$	+	Labor $(1-\alpha)\Delta K/K$	+	Total Factor Productivity $\Delta A/A$
			(average percentage increase per year)				
1950–1999	3.6		1.2		1.3		1.1
1950–1960	3.3		1.0		1.0		1.3
1960–1970	4.4		1.4		1.2		1.8
1970–1980	3.6		1.4		1.2		1.0
1980–1990	3.4		1.2		1.6		0.6
1990–1999	3.7		1.2		1.6		0.9

Source: U.S. Department of Commerce, U.S. Department of Labor, and the author's calculations. The parameter α is set to equal 0.3.

CASE STUDY

Growth in the East Asian Tigers

Perhaps the most spectacular growth experiences in recent history have been those of the "Tigers" of East Asia: Hong Kong, Singapore, South Korea, and Taiwan. From 1966 to 1990, while real income per person was growing about 2 percent per year in the United States, it grew more than 7 percent per year in each of these countries. In the course of a single generation, real income per person increased fivefold, moving the Tigers from among the world's poorest countries to among the richest. (In the late 1990s, a period of pronounced financial turmoil tarnished the reputation of some of these economies. But this short-run problem, which we examine in a case study in Chapter 12, does not come close to reversing the spectacular long-run growth performance that the Asian Tigers have experienced.)

What accounts for these growth miracles? Some commentators have argued that the success of these four countries is hard to reconcile with basic growth theory, such as the Solow growth model, which has technology growing at a constant, exogenous rate. They have suggested that these countries' rapid growth is explained by their ability to imitate foreign technologies. By adopting technology developed abroad, the argument goes, these countries managed to improve their production functions substantially in a relatively short period of time. If this argument is correct, these countries should have experienced unusually rapid growth in total factor productivity.

One recent study shed light on this issue by examining in detail the data from these four countries. The study found that their exceptional growth can be traced to large increases in measured factor inputs: increases in labor-force participation, increases in the capital stock, and increases in educational attainment. In South Korea, for example, the investment–GDP ratio rose from about 5 percent in the 1950s to about 30 percent in the 1980s; the percentage of the working population with at least a high-school education went from 26 percent in 1966 to 75 percent in 1991.

Once we account for growth in labor, capital, and human capital, little of the growth in output is left to explain. None of these four countries experienced unusually rapid growth in total factor productivity. Indeed, the average growth in total factor productivity in the East Asian Tigers was almost exactly the same as in the United States. Thus, although these countries' rapid growth has been truly impressive, it is easy to explain using the tools of basic growth theory.[13]

[13] Alwyn Young, "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience," *Quarterly Journal of Economics* 110 (August 1995): 641–680.

MORE PROBLEMS AND APPLICATIONS

1. In the economy of Solovia, the owners of capital get two-thirds of national income, and the workers receive one-third.
   a. The men of Solovia stay at home performing household chores, while the women work in factories. If some of the men started working outside the home so that the labor force increased by 5 percent, what would happen to the measured output of the economy? Does labor productivity—defined as output per worker—increase, decrease, or stay the same? Does total factor productivity increase, decrease, or stay the same?
   b. In year 1, the capital stock was 6, the labor input was 3, and output was 12. In year 2, the capital stock was 7, the labor input was 4, and output was 14. What happened to total factor productivity between the two years?
2. Labor productivity is defined as Y/L, the amount of output divided by the amount of labor input. Start with the growth-accounting equation and show that the growth in labor productivity depends on growth in total factor productivity and growth in the capital–labor ratio. In particular, show that

$$\frac{\Delta(Y/L)}{Y/L} = \frac{\Delta A}{A} + \alpha \frac{\Delta(K/L)}{K/L}.$$

(Hint: You may find the following mathematical trick helpful. If $z = wx$, then the growth rate of z is approximately the growth rate of w plus the growth rate of x. That is, $\Delta z/z \approx \Delta w/w + \Delta x/x$.)

3. Suppose an economy described by the Solow model is in a steady state with population growth n of 1.0 percent per year and technological progress g of 2.0 percent per year. Total output and total capital grow at 3.0 percent per year. Suppose further that the capital share of output is 0.3. If you used the growth-accounting equation to divide output growth into three sources—capital, labor, and total factor productivity—how much would you attribute to each source? Compare your results to the figures we found for the United States in Table 8-3.

part IV

Business Cycle Theory: The Economy in the Short Run

CHAPTER 9 NINE

Introduction to Economic Fluctuations

> The modern world regards business cycles much as the ancient Egyptians regarded the overflowing of the Nile. The phenomenon recurs at intervals, it is of great importance to everyone, and natural causes of it are not in sight.
>
> — John Bates Clark, 1898

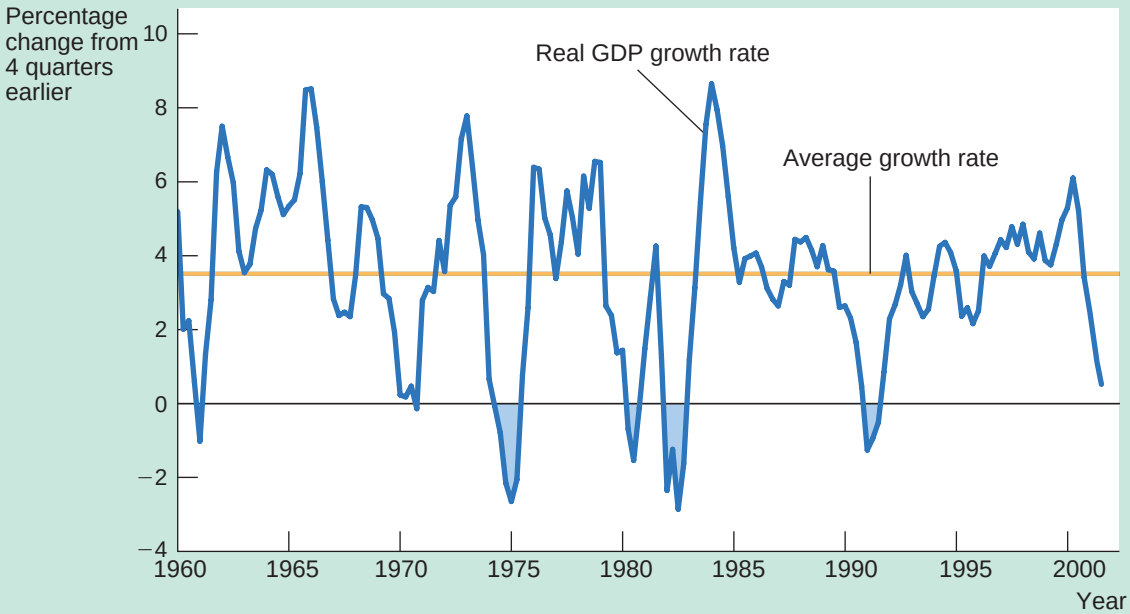
Economic fluctuations present a recurring problem for economists and policymakers. This problem is illustrated in Figure 9-1, which shows growth in real GDP for the U.S. economy. As you can see, although the economy experiences long-run growth that averages about 3.5 percent per year, this growth is not at all steady. Recessions—periods of falling incomes and rising unemployment—are frequent. In the recession of 1990, for instance, real GDP fell 2.2 percent from its peak to its trough, and the unemployment rate rose to 7.7 percent. During recessions, not only are more people unemployed, but those who are employed have shorter workweeks, as more workers have to accept part-time jobs and fewer workers have the opportunity to work overtime. When recessions end and the economy enters a boom, these effects work in reverse: incomes rise, unemployment falls, and workweeks expand.

Economists call these short-run fluctuations in output and employment the *business cycle*. Although this term suggests that economic fluctuations are regular and predictable, they are not. Recessions are as irregular as they are common. Sometimes they are close together, such as the recessions of 1980 and 1982. Sometimes they are far apart, such as the recessions of 1982 and 1990.

In Parts II and III of this book, we developed theories to explain how the economy behaves in the long run. Those theories were based on the classical dichotomy—the premise that real variables, such as output and employment, are not affected by what happens to nominal variables, such as the money supply and the price level. Although classical theories are useful for explaining long-run trends, including the economic growth we observe from decade to decade, most economists believe that the classical dichotomy does not hold in the short run and, therefore, that classical theories cannot explain year-to-year fluctuations in output and employment.

Here, in Part IV, we see how economists explain these short-run fluctuations. This chapter begins our analysis by discussing the key differences between the long run and the short run and by introducing the model of aggregate supply

figure 9-1



Real GDP Growth in the United States Growth in real GDP averages about 3.5 percent per year, as indicated by the orange line, but there are substantial fluctuations around this average. Recessions are periods when the production of goods and services is declining, represented here by negative growth in real GDP.

Source: U.S. Department of Commerce.

and aggregate demand. With this model we can show how shocks to the economy lead to short-run fluctuations in output and employment.

Just as Egypt now controls the flooding of the Nile Valley with the Aswan Dam, modern society tries to control the business cycle with appropriate economic policies. The model we develop over the next several chapters shows how monetary and fiscal policies influence the business cycle. We will see that these policies can potentially stabilize the economy or, if poorly conducted, make the problem of economic instability even worse.

9-1 Time Horizons in Macroeconomics

Before we start building a model of short-run economic fluctuations, let's step back and ask a fundamental question: Why do economists need different models for different time horizons? Why can't we stop the course here and be content with the classical models developed in Chapters 3 through 8? The answer, as this book has consistently reminded its reader, is that classical macroeconomic theory applies to the long run but not to the short run. But why is this so?

How the Short Run and Long Run Differ

Most macroeconomists believe that the key difference between the short run and the long run is the behavior of prices. In the long run, prices are flexible and can respond to changes in supply or demand. In the short run, many prices are "sticky" at some predetermined level. Because prices behave differently in the short run than in the long run, economic policies have different effects over different time horizons.

To see how the short run and the long run differ, consider the effects of a change in monetary policy. Suppose that the Federal Reserve suddenly reduced the money supply by 5 percent. According to the classical model, which almost all economists agree describes the economy in the long run, the money supply affects nominal variables—variables measured in terms of money—but not real variables. In the long run, a 5-percent reduction in the money supply lowers all prices (including nominal wages) by 5 percent whereas all real variables remain the same. Thus, in the long run, changes in the money supply do not cause fluctuations in output or employment.

In the short run, however, many prices do not respond to changes in monetary policy. A reduction in the money supply does not immediately cause all firms to cut the wages they pay, all stores to change the price tags on their goods, all mail-order firms to issue new catalogs, and all restaurants to print new menus. Instead, there is little immediate change in many prices; that is, many prices are sticky. This short-run price stickiness implies that the short-run impact of a change in the money supply is not the same as the long-run impact.

A model of economic fluctuations must take into account this short-run price stickiness. We will see that the failure of prices to adjust quickly and completely means that, in the short run, output and employment must do some of the adjusting instead. In other words, during the time horizon over which prices are sticky, the classical dichotomy no longer holds: nominal variables can influence real variables, and the economy can deviate from the equilibrium predicted by the classical model.

CASE STUDY

The Puzzle of Sticky Magazine Prices

How sticky are prices? The answer to this question depends on what price we consider. Some commodities, such as wheat, soybeans, and pork bellies, are traded on organized exchanges, and their prices change every minute. No one would call these prices sticky. Yet the prices of most goods and services change much less frequently. One survey found that 39 percent of firms change their prices once a year, and another 10 percent change their prices less than once a year.[1]

The reasons for price stickiness are not always apparent. Consider, for example, the market for magazines. A study has documented that magazines change their newsstand prices very infrequently. The typical magazine allows inflation to erode

[1] Alan S. Blinder, "On Sticky Prices: Academic Theories Meet the Real World," in N. G. Mankiw, ed., *Monetary Policy* (Chicago: University of Chicago Press, 1994), 117–154. A case study in Chapter 19 discusses this survey in more detail.

its real price by about 25 percent before it raises its nominal price. That is, when inflation is 4 percent per year, the typical magazine changes its price about every six years.[2]

Why do magazines keep their prices the same for so long? Economists do not have a definitive answer. The question is puzzling because it would seem that for magazines, the cost of a price change is small. To change prices, a mail-order firm must issue a new catalog and a restaurant must print a new menu, but a magazine publisher can simply print a new price on the cover of the next issue. Perhaps the cost to the publisher of charging the wrong price is also small. Or maybe customers would find it inconvenient if the price of their favorite magazine changed every month.

As the magazine example shows, explaining at the microeconomic level why prices are sticky is sometimes hard. The cause of price stickiness is, therefore, an active area of research, which we discuss more fully in Chapter 19. In this chapter, however, we simply assume that prices are sticky so we can start developing the link between sticky prices and the business cycle. Although not yet fully explained, short-run price stickiness is widely believed to be crucial for understanding short-run economic fluctuations.

The Model of Aggregate Supply and Aggregate Demand

How does introducing sticky prices change our view of how the economy works? We can answer this question by considering economists' two favorite words—supply and demand.

In classical macroeconomic theory, the amount of output depends on the economy's ability to supply goods and services, which in turn depends on the supplies of capital and labor and on the available production technology. This is the essence of the basic classical model in Chapter 3, as well as of the Solow growth model in Chapters 7 and 8. Flexible prices are a crucial assumption of classical theory. The theory posits, sometimes implicitly, that prices adjust to ensure that the quantity of output demanded equals the quantity supplied.

The economy works quite differently when prices are sticky. In this case, as we will see, output also depends on the demand for goods and services. Demand, in turn, is influenced by monetary policy, fiscal policy, and various other factors. Because monetary and fiscal policy can influence the economy's output over the time horizon when prices are sticky, price stickiness provides a rationale for why these policies may be useful in stabilizing the economy in the short run.

In the rest of this chapter, we develop a model that makes these ideas more precise. The model of supply and demand, which we used in Chapter 1 to discuss the market for pizza, offers some of the most fundamental insights in economics. This model shows how the supply and demand for any good jointly determine the

[2] Stephen G. Cecchetti, "The Frequency of Price Adjustment: A Study of the Newsstand Prices of Magazines," *Journal of Econometrics* 31 (1986): 255–274.

goods price and the quantity sold, and how shifts in supply and demand affect the price and quantity. In the rest of this chapter, we introduce the economy-size version of this model—the model of aggregate supply and aggregate demand. This macroeconomic model allows us to study how the aggregate price level and the quantity of aggregate output are determined. It also provides a way to contrast how the economy behaves in the long run and how it behaves in the short run.

Although the model of aggregate supply and aggregate demand resembles the model of supply and demand for a single good, the analogy is not exact. The model of supply and demand for a single good considers only one good within a large economy. By contrast, as we will see in the coming chapters, the model of aggregate supply and aggregate demand is a sophisticated model that incorporates the interactions among many markets.

9-2 Aggregate Demand

Aggregate demand (AD) is the relationship between the quantity of output demanded and the aggregate price level. In other words, the aggregate demand curve tells us the quantity of goods and services people want to buy at any given level of prices. We examine the theory of aggregate demand in detail in Chapters 10 through 12. Here we use the quantity theory of money to provide a simple, although incomplete, derivation of the aggregate demand curve.

The Quantity Equation as Aggregate Demand

Recall from Chapter 4 that the quantity theory says that

$$MV = PY,$$

where M is the money supply, V is the velocity of money, P is the price level, and Y is the amount of output. If the velocity of money is constant, then this equation states that the money supply determines the nominal value of output, which in turn is the product of the price level and the amount of output.

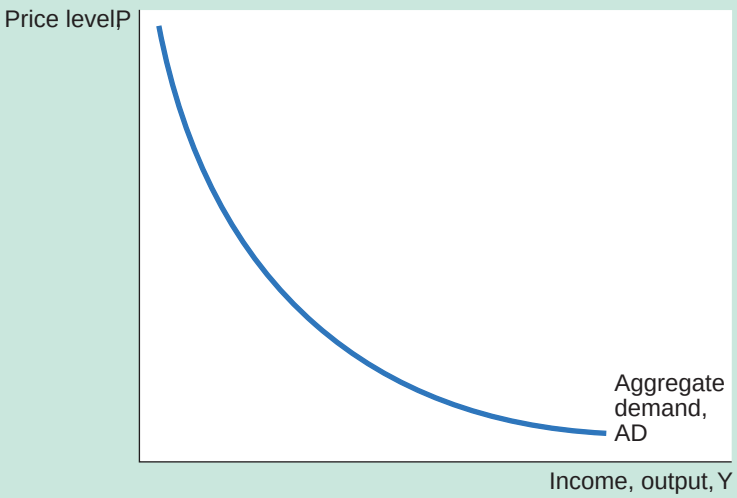
You might recall that the quantity equation can be rewritten in terms of the supply and demand for real money balances:

$$M/P = (M/P)^d = kY,$$

where $k = 1/V$ is a parameter determining how much money people want to hold for every dollar of income. In this form, the quantity equation states that the supply of real money balances M/P equals the demand $(M/P)^d$ and that the demand is proportional to output Y. The velocity of money V is the "flip side" of the money demand parameter k.

For any fixed money supply and velocity, the quantity equation yields a negative relationship between the price level P and output Y. Figure 9-2 graphs the combinations of P and Y that satisfy the quantity equation holding M and V constant. This downward-sloping curve is called the aggregate demand curve.

figure 9-2



The Aggregate Demand Curve
The aggregate demand curve AD shows the relationship between the price level P and the quantity of goods and services demanded Y. It is drawn for a given value of the money supply M. The aggregate demand curve slopes downward: the higher the price level P, the lower the level of real balances M/P, and therefore the lower the quantity of goods and services demanded Y.

Why the Aggregate Demand Curve Slopes Downward

As a strictly mathematical matter, the quantity equation explains the downward slope of the aggregate demand curve very simply. The money supply M and the velocity of money V determine the nominal value of output PY. Once PY is fixed, if P goes up, Y must go down.

What is the economics that lies behind this mathematical relationship? For a complete answer, we have to wait for a couple of chapters. For now, however, consider the following logic: Because we have assumed the velocity of money is fixed, the money supply determines the dollar value of all transactions in the economy. (This conclusion should be familiar from Chapter 4.) If the price level rises, each transaction requires more dollars, so the number of transactions and thus the quantity of goods and services purchased must fall.

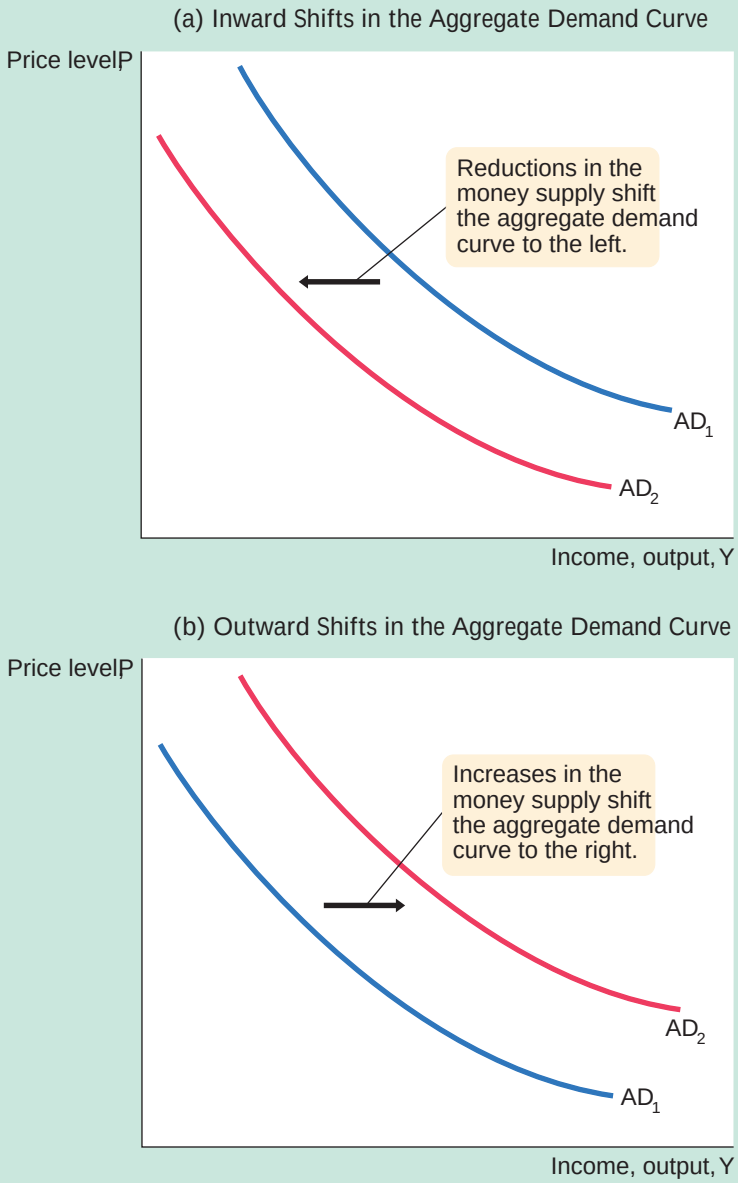
We can also explain the downward slope of the aggregate demand curve by thinking about the supply and demand for real money balances. If output is higher, people engage in more transactions and need higher real balances M/P. For a fixed money supply M, higher real balances imply a lower price level. Conversely, if the price level is lower, real money balances are higher; the higher level of real balances allows a greater volume of transactions, which means a greater quantity of output is demanded.

Shifts in the Aggregate Demand Curve

The aggregate demand curve is drawn for a fixed value of the money supply. In other words, it tells us the possible combinations of P and Y for a given value of M. If the Fed changes the money supply, then the possible combinations of P and Y change, which means the aggregate demand curve shifts.

For example, consider what happens if the Fed reduces the money supply. The quantity equation, $MV = PY$, tells us that the reduction in the money supply leads to a proportionate reduction in the nominal value of output PY. For any given price level, the amount of output is lower, and for any given amount of output, the price level is lower. As in Figure 9-3(a), the aggregate demand curve relating P and Y shifts inward.

figure 9-3



Shifts in the Aggregate Demand Curve Changes in the money supply shift the aggregate demand curve. In panel (a), a decrease in the money supply (M) reduces the nominal value of output (PY). For any given price level (P), output (Y) is lower. Thus, a decrease in the money supply shifts the aggregate demand curve inward from AD_1 to AD_2. In panel (b), an increase in the money supply (M) raises the nominal value of output (PY). For any given price level (P), output (Y) is higher. Thus, an increase in the money supply shifts the aggregate demand curve outward from AD_1 to AD_2.

The opposite occurs if the Fed increases the money supply. The quantity equation tells us that an increase in M leads to an increase in PY. For any given price level, the amount of output is higher, and for any given amount of output, the price level is higher. As shown in Figure 9-3(b), the aggregate demand curve shifts outward.

Fluctuations in the money supply are not the only source of fluctuations in aggregate demand. Even if the money supply is held constant, the aggregate demand curve shifts if some event causes a change in the velocity of money. Over the next three chapters, we consider many possible reasons for shifts in the aggregate demand curve.

9-3 Aggregate Supply

By itself, the aggregate demand curve does not tell us the price level or the amount of output; it merely gives a relationship between these two variables. To accompany the aggregate demand curve, we need another relationship between P and Y that crosses the aggregate demand curve—an aggregate supply curve. The aggregate demand and aggregate supply curves together pin down the economy's price level and quantity of output.

Aggregate supply (AS) is the relationship between the quantity of goods and services supplied and the price level. Because firms that supply goods and services have flexible prices in the long run but sticky prices in the short run, the aggregate supply relationship depends on the time horizon. We need to discuss two different aggregate supply curves: the long-run aggregate supply curve LRAS and the short-run aggregate supply curve SRAS. We also need to discuss how the economy makes the transition from the short run to the long run.

The Long Run: The Vertical Aggregate Supply Curve

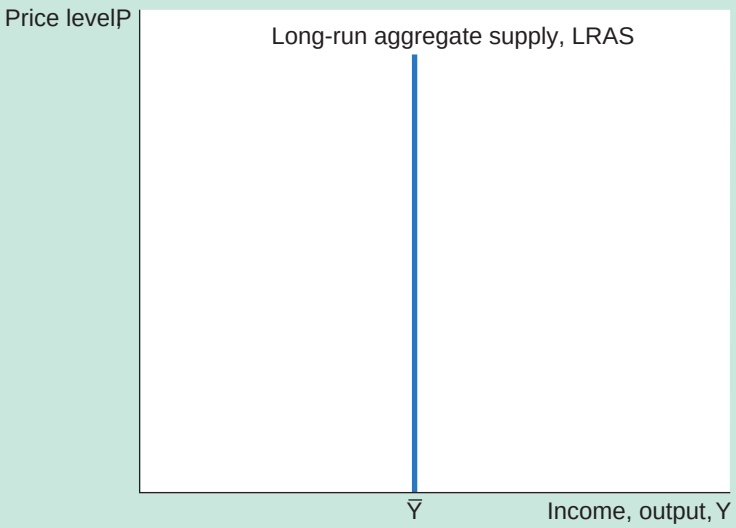
Because the classical model describes how the economy behaves in the long run, we derive the long-run aggregate supply curve from the classical model. Recall from Chapter 3 that the amount of output produced depends on the fixed amounts of capital and labor and on the available technology. To show this, we write

$$\begin{aligned} Y &= F(\bar{K}, \bar{L}) \\ &= \bar{Y}. \end{aligned}$$

According to the classical model, output does not depend on the price level. To show that output is the same for all price levels, we draw a vertical aggregate supply curve, as in Figure 9-4. The intersection of the aggregate demand curve with this vertical aggregate supply curve determines the price level.

If the aggregate supply curve is vertical, then changes in aggregate demand affect prices but not output. For example, if the money supply falls, the aggregate

figure 9-4

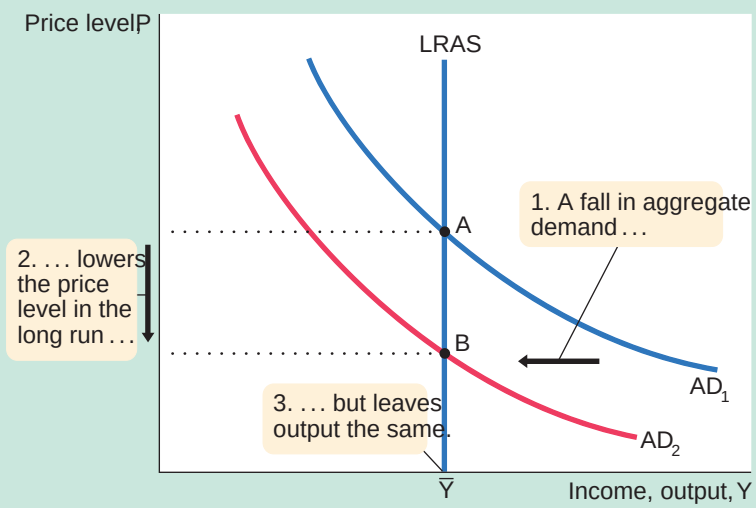


The Long-Run Aggregate Supply Curve In the long run, the level of output is determined by the amounts of capital and labor and by the available technology; it does not depend on the price level. The long-run aggregate supply curve $LRAS$ is vertical.

demand curve shifts downward, as in Figure 9-5. The economy moves from the old intersection of aggregate supply and aggregate demand, point A, to the new intersection, point B. The shift in aggregate demand affects only prices.

The vertical aggregate supply curve reflects the classical dichotomy, because it implies that the level of output is independent of the money supply. This long-run level of output, $\bar{Y}$, is called the *full-employment* or *natural* level of output. It is the level of output at which the economy's resources are fully employed or, more realistically, at which unemployment is at its natural rate.

figure 9-5



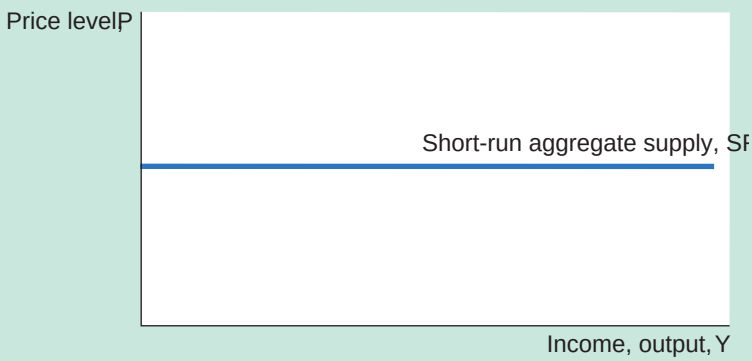
Shifts in Aggregate Demand in the Long Run A reduction in the money supply shifts the aggregate demand curve downward from AD_1 to AD_2. The equilibrium for the economy moves from point A to point B. Since the aggregate supply curve is vertical in the long run, the reduction in aggregate demand affects the price level but not the level of output.

The Short Run: The Horizontal Aggregate Supply Curve

The classical model and the vertical aggregate supply curve apply only in the long run. In the short run, some prices are sticky and, therefore, do not adjust to changes in demand. Because of this price stickiness, the short-run aggregate supply curve is not vertical.

As an extreme example, suppose that all firms have issued price catalogs and that it is costly for them to issue new ones. Thus, all prices are stuck at predetermined levels. At these prices, firms are willing to sell as much as their customers are willing to buy, and they hire just enough labor to produce the amount demanded. Because the price level is fixed, we represent this situation in Figure 9-6 with a horizontal aggregate supply curve.

figure 9-6



The Short-Run Aggregate Supply Curve In this extreme example, all prices are fixed in the short run. Therefore, the short-run aggregate supply curve, SRAS, is horizontal.

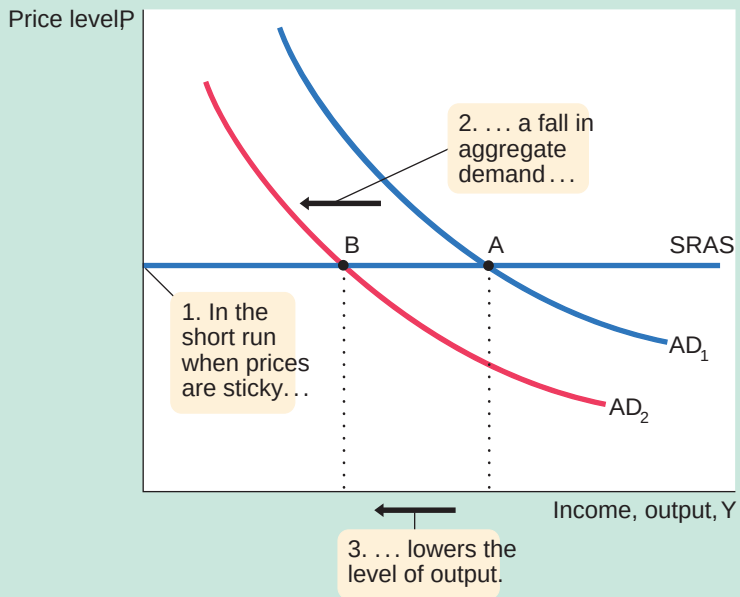
The short-run equilibrium of the economy is the intersection of the aggregate demand curve and this horizontal short-run aggregate supply curve. In this case, changes in aggregate demand do affect the level of output. For example, if the Fed suddenly reduces the money supply, the aggregate demand curve shifts inward, as in Figure 9-7. The economy moves from the old intersection of aggregate demand and aggregate supply, point A, to the new intersection, point B. The movement from point A to point B represents a decline in output at a fixed price level.

Thus, a fall in aggregate demand reduces output in the short run because prices do not adjust instantly. After the sudden fall in aggregate demand, firms are stuck with prices that are too high. With demand low and prices high, firms sell less of their product, so they reduce production and lay off workers. The economy experiences a recession.

From the Short Run to the Long Run

We can summarize our analysis so far as follows: Over long periods of time, prices are flexible, the aggregate supply curve is vertical, and changes in aggregate demand affect the price level but not output. Over short periods of time, prices are sticky, the aggregate supply curve is flat, and changes in aggregate demand do affect the output of goods and services.

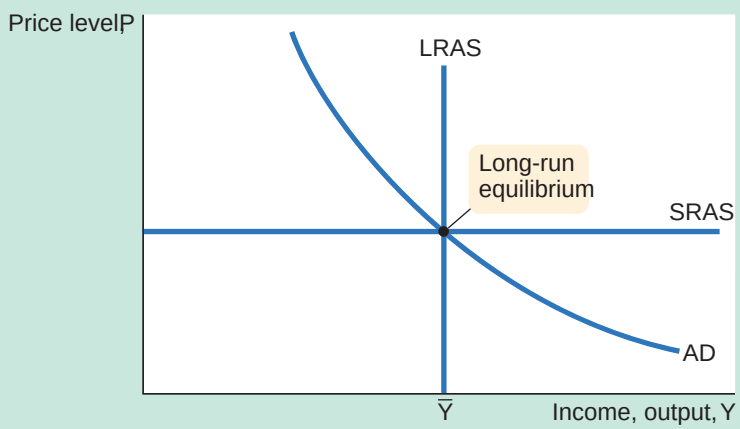
figure 9-7



Shifts in Aggregate Demand in the Short Run A reduction in the money supply shifts the aggregate demand curve downward from AD_1 to AD_2. The equilibrium for the economy moves from point A to point B. Since the aggregate supply curve is horizontal in the short run, the reduction in aggregate demand reduces the level of output.

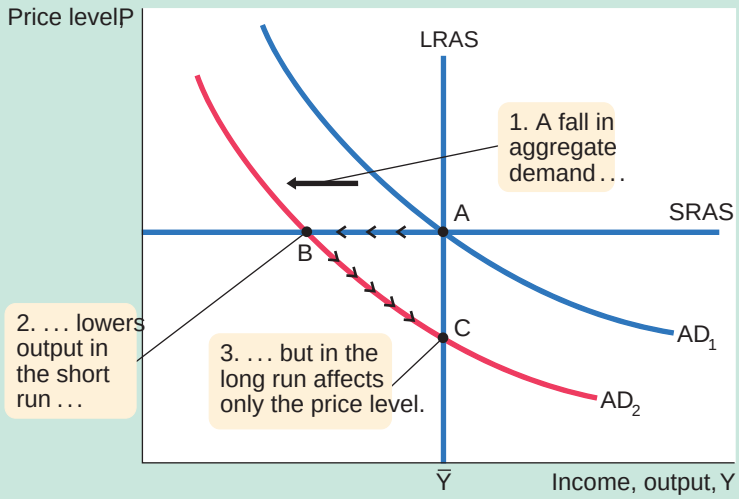
How does the economy make the transition from the short run to the long run? Let's trace the effects over time of a fall in aggregate demand. Suppose that the economy is initially in long-run equilibrium, as shown in Figure 9-8. In this figure, there are three curves: the aggregate demand curve, the long-run aggregate supply curve, and the short-run aggregate supply curve. The long-run equilibrium is the point at which aggregate demand crosses the long-run aggregate supply curve. Prices have adjusted to reach this equilibrium. Therefore, when the

figure 9-8



Long-Run Equilibrium In the long run, the economy finds itself at the intersection of the long-run aggregate supply curve and the aggregate demand curve. Because prices have adjusted to this level, the short-run aggregate supply curve crosses this point as well.

figure 9-9



A Reduction in Aggregate Demand The economy begins in long-run equilibrium at point A. A reduction in aggregate demand, perhaps caused by a decrease in the money supply, moves the economy from point A to point B, where output is below its natural level. As prices fall, the economy gradually recovers from the recession, moving from point B to point C.

economy is in its long-run equilibrium, the short-run aggregate supply curve must cross this point as well.

Now suppose that the Fed reduces the money supply and the aggregate demand curve shifts downward, as in Figure 9-9. In the short run, prices are sticky, so the economy moves from point A to point B. Output and employment fall below their natural levels, which means the economy is in a recession. Over time, in response to the low demand, wages and prices fall. The gradual reduction in the price level moves the economy downward along the aggregate demand curve to point C, which is the new long-run equilibrium. In the new long-run equilibrium (point C), output and employment are back to their natural levels, but prices are lower than in the old long-run equilibrium (point A). Thus, a shift in aggregate demand affects output in the short run, but this effect dissipates over time as firms adjust their prices.

CASE STUDY

Gold, Greenbacks, and the Contraction of the 1870s

The aftermath of the Civil War in the United States provides a vivid example of how contractionary monetary policy affects the economy. Before the war, the United States was on a gold standard. Paper dollars were readily convertible into gold. Under this policy, the quantity of gold determined the money supply and the price level.

In 1862, after the Civil War broke out, the Treasury announced that it would no longer redeem dollars for gold. In essence, this act replaced the gold standard with a system of fiat money. Over the next few years, the government printed large quantities of paper currency called greenbacks for their color—and used

the seigniorage to finance wartime expenditure. Because of this increase in the money supply, the price level approximately doubled during the war.

When the war was over, much political debate centered on the question of whether to return to the gold standard. The Greenback party was formed with the primary goal of maintaining the system of fiat money. Eventually, however, the Greenback party lost the debate. Policymakers decided to retire the greenbacks over time in order to reinstate the gold standard at the rate of exchange between dollars and gold that had prevailed before the war. Their goal was to return the value of the dollar to its former level.

Returning to the gold standard in this way required reversing the wartime rise in prices, which meant aggregate demand had to fall. (To be more precise, the growth in aggregate demand needed to fall short of the growth in the natural rate of output.) As the price level fell, the economy experienced a recession from 1873 to 1879, the longest on record. By 1879, the price level was back to its level before the war, and the gold standard was reinstated.

9-4 Stabilization Policy

Fluctuations in the economy as a whole come from changes in aggregate supply or aggregate demand. Economists call exogenous changes in these curves *shocks* to the economy. A shock that shifts the aggregate demand curve is called a *demand shock*, and a shock that shifts the aggregate supply curve is called a *supply shock*. These shocks disrupt economic well-being by pushing output and employment away from their natural rates. One goal of the model of aggregate supply and aggregate demand is to show how shocks cause economic fluctuations.

Another goal of the model is to evaluate how macroeconomic policy can respond to these shocks. Economists use the term *stabilization policy* to refer to policy actions aimed at reducing the severity of short-run economic fluctuations. Because output and employment fluctuate around their long-run natural rates, stabilization policy dampens the business cycle by keeping output and employment as close to their natural rates as possible.

In the coming chapters, we examine in detail how stabilization policy works and what practical problems arise in its use. Here we begin our analysis of stabilization policy by examining how monetary policy might respond to shocks. Monetary policy is an important component of stabilization policy because, as we have seen, the money supply has a powerful impact on aggregate demand.

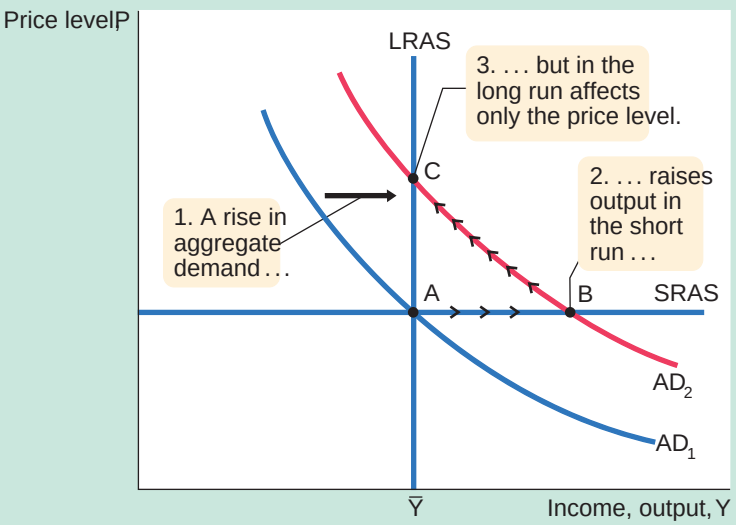
Shocks to Aggregate Demand

Consider an example of a demand shock: the introduction and expanded availability of credit cards. Because credit cards are often a more convenient way to make purchases than using cash, they reduce the quantity of money that people choose to hold. This reduction in money demand is equivalent to an increase in

the velocity of money. When each person holds less money, the money demand parameter k falls. This means that each dollar of money moves from hand to hand more quickly, so velocity V (= 1/k) rises.

If the money supply is held constant, the increase in velocity causes nominal spending to rise and the aggregate demand curve to shift outward, as in Figure 9-10. In the short run, the increase in demand raises the output of the economy and it causes an economic boom. At the old prices, firms now sell more output. Therefore, they hire more workers, ask their existing workers to work longer hours, and make greater use of their factories and equipment.

figure 9-10



An Increase in Aggregate Demand The economy begins in long-run equilibrium at point A. An increase in aggregate demand, due to an increase in the velocity of money, moves the economy from point A to point B, where output is above its natural level. As prices rise, output gradually returns to its natural rate, and the economy moves from point B to point C.

Over time, the high level of aggregate demand pulls up wages and prices. As the price level rises, the quantity of output demanded declines, and the economy gradually approaches the natural rate of production. But during the transition to the higher price level, the economy's output is higher than the natural rate.

What can the Fed do to dampen this boom and keep output closer to the natural rate? The Fed might reduce the money supply to offset the increase in velocity. Offsetting the change in velocity would stabilize aggregate demand. Thus, the Fed can reduce or even eliminate the impact of demand shocks on output and employment if it can skillfully control the money supply. Whether the Fed in fact has the necessary skill is a more difficult question, which we take up in Chapter 14.

Shocks to Aggregate Supply

Shocks to aggregate supply, as well as shocks to aggregate demand, can cause economic fluctuations. A supply shock is a shock to the economy that alters the cost of producing goods and services and, as a result, the prices firms charge.

Because supply shocks have a direct impact on the price level, they are sometimes called *price shocks*. Here are some examples:

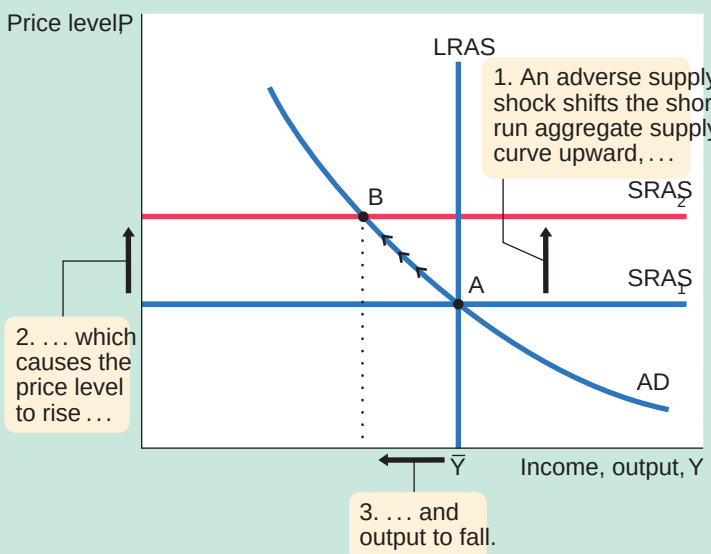
- A drought that destroys crops. The reduction in food supply pushes up food prices.
- A new environmental protection law that requires firms to reduce their emissions of pollutants. Firms pass on the added costs to customers in the form of higher prices.
- An increase in union aggressiveness. This pushes up wages and the prices of the goods produced by union workers.
- The organization of an international oil cartel. By curtailing competition, the major oil producers can raise the world price of oil.

All these events are *adverse supply shocks*, which means they push costs and prices upward. A favorable supply shock, such as the breakup of an international oil cartel, reduces costs and prices.

Figure 9-11 shows how an adverse supply shock affects the economy. The short-run aggregate supply curve shifts upward. (The supply shock may also lower the natural level of output and thus shift the long-run aggregate supply curve to the left, but we ignore that effect here.) If aggregate demand is held constant, the economy moves from point A to point B: the price level rises and the amount of output falls below the natural rate. An experience like this is called *stagflation*, because it combines stagnation (falling output) with inflation (rising prices).

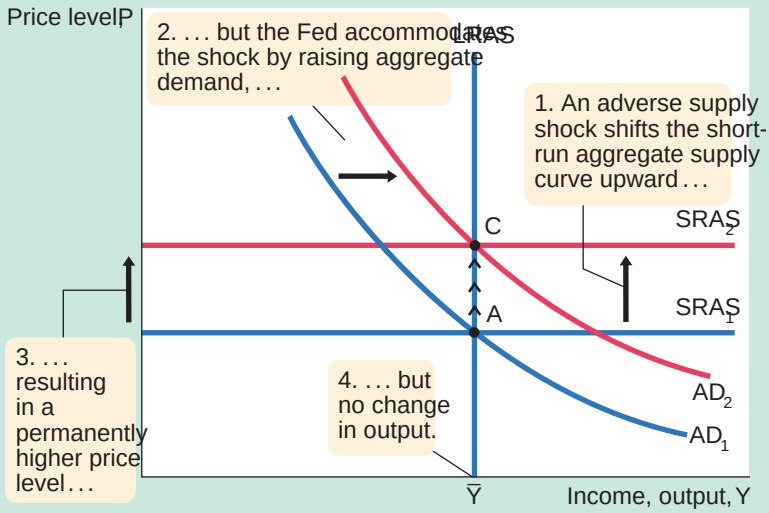
Faced with an adverse supply shock, a policymaker controlling aggregate demand, such as the Fed, has a difficult choice between two options. The first option, implicit in Figure 9-11, is to hold aggregate demand constant. In this case, output and employment are lower than the natural rate. Eventually, prices

figure 9-11



An Adverse Supply Shock An adverse supply shock pushes up costs and thus prices. If aggregate demand is held constant, the economy moves from point A to point B, leading to stagflation—a combination of increasing prices and falling output. Eventually, as prices fall, the economy returns to the natural rate, point A.

figure 9-12



Accommodating an Adverse Supply Shock In response to an adverse supply shock, the Fed can increase aggregate demand to prevent a reduction in output. The economy moves from point A to point C. The cost of this policy is a permanently higher level of prices.

will fall to restore full employment at the old price level (point A). But the cost of this adjustment process is a painful recession.

The second option, illustrated in Figure 9-12, is to expand aggregate demand to bring the economy toward the natural rate more quickly. If the increase in aggregate demand coincides with the shock to aggregate supply, the economy goes immediately from point A to point C. In this case, the Fed is said to accommodate the supply shock. The drawback of this option, of course, is that the price level is permanently higher. There is no way to adjust aggregate demand to maintain full employment and keep the price level stable.

CASE STUDY

How OPEC Helped Cause Stagflation in the 1970s and Euphoria in the 1980s

The most disruptive supply shocks in recent history were caused by OPEC, the Organization of Petroleum Exporting Countries. In the early 1970s, OPEC coordinated reduction in the supply of oil nearly doubled the world price. This increase in oil prices caused stagflation in most industrial countries. These statistics show what happened in the United States:

Year	Change in Oil Prices	Inflation Rate (CPI)	Unemployment Rate
1973	11.0%	6.2%	4.9%
1974	68.0	11.0	5.6
1975	16.0	9.1	8.5
1976	3.3	5.8	7.7
1977	8.1	6.5	7.1

The 68-percent increase in the price of oil in 1974 was an adverse supply shock of major proportions. As one would have expected, it led to both higher inflation and higher unemployment.

A few years later, when the world economy had nearly recovered from the first OPEC recession, almost the same thing happened again. OPEC raised oil prices, causing further stagflation. Here are the statistics for the United States:

Year	Change in Oil Prices	Inflation Rate (CPI)	Unemployment Rate
1978	9.4%	7.7%	6.1%
1979	25.4	11.3	5.8
1980	47.8	13.5	7.0
1981	44.4	10.3	7.5
1982	–8.7	6.1	9.5

The increases in oil prices in 1979, 1980, and 1981 again led to double-digit inflation and higher unemployment.

In the mid-1980s, political turmoil among the Arab countries weakened OPEC's ability to restrain supplies of oil. Oil prices fell, reversing the trend of the 1970s and the early 1980s. Here's what happened:

Year	Change in Oil Prices	Inflation Rate (CPI)	Unemployment Rate
1983	–7.1%	3.2%	9.5%
1984	–1.7	4.3	7.4
1985	–7.5	3.6	7.1
1986	–44.5	1.9	6.9
1987	18.3	3.6	6.1

In 1986 oil prices fell by nearly half. This favorable supply shock led to one of the lowest inflation rates experienced in recent U.S. history and to falling unemployment.

More recently, OPEC has not been a major cause of economic fluctuations. This is in part because OPEC has been less successful at raising the price of oil. Although world oil prices have fluctuated, the changes have not been as large as those experienced during the 1970s, and the real price of oil has never returned to the peaks reached in the early 1980s. Moreover, conservation efforts and technological changes have made the economy less susceptible to oil shocks. The amount of oil consumed per unit of real GDP has fallen about 40 percent over the past three decades.

But we should not be too sanguine. The experiences of the 1970s and 1980s could always be repeated. Events in the Middle East are a potential source of shocks to economies around the world.[3]

[3] Some economists have suggested that changes in oil prices played a major role in economic fluctuations even before the 1970s. See James D. Hamilton, "Oil and the Macroeconomy Since World War II," Journal of Political Economy 91 (April 1983): 228–248.

9-5 Conclusion

This chapter introduced a framework to study economic fluctuations: the model of aggregate supply and aggregate demand. The model is built on the assumption that prices are sticky in the short run and flexible in the long run. It shows how shocks to the economy cause output to deviate temporarily from the level implied by the classical model.

The model also highlights the role of monetary policy. Poor monetary policy can be a source of shocks to the economy. A well-run monetary policy can respond to shocks and stabilize the economy.

In the chapters that follow, we refine our understanding of this model and our analysis of stabilization policy. Chapters 10 through 12 go beyond the quantity equation to refine our theory of aggregate demand. This refinement shows that aggregate demand depends on fiscal policy as well as monetary policy. Chapter 13 examines aggregate supply in more detail. Chapter 14 examines the debate over the virtues and limits of stabilization policy.

Summary

1. The crucial difference between the long run and the short run is that prices are flexible in the long run but sticky in the short run. The model of aggregate supply and aggregate demand provides a framework to analyze economic fluctuations and see how the impact of policies varies over different time horizons.
2. The aggregate demand curve slopes downward. It tells us that the lower the price level, the greater the aggregate quantity of goods and services demanded.
3. In the long run, the aggregate supply curve is vertical because output is determined by the amounts of capital and labor and by the available technology, but not by the level of prices. Therefore, shifts in aggregate demand affect the price level but not output or employment.
4. In the short run, the aggregate supply curve is horizontal, because wages and prices are sticky at predetermined levels. Therefore, shifts in aggregate demand affect output and employment.
5. Shocks to aggregate demand and aggregate supply cause economic fluctuations. Because the Fed can shift the aggregate demand curve, it can attempt to offset these shocks to maintain output and employment at their natural rates.

KEY CONCEPTS

Aggregate demand
Aggregate supply
Shocks
Demand shocks
Supply shocks
Stabilization policy

QUESTIONS FOR REVIEW

1. Give an example of a price that is sticky in the short run and flexible in the long run.
2. Why does the aggregate demand curve slope downward?
3. Explain the impact of an increase in the money supply in the short run and in the long run.
4. Why is it easier for the Fed to deal with demand shocks than with supply shocks?

PROBLEMS AND APPLICATIONS

1. Suppose that a change in government regulations allows banks to start paying interest on checking accounts. Recall that the money stock is the sum of currency and demand deposits, including checking accounts, so this regulatory change makes holding money more attractive.
   a. How does this change affect the demand for money?
   b. What happens to the velocity of money?
   c. If the Fed keeps the money supply constant, what will happen to output and prices in the short run and in the long run?
   d. Should the Fed keep the money supply constant in response to this regulatory change? Why or why not?
2. Suppose the Fed reduces the money supply by 5 percent.
   a. What happens to the aggregate demand curve?
   b. What happens to the level of output and the price level in the short run and in the long run?
   c. According to Okun's law, what happens to unemployment in the short run and in the long run? (Hint: Okun's law is the relationship between output and unemployment discussed in Chapter 2.)
   d. What happens to the real interest rate in the short run and in the long run? (Hint: Use the model of the real interest rate in Chapter 3 to see what happens when output changes.)
3. Let's examine how the goals of the Federal Reserve influence its response to shocks. Suppose Fed A cares only about keeping the price level stable, and Fed B cares only about keeping output and employment at their natural rates. Explain how each Fed would respond to
   a. An exogenous decrease in the velocity of money.
   b. An exogenous increase in the price of oil.
4. The official arbiter of when recessions begin and end is the National Bureau of Economic Research, a nonprofit economics research group. Go to the NBER's Web site (www.nber.org) and find the latest turning point in the business cycle. When did it occur? Was this a switch from expansion to contraction or the other way around? List all the recessions (contractions) that have occurred during your lifetime and the dates when they began and ended.

CHAPTER 10 TEN

Aggregate Demand I

> I shall argue that the postulates of the classical theory are applicable to a special case only and not to the general case. . . . Moreover, the characteristics of the special case assumed by the classical theory happen not to be those of the economic society in which we actually live, with the result that its teaching is misleading and disastrous if we attempt to apply it to the facts of experience.
>
> — John Maynard Keynes, The General Theory

Of all the economic fluctuations in world history, the one that stands out as particularly large, painful, and intellectually significant is the Great Depression of the 1930s. During this time, the United States and many other countries experienced massive unemployment and greatly reduced incomes. In the worst year, 1933, one-fourth of the U.S. labor force was unemployed, and real GDP was 30 percent below its 1929 level.

This devastating episode caused many economists to question the validity of classical economic theory—the theory we examined in Chapters 3 through 6. Classical theory seemed incapable of explaining the Depression. According to that theory, national income depends on factor supplies and the available technology, neither of which changed substantially from 1929 to 1933. After the onset of the Depression, many economists believed that a new model was needed to explain such a large and sudden economic downturn and to suggest government policies that might reduce the economic hardship so many people faced.

In 1936 the British economist John Maynard Keynes revolutionized economics with his book The General Theory of Employment, Interest, and Money. Keynes proposed a new way to analyze the economy, which he presented as an alternative to classical theory. His vision of how the economy works quickly became a center of controversy. Yet, as economists debated The General Theory, a new understanding of economic fluctuations gradually developed.

Keynes proposed that low aggregate demand is responsible for the low income and high unemployment that characterize economic downturns. He criticized classical theory for assuming that aggregate supply alone—capital, labor, and technology—determines national income. Economists today reconcile these two views

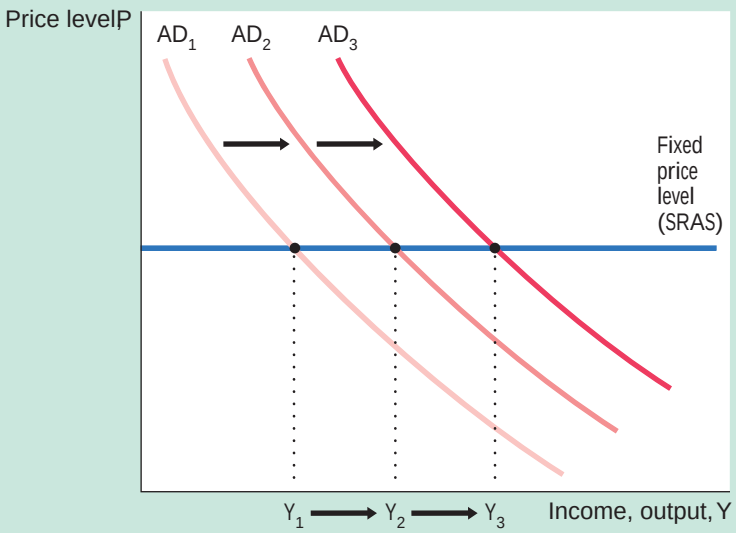
with the model of aggregate demand and aggregate supply introduced in Chapter 9. In the long run, prices are flexible, and aggregate supply determines income. But in the short run, prices are sticky, so changes in aggregate demand influence income.

In this chapter and the next, we continue our study of economic fluctuations by looking more closely at aggregate demand. Our goal is to identify the variables that shift the aggregate demand curve, causing fluctuations in national income. We also examine more fully the tools policymakers can use to influence aggregate demand. In Chapter 9 we derived the aggregate demand curve from the quantity theory of money, and we showed that monetary policy can shift the aggregate demand curve. In this chapter we see that the government can influence aggregate demand with both monetary and fiscal policy.

The model of aggregate demand developed in this chapter, called the ***IS–LM*** model, is the leading interpretation of Keynes's theory. The goal of the model is to show what determines national income for any given price level. There are two ways to view this exercise. We can view the IS–LM model as showing what causes income to change in the short run when the price level is fixed. Or we can view the model as showing what causes the aggregate demand curve to shift. These two views of the model are equivalent: as Figure 10-1 shows, in the short run when the price level is fixed, shifts in the aggregate demand curve lead to changes in national income.

The two parts of the IS–LM model are, not surprisingly, the ***IS*** curve and the ***LM*** curve. IS stands for "investment" and "saving," and the IS curve represents what's going on in the market for goods and services (which we first discussed in Chapter 3). LM stands for "liquidity" and "money," and the LM curve represents what's happening to the supply and demand for money (which we first discussed in Chapter 4). Because the interest rate influences both investment and money

figure 10-1



Shifts in Aggregate Demand For a given price level, national income fluctuates because of shifts in the aggregate demand curve. The IS–LM model takes the price level as given and shows what causes income to change. The model therefore shows what causes aggregate demand to shift.

demand, it is the variable that links the two halves of the IS–LM model. The model shows how interactions between these markets determine the position and slope of the aggregate demand curve and, therefore, the level of national income in the short run.[1]

10-1 The Goods Market and the *IS* Curve

The IS curve plots the relationship between the interest rate and the level of income that arises in the market for goods and services. To develop this relationship, we start with a basic model called the Keynesian cross. This model is the simplest interpretation of Keynes's theory of national income and is a building block for the more complex and realistic IS–LM model.

The Keynesian Cross

In *The General Theory*, Keynes proposed that an economy's total income was, in the short run, determined largely by the desire to spend by households, firms, and the government. The more people want to spend, the more goods and services firms can sell. The more firms can sell, the more output they will choose to produce and the more workers they will choose to hire. Thus, the problem during recessions and depressions, according to Keynes, was inadequate spending. The Keynesian cross is an attempt to model this insight.

Planned Expenditure We begin our derivation of the Keynesian cross by drawing a distinction between actual and planned expenditure. *Actual expenditure* is the amount households, firms, and the government spend on goods and services, and as we first saw in Chapter 2, it equals the economy's gross domestic product (GDP). *Planned expenditure* is the amount households, firms, and the government would like to spend on goods and services.

Why would actual expenditure ever differ from planned expenditure? The answer is that firms might engage in unplanned inventory investment because their sales do not meet their expectations. When firms sell less of their product than they planned, their stock of inventories automatically rises; conversely, when firms sell more than planned, their stock of inventories falls. Because these unplanned changes in inventory are counted as investment spending by firms, actual expenditure can be either above or below planned expenditure.

Now consider the determinants of planned expenditure. Assuming that the economy is closed, so that net exports are zero, we write planned expenditure E as the sum of consumption C, planned investment I, and government purchases G:

$$E = C + I + G.$$

[1] The IS–LM model was introduced in a classic article by the Nobel-Prize-winning economist John R. Hicks, "Mr. Keynes and the Classics: A Suggested Interpretation," *Econometrica* 5 (1937): 147–159.

To this equation, we add the consumption function

$$C = C(Y - T).$$

This equation states that consumption depends on disposable income ($Y - T$), which is total income Y minus taxes T. To keep things simple, for now we take planned investment as exogenously fixed:

$$I = \bar{I}.$$

And as in Chapter 3, we assume that fiscal policy—the levels of government purchases and taxes—is fixed:

$$G = \bar{G},$$
$$T = \bar{T}.$$

Combining these five equations, we obtain

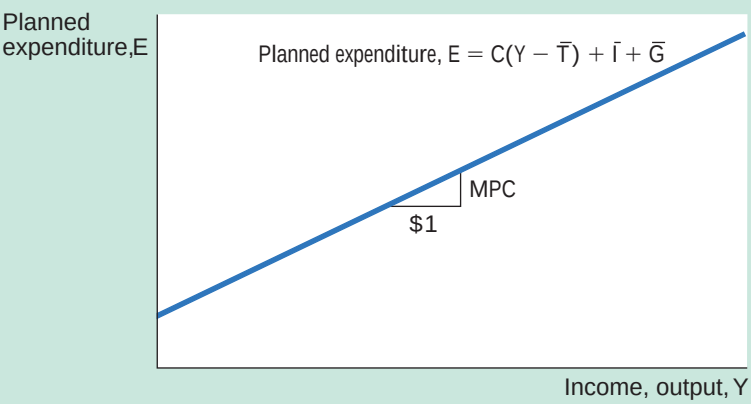
$$E = C(Y - \bar{T}) + \bar{I} + \bar{G}.$$

This equation shows that planned expenditure is a function of income Y, the level of planned investment $\bar{I}$, and the fiscal policy variables $\bar{G}$ and $\bar{T}$.

Figure 10-2 graphs planned expenditure as a function of the level of income. This line slopes upward because higher income leads to higher consumption and thus higher planned expenditure. The slope of this line is the marginal propensity to consume, the MPC: it shows how much planned expenditure increases when income rises by $1. This planned-expenditure function is the first piece of the model called the Keynesian cross.

The Economy in Equilibrium The next piece of the Keynesian cross is the assumption that the economy is in equilibrium when actual expenditure equals planned expenditure. This assumption is based on the idea that when people's plans have been realized, they have no reason to change what they are doing.

figure 10-2



Planned Expenditure as a Function of Income Planned expenditure depends on income because higher income leads to higher consumption, which is part of planned expenditure. The slope of this planned-expenditure function is the marginal propensity to consume, MPC.

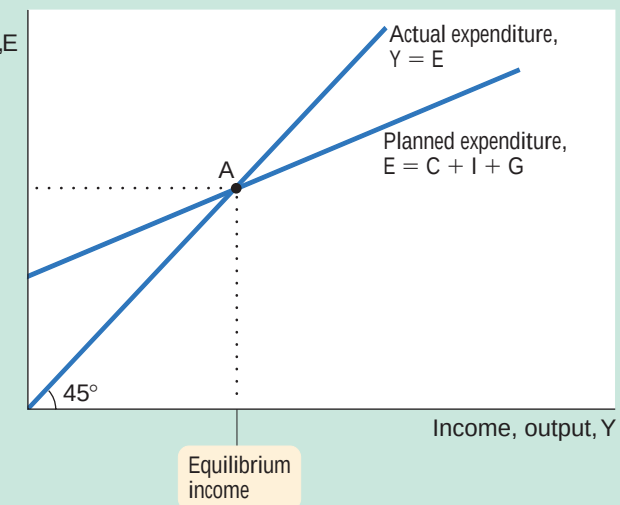
Recalling that Y as GDP equals not only total income but also total actual expenditure on goods and services, we can write this equilibrium condition as

$$\text{Actual Expenditure} = \text{Planned Expenditure}$$
$$Y = E.$$

The 45-degree line in Figure 10-3 plots the points where this condition holds. With the addition of the planned-expenditure function, this diagram becomes the Keynesian cross. The equilibrium of this economy is at point A, where the planned-expenditure function crosses the 45-degree line.

How does the economy get to the equilibrium? In this model, inventories play an important role in the adjustment process. Whenever the economy is not in equilibrium, firms experience unplanned changes in inventories, and this induces them to change production levels. Changes in production in turn influence total income and expenditure, moving the economy toward equilibrium.

figure 10-3

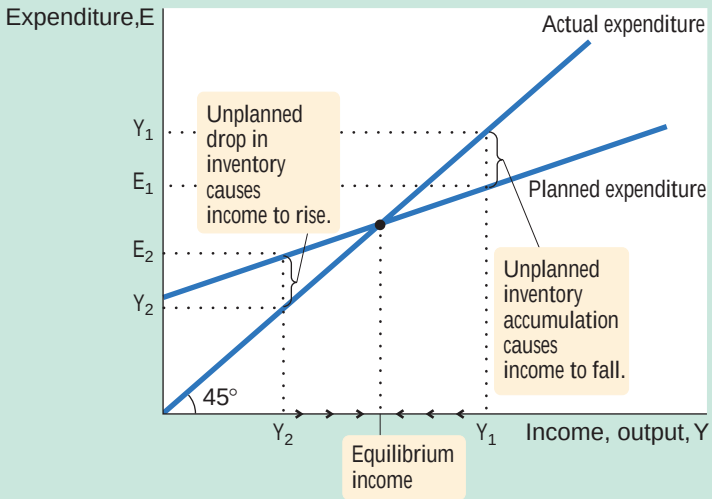


The Keynesian Cross The equilibrium in the Keynesian cross is at point A, where income (actual expenditure) equals planned expenditure.

For example, suppose the economy were ever to find itself with GDP at a level greater than the equilibrium level, such as the level Y_1 in Figure 10-4. In this case, planned expenditure E_1 is less than production Y_1, so firms are selling less than they are producing. Firms add the unsold goods to their stock of inventories. This unplanned rise in inventories induces firms to lay off workers and reduce production, and these actions in turn reduce GDP. This process of unintended inventory accumulation and falling income continues until income falls to the equilibrium level Y.

Similarly, suppose GDP were at a level lower than the equilibrium level, such as the level Y_2 in Figure 10-4. In this case, planned expenditure E_2 is greater than production Y_2. Firms meet the high level of sales by drawing down their inventories.

figure 10-4



The Adjustment to Equilibrium in the Keynesian Cross If firms were producing at level Y_1, then planned expenditure E_1 would fall short of production, and firms would accumulate inventories. This inventory accumulation would induce firms to reduce production. Similarly, if firms were producing at level Y_2, then planned expenditure E_2 would exceed production, and firms would run down their inventories. This fall in inventories would induce firms to raise production. In both cases, the firms' decisions drive the economy toward equilibrium.

But when firms see their stock of inventories dwindle, they hire more workers and increase production. GDP rises, and the economy approaches the equilibrium.

In summary, the Keynesian cross shows how income is determined for given levels of planned investment and fiscal policy G and T. We can use this model to show how income changes when one of these exogenous variables changes.

Fiscal Policy and the Multiplier: Government Purchases Consider how changes in government purchases affect the economy. Because government purchases are one component of expenditure, higher government purchases result in higher planned expenditure for any given level of income. If government purchases rise by ΔG, then the planned-expenditure schedule shifts upward by ΔG, as in Figure 10-5. The equilibrium of the economy moves from point A to point B.

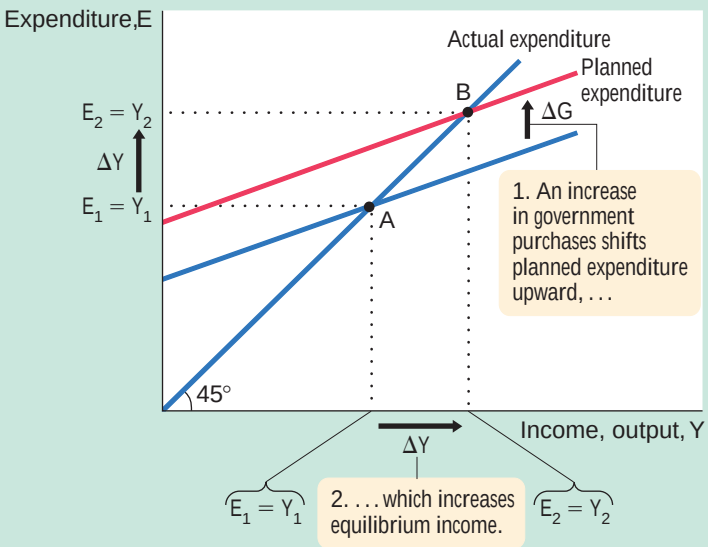
© The New Yorker Collection, 1992 Dana Fradon from cartoonbank.com. All Rights Reserved.

"Your Majesty, my voyage will not only forge a new route to the spice f the East but also create over three thousand new jobs.

This graph shows that an increase in government purchases leads to an even greater increase in income. That is, ΔY is larger than ΔG. The ratio $\Delta Y / \Delta G$ is called the government-purchases multiplier; it tells us how much income rises in response to a $1 increase in government purchases. An implication of the Keynesian cross is that the government-purchases multiplier is larger than 1.

Why does fiscal policy have a multiplied effect on income?

figure 10-5



An Increase in Government Purchases in the Keynesian Cross
An increase in government purchases of ΔG raises planned expenditure by that amount for any given level of income. The equilibrium moves from point A to point B, and income rises from Y_1 to Y_2. Note that the increase in income ΔY exceeds the increase in government purchases ΔG. Thus, fiscal policy has a multiplied effect on income.

The reason is that, according to the consumption function $C = C(Y - T)$, higher income causes higher consumption. When an increase in government purchases raises income, it also raises consumption, which further raises income, which further raises consumption, and so on. Therefore, in this model, an increase in government purchases causes a greater increase in income.

How big is the multiplier? To answer this question, we trace through each step of the change in income. The process begins when expenditure rises by ΔG, which implies that income rises by ΔG as well. This increase in income in turn raises consumption by $MPC \times \Delta G$, where MPC is the marginal propensity to consume. This increase in consumption raises expenditure and income once again. This second increase in income of $MPC \times \Delta G$ again raises consumption, this time by $MPC \times (MPC \times \Delta G)$, which again raises expenditure and income, and so on. This feedback from consumption to income to consumption continues indefinitely. The total effect on income is

$$\begin{aligned}
\text{Initial Change in Government Purchases} &= \Delta G \\
\text{First Change in Consumption} &= MPC \times \Delta G \\
\text{Second Change in Consumption} &= MPC^2 \times \Delta G \\
\text{Third Change in Consumption} &= MPC^3 \times \Delta G \\
&\vdots \qquad\qquad \vdots \\
\hline
\Delta Y &= (1 + MPC + MPC^2 + MPC^3 + \cdots)\Delta G.
\end{aligned}$$

The government-purchases multiplier is

$$\Delta Y / \Delta G = 1 + MPC + MPC^2 + MPC^3 + \cdots$$

This expression for the multiplier is an example of an *infinite geometric series*. A result from algebra allows us to write the multiplier as[2]

$$\Delta Y/\Delta G = 1/(1 - MPC).$$

For example, if the marginal propensity to consume is 0.6, the multiplier is

$$\begin{aligned}\Delta Y/\Delta G &= 1 + 0.6 + 0.6^2 + 0.6^3 + \cdots \\ &= 1/(1 - 0.6) \\ &= 2.5.\end{aligned}$$

In this case, a $1.00 increase in government purchases raises equilibrium income by $2.50.[3]

Fiscal Policy and the Multiplier: Taxes Consider now how changes in taxes affect equilibrium income. A decrease in taxes of ΔT immediately raises disposable income $Y - T$ by ΔT and, therefore, increases consumption by $MPC \times \Delta T$. For any given level of income Y, planned expenditure is now higher. As Figure 10-6 shows, the planned-expenditure schedule shifts upward by $MPC \times \Delta T$. The equilibrium of the economy moves from point A to point B.

---

[2] *Mathematical note:* We prove this algebraic result as follows. Let

$$z = 1 + x + x^2 + \cdots.$$

Multiply both sides of this equation by x:

$$xz = x + x^2 + x^3 + \cdots.$$

Subtract the second equation from the first:

$$z - xz = 1.$$

Rearrange this last equation to obtain

$$z(1 - x) = 1,$$

which implies

$$z = 1/(1 - x).$$

This completes the proof.

[3] *Mathematical note:* The government-purchases multiplier is most easily derived using a little calculus. Begin with the equation

$$Y = C(Y - T) + I + G.$$

Holding T and I fixed, differentiate to obtain

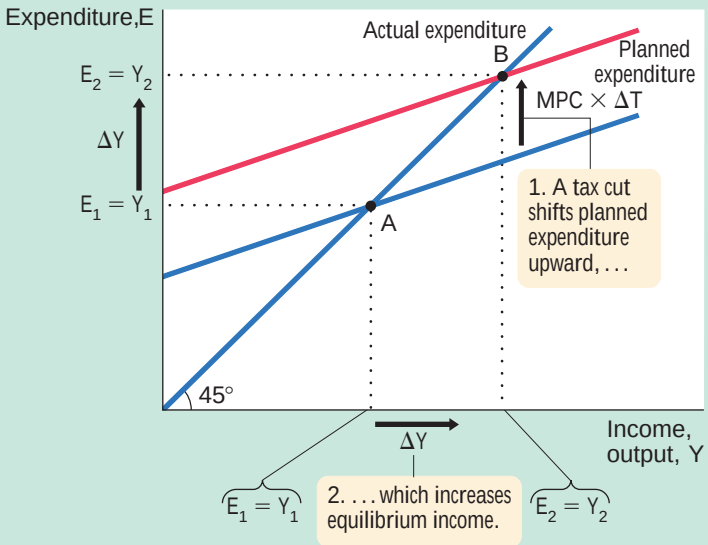
$$dY = C' dY + dG,$$

and then rearrange to find

$$dY/dG = 1/(1 - C').$$

This is the same as the equation in the text.

figure 10-6



A Decrease in Taxes in the Keynesian Cross A decrease in taxes of ΔT raises planned expenditure by $MPC \times \Delta T$ for any given level of income. The equilibrium moves from point A to point B, and income rises from Y_1 to Y_2. Again, fiscal policy has a multiplied effect on income.

Just as an increase in government purchases has a multiplied effect on income, so does a decrease in taxes. As before, the initial change in expenditure, now $MPC \times \Delta T$, is multiplied by $1/(1 - MPC)$. The overall effect on income of the change in taxes is

$$\Delta Y/\Delta T = -MPC/(1 - MPC).$$

This expression is the *tax multiplier*, the amount income changes in response to a $1 change in taxes. For example, if the marginal propensity to consume is 0.6, then the tax multiplier is

$$\Delta Y/\Delta T = -0.6/(1 - 0.6) = -1.5.$$

In this example, a $1.00 cut in taxes raises equilibrium income by $1.50.[4]

---

[4] Mathematical note: As before, the multiplier is most easily derived using a little calculus. Begin with the equation

$$Y = C(Y - T) + I + G.$$

Holding I and G fixed, differentiate to obtain

$$dY = C'(dY - dT),$$

and then rearrange to find

$$dY/dT = -C'/(1 - C').$$

This is the same as the equation in the text.

CASE STUDY

Cutting Taxes to Stimulate the Economy

When John F. Kennedy became president of the United States in 1961, he brought to Washington some of the brightest young economists of the day to work on his Council of Economic Advisers. These economists, who had been schooled in the economics of Keynes, brought Keynesian ideas to discussions of economic policy at the highest level.

One of the council's first proposals was to expand national income by reducing taxes. This eventually led to a substantial cut in personal and corporate income taxes in 1964. The tax cut was intended to stimulate expenditure on consumption and investment and thus lead to higher levels of income and employment. When a reporter asked Kennedy why he advocated a tax cut, Kennedy replied, "To stimulate the economy. Don't you remember your Economics 101?"

As Kennedy's economic advisers predicted, the passage of the tax cut was followed by an economic boom. Growth in real GDP was 5.3 percent in 1964 and 6.0 percent in 1965. The unemployment rate fell from 5.7 percent in 1963 to 5.2 percent in 1964 and then to 4.5 percent in 1965.[5]

Economists continue to debate the source of this rapid growth in the early 1960s. A group called *supply-siders* argues that the economic boom resulted from the incentive effects of the cut in income tax rates. According to supply-siders, when workers are allowed to keep a higher fraction of their earnings, they supply substantially more labor and expand the aggregate supply of goods and services. Keynesians, however, emphasize the impact of tax cuts on aggregate demand. Most likely, both views have some truth. Tax cuts stimuate aggregate supply by improving workers' incentives and expand aggregate demand by raising households' disposable income.

When George W. Bush was elected president in 2001, a major element of his platform was a cut in income taxes. Bush and his advisers used both supply-side and Keynesian rhetoric to make the case for their policy. During the campaign, when the economy was doing fine, they argued that lower marginal tax rates would improve work incentives. But then the economy started to slow: unemployment rose from 3.9 percent in October 2000 to 4.5 percent in April 2001. The argument shifted to emphasize that the tax cut would stimulate spending and reduce the risk of recession.

Congress passed the tax cut in May 2001. Compared to the original Bush proposal, the bill cut tax rates less in the long run. But it added an immediate tax rebate of $600 per family ($300 for single taxpayers) that was mailed out in the summer of 2001. Consistent with Keynesian theory, the goal of the rebate was to provide an immediate stimulus to aggregate demand.

---

[5] For an analysis of the 1964 tax cut by one of Kennedy's economists, see Arthur Okun, "Measuring the Impact of the 1964 Tax Reduction," in W. W. Heller, ed., *Perspectives on Economic Growth* (New York: Random House, 1968); reprinted in Arthur M. Okun, *Economics for Policymaking* (Cambridge, MA: MIT Press, 1983), 409–423.

The Interest Rate, Investment, and the *IS* Curve

The Keynesian cross is only a steppingstone on our path to the *IS–LM* model. The Keynesian cross is useful because it shows how the spending plans of households, firms, and the government determine the economy's income. Yet it makes the simplifying assumption that the level of planned investment is fixed. As we discussed in Chapter 3, an important macroeconomic relationship is that planned investment depends on the interest rate.

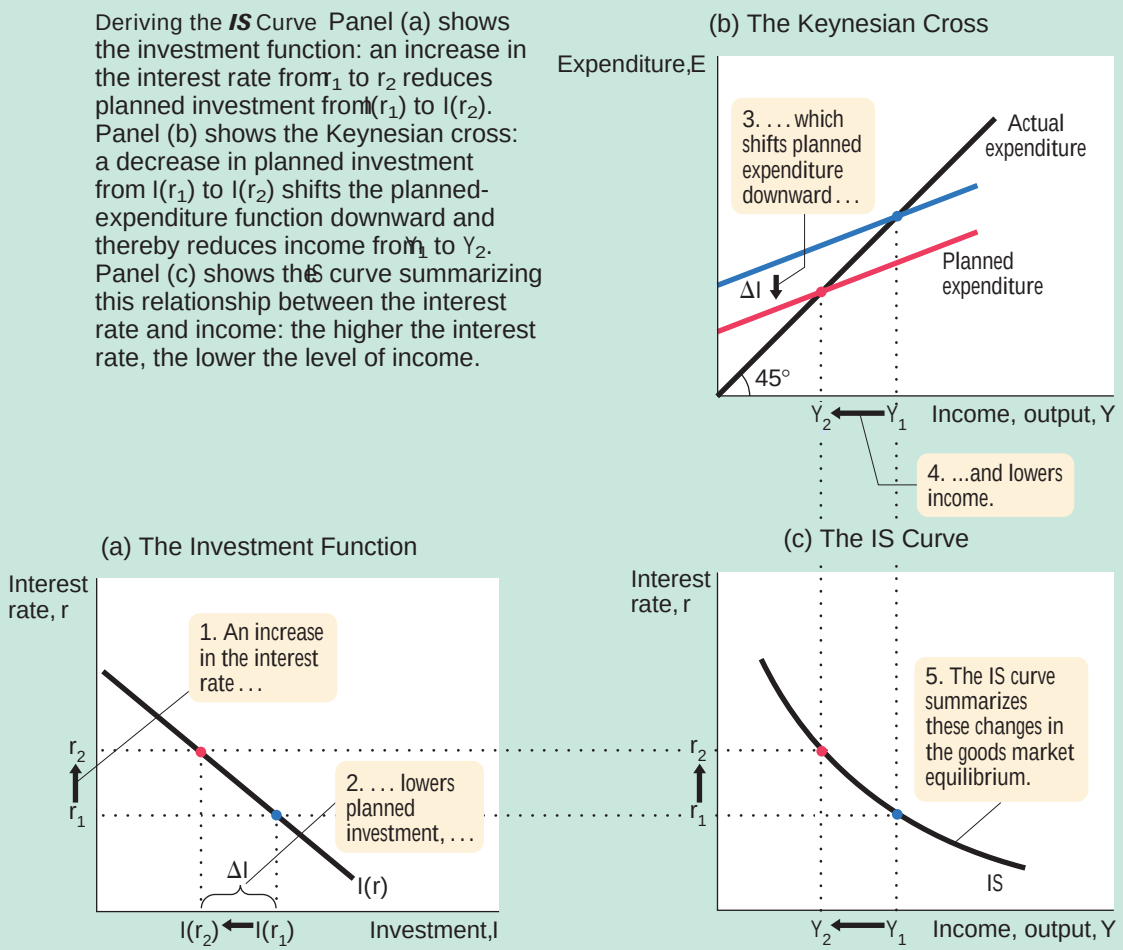
To add this relationship between the interest rate and investment to our model, we write the level of planned investment as

$$I = I(r).$$

This investment function is graphed in panel (a) of Figure 10-7. Because the interest rate is the cost of borrowing to finance investment projects, an increase in

figure 10-7

Deriving the *IS* Curve Panel (a) shows the investment function: an increase in the interest rate from r_1 to r_2 reduces planned investment from $I(r_1)$ to $I(r_2)$. Panel (b) shows the Keynesian cross: a decrease in planned investment from $I(r_1)$ to $I(r_2)$ shifts the planned-expenditure function downward and thereby reduces income from Y_1 to Y_2. Panel (c) shows the *IS* curve summarizing this relationship between the interest rate and income: the higher the interest rate, the lower the level of income.



the interest rate reduces planned investment. As a result, the investment function slopes downward.

To determine how income changes when the interest rate changes, we can combine the investment function with the Keynesian-cross diagram. Because investment is inversely related to the interest rate, an increase in the interest rate from r_1 to r_2 reduces the quantity of investment from $I(r_1)$ to $I(r_2)$. The reduction in planned investment, in turn, shifts the planned-expenditure function downward, as in panel (b) of Figure 10-7. The shift in the planned-expenditure function causes the level of income to fall from Y_1 to Y_2. Hence, an increase in the interest rate lowers income.

The IS curve, shown in panel (c) of Figure 10-7, summarizes this relationship between the interest rate and the level of income. In essence, the IS curve combines the interaction between r and I expressed by the investment function and the interaction between I and Y demonstrated by the Keynesian cross. Because an increase in the interest rate causes planned investment to fall, which in turn causes income to fall, the IS curve slopes downward.

How Fiscal Policy Shifts the *IS* Curve

The IS curve shows us, for any given interest rate, the level of income that brings the goods market into equilibrium. As we learned from the Keynesian cross, the level of income also depends on fiscal policy. The IS curve is drawn for a given fiscal policy; that is, when we construct the IS curve, we hold G and T fixed. When fiscal policy changes, the IS curve shifts.

Figure 10-8 uses the Keynesian cross to show how an increase in government purchases by ΔG shifts the IS curve. This figure is drawn for a given interest rate r and thus for a given level of planned investment. The Keynesian cross shows that this change in fiscal policy raises planned expenditure and thereby increases equilibrium income from Y_1 to Y_2. Therefore, an increase in government purchases shifts the IS curve outward.

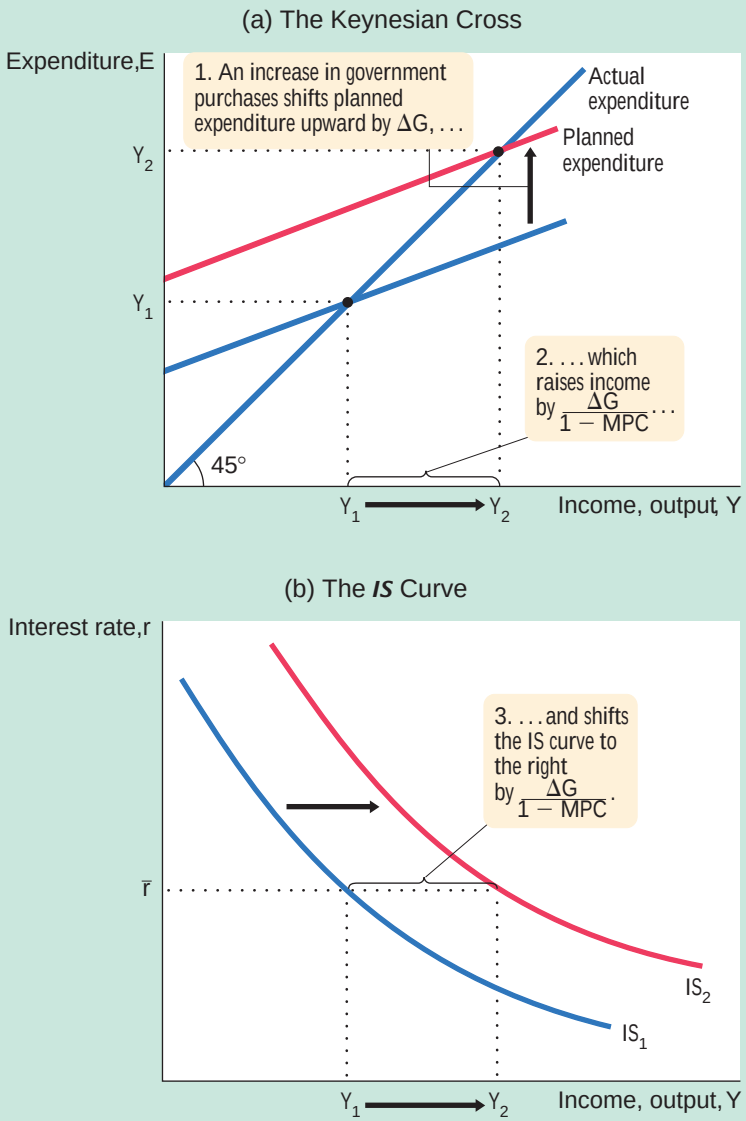
We can use the Keynesian cross to see how other changes in fiscal policy shift the IS curve. Because a decrease in taxes also expands expenditure and income, it too shifts the IS curve outward. A decrease in government purchases or an increase in taxes reduces income; therefore, such a change in fiscal policy shifts the IS curve inward.

In summary, the IS curve shows the combinations of the interest rate and the level of income that are consistent with equilibrium in the market for goods and services. The IS curve is drawn for a given fiscal policy. Changes in fiscal policy that raise the demand for goods and services shift the IS curve to the right. Changes in fiscal policy that reduce the demand for goods and services shift the IS curve to the left.

A Loanable-Funds Interpretation of the *IS* Curve

When we first studied the market for goods and services in Chapter 3, we noted an equivalence between the supply and demand for goods and services and the supply and demand for loanable funds. This equivalence provides another way to interpret the IS curve.

figure 10-8



An Increase in Government Purchases Shifts the ***IS*** Curve Outward Panel (a) shows that an increase in government purchases raises planned expenditure. For any given interest rate, the upward shift in planned expenditure of ΔG leads to an increase in income Y of $\Delta G/(1 - MPC)$. Therefore, in panel (b), the IS curve shifts to the right by this amount.

Recall that the national income accounts identity can be written as

$$Y - C - G = I$$
$$S = I.$$

The left-hand side of this equation is national saving, or S, and the right-hand side is investment, I. National saving represents the supply of loanable funds, and investment represents the demand for these funds.

To see how the market for loanable funds produces the *IS* curve, substitute the consumption function for C and the investment function for I:

$$Y - C(Y - T) - G = I(r).$$

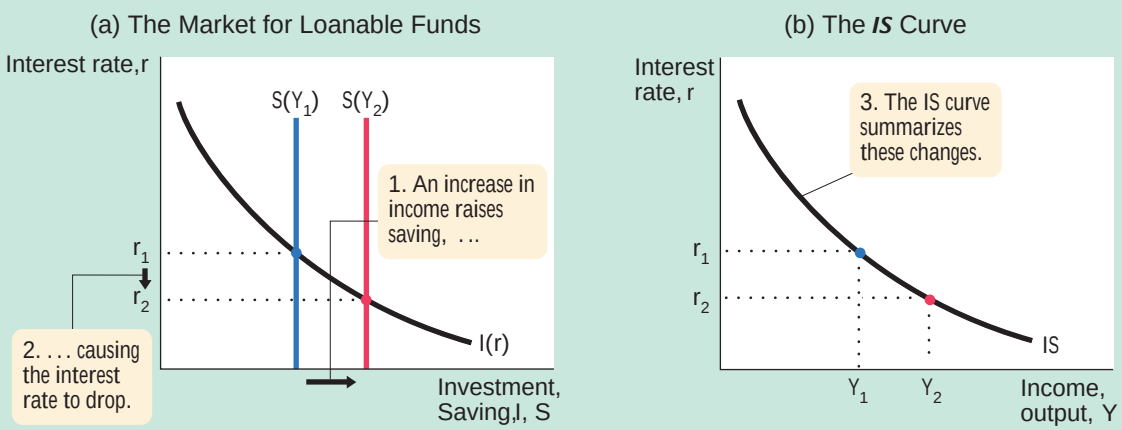
The left-hand side of this equation shows that the supply of loanable funds depends on income and fiscal policy. The right-hand side shows that the demand for loanable funds depends on the interest rate. The interest rate adjusts to equilibrate the supply and demand for loans.

As Figure 10-9 illustrates, we can interpret the *IS* curve as showing the interest rate that equilibrates the market for loanable funds for any given level of income. When income rises from Y_1 to Y_2, national saving, which equals $Y - C - G$, increases. (Consumption rises by less than income, because the marginal propensity to consume is less than 1.) As panel (a) shows, the increased supply of loanable funds drives down the interest rate from r_1 to r_2. The IS curve in panel (b) summarizes this relationship: higher income implies higher saving, which in turn implies a lower equilibrium interest rate. For this reason, the *IS* curve slopes downward.

This alternative interpretation of the *IS* curve also explains why a change in fiscal policy shifts the *IS* curve. An increase in government purchases or a decrease in taxes reduces national saving for any given level of income. The reduced supply of loanable funds raises the interest rate that equilibrates the market. Because the interest rate is now higher for any given level of income, the *IS* curve shifts upward in response to the expansionary change in fiscal policy.

Finally, note that the *IS* curve does not determine either income or the interest rate. Instead, the *IS* curve is a relationship between Y and r arising in the

figure 10-9



A Loanable-Funds Interpretation of the *IS* Curve Panel (a) shows that an increase in income from Y_1 to Y_2 raises saving and thus lowers the interest rate that equilibrates the supply and demand for loanable funds. The *IS* curve in panel (b) expresses this negative relationship between income and the interest rate.

market for goods and services or, equivalently, the market for loanable funds. To determine the equilibrium of the economy, we need another relationship between these two variables, to which we now turn.

10-2 The Money Market and the *LM* Curve

The LM curve plots the relationship between the interest rate and the level of income that arises in the market for money balances. To understand this relationship, we begin by looking at a theory of the interest rate, called the theory of liquidity preference.

The Theory of Liquidity Preference

In his classic work The General Theory, Keynes offered his view of how the interest rate is determined in the short run. That explanation is called the theory of liquidity preference, because it posits that the interest rate adjusts to balance the supply and demand for the economy's most liquid asset—money. Just as the Keynesian cross is a building block for the *IS* curve, the theory of liquidity preference is a building block for the *LM* curve.

To develop this theory, we begin with the supply of real money balances. If M stands for the supply of money and P stands for the price level, then M/P is the supply of real money balances. The theory of liquidity preference assumes there is a fixed supply of real money balances. That is,

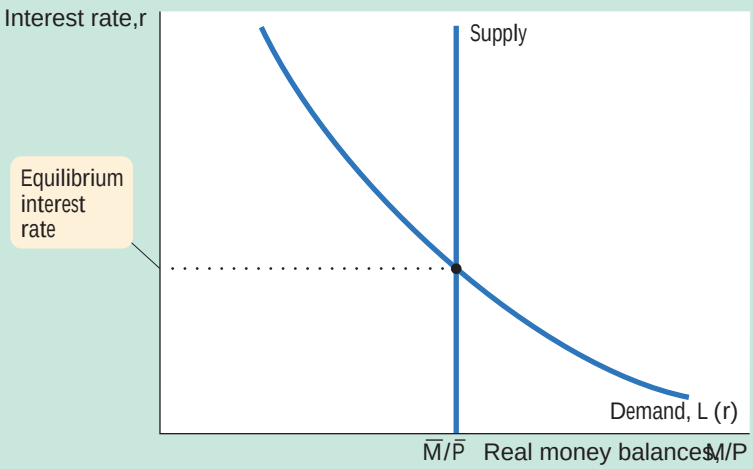
$$(M/P)^s = \bar{M}/\bar{P}.$$

The money supply M is an exogenous policy variable chosen by a central bank, such as the Federal Reserve. The price level P is also an exogenous variable in this model. (We take the price level as given because it is determined in the *IS–LM* model—our ultimate goal in this chapter—over longer periods than we are considering here. The *IS–LM* model explains the short run when the price level is fixed.) These assumptions imply that the supply of real money balances is fixed and, in particular, does not depend on the interest rate. Thus, when we plot the supply of real money balances against the interest rate in Figure 10-10, we obtain a vertical supply curve.

Next, consider the demand for real money balances. The theory of liquidity preference posits that the interest rate is one determinant of how much money people choose to hold. The reason is that the interest rate is the opportunity cost of holding money: it is what you forgo by holding some of your assets as money, which does not bear interest, instead of as interest-bearing bank deposits or bonds. When the interest rate rises, people want to hold less of their wealth in the form of money. We can write the demand for real money balances as

$$(M/P)^d = L(r),$$

figure 10-10



The Theory of Liquidity Preference The supply and demand for real money balances determine the interest rate. The supply curve for real money balances is vertical because the supply does not depend on the interest rate. The demand curve is downward sloping because a higher interest rate raises the cost of holding money and thus lowers the quantity demanded. At the equilibrium interest rate, the quantity of real money balances demanded equals the quantity supplied.

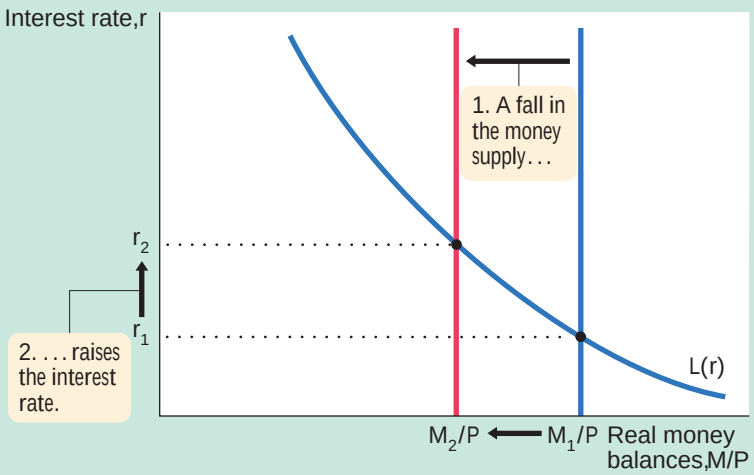
where the function $L(\ )$ shows that the quantity of money demanded depends on the interest rate. Thus, the demand curve in Figure 10-10 slopes downward because higher interest rates reduce the quantity of real money balances demanded.[6]

According to the theory of liquidity preference, the supply and demand for real money balances determine what interest rate prevails in the economy. That is, the interest rate adjusts to equilibrate the money market. As the figure shows, at the equilibrium interest rate, the quantity of real money balances demanded equals the quantity supplied.

How does the interest rate get to this equilibrium of money supply and money demand? The adjustment occurs because whenever the money market is not in equilibrium, people try to adjust their portfolios of assets and, in the process, alter the interest rate. For instance, if the interest rate is above the equilibrium level, the quantity of real money balances supplied exceeds the quantity demanded. Individuals holding the excess supply of money try to convert some of their non-interest-bearing money into interest-bearing bank deposits or bonds. Banks and bond issuers, who prefer to pay lower interest rates, respond to this excess supply of money by lowering the interest rates they offer. Conversely, if the interest rate is below the equilibrium level, so that the quantity of money demanded exceeds the quantity supplied, individuals try to obtain money by selling bonds or making bank withdrawals. To attract now-scarcer funds, banks and bond issuers respond by increasing the interest rates they offer. Eventually, the

[6] Note that r is being used to denote the interest rate here, as it was in our discussion of the IS curve. More accurately, it is the nominal interest rate that determines money demand and the real interest rate that determines investment. To keep things simple, we are ignoring expected inflation, which creates the difference between the real and nominal interest rates. The role of expected inflation in the IS–LM model is explored in Chapter 11.

figure 10-11



A Reduction in the Money Supply in the Theory of Liquidity Preference If the price level is fixed, a reduction in the money supply from M_1 to M_2 reduces the supply of real money balances. The equilibrium interest rate therefore rises from r_1 to r_2.

interest rate reaches the equilibrium level, at which people are content with their portfolios of monetary and nonmonetary assets.

Now that we have seen how the interest rate is determined, we can use the theory of liquidity preference to show how the interest rate responds to changes in the supply of money. Suppose, for instance, that the Fed suddenly decreases the money supply. A fall in M reduces M/P, because P is fixed in the model. The supply of real money balances shifts to the left, as in Figure 10-11. The equilibrium interest rate rises from r_1 to r_2, and the higher interest rate makes people satisfied to hold the smaller quantity of real money balances. The opposite would occur if the Fed had suddenly increased the money supply. Thus, according to the theory of liquidity preference, a decrease in the money supply raises the interest rate, and an increase in the money supply lowers the interest rate.

CASE STUDY

Did Paul Volcker's Monetary Tightening Raise or Lower Interest Rates?

The early 1980s saw the largest and quickest reduction in inflation in recent U.S. history. By the late 1970s inflation had reached the double-digit range; in 1979, consumer prices were rising at a rate of 11.3 percent per year. In October 1979, only two months after becoming the chairman of the Federal Reserve, Paul Volcker announced that monetary policy would aim to reduce the rate of inflation. This announcement began a period of tight money that, by 1983, brought the inflation rate down to about 3 percent.

How does such a monetary tightening influence interest rates? According to the theories we have been developing, the answer depends on the time horizon. Our analysis of the Fisher effect in Chapter 4 suggests that in the long run

Volcker's change in monetary policy would lower inflation, and this in turn would lead to lower nominal interest rates. Yet the theory of liquidity preference predicts that, in the short run when prices are sticky, anti-inflationary monetary policy would lead to falling real money balances and higher nominal interest rates.

Both conclusions are consistent with experience. Nominal interest rates did fall in the 1980s as inflation fell. But comparing the year before the October 1979 announcement and the year after, we find that real money balances (M1 divided by the CPI) fell 8.3 percent and the nominal interest rate (on short-term commercial loans) rose from 10.1 percent to 11.9 percent. Hence, although a monetary tightening leads to lower nominal interest rates in the long run, it leads to higher nominal interest rates in the short run.

Income, Money Demand, and the *LM* Curve

Having developed the theory of liquidity preference as an explanation for what determines the interest rate, we can now use the theory to derive the *LM* curve. We begin by considering the following question: How does a change in the economy's level of income Y affect the market for real money balances? The answer (which should be familiar from Chapter 4) is that the level of income affects the demand for money. When income is high, expenditure is high, so people engage in more transactions that require the use of money. Thus, greater income implies greater money demand. We can express these ideas by writing the money demand function as

$$(M/P)^d = L(r, Y).$$

The quantity of real money balances demanded is negatively related to the interest rate and positively related to income.

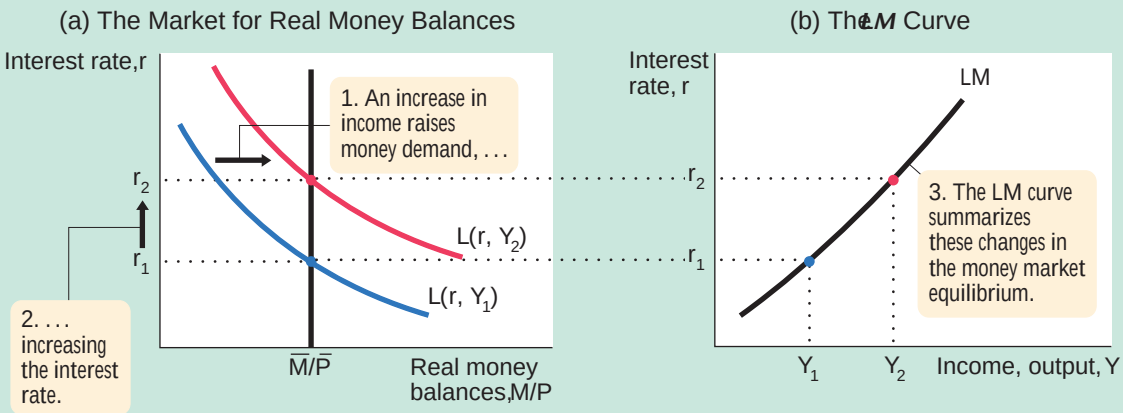
Using the theory of liquidity preference, we can figure out what happens to the equilibrium interest rate when the level of income changes. For example, consider what happens in Figure 10-12 when income increases from Y_1 to Y_2. As panel (a) illustrates, this increase in income shifts the money demand curve to the right. With the supply of real money balances unchanged, the interest rate must rise from r_1 to r_2 to equilibrate the money market. Therefore, according to the theory of liquidity preference, higher income leads to a higher interest rate.

The *LM* curve plots this relationship between the level of income and the interest rate. The higher the level of income, the higher the demand for real money balances, and the higher the equilibrium interest rate. For this reason, the *LM* curve slopes upward, as in panel (b) of Figure 10-12.

How Monetary Policy Shifts the *LM* Curve

The *LM* curve tells us the interest rate that equilibrates the money market at any level of income. Yet, as we saw earlier, the equilibrium interest rate also depends on the supply of real money balances, M/P. This means that the *LM* curve is

figure 10-12



Deriving the *LM* Curve Panel (a) shows the market for real money balances: an increase in income from Y_1 to Y_2 raises the demand for money and thus raises the interest rate from r_1 to r_2. Panel (b) shows the LM curve summarizing this relationship between the interest rate and income: the higher the level of income, the higher the interest rate.

drawn for a given supply of real money balances. If real money balances change, for example, if the Fed alters the money supply, the LM curve shifts.

We can use the theory of liquidity preference to understand how monetary policy shifts the LM curve. Suppose that the Fed decreases the money supply from M_1 to M_2, which causes the supply of real money balances to fall from M_1/P to M_2/P. Figure 10-13 shows what happens. Holding constant the amount of income and thus the demand curve for real money balances, we see that a reduction in the supply of real money balances raises the interest rate that equilibrates the money market. Hence, a decrease in the money supply shifts the LM curve upward.

In summary, the LM curve shows the combinations of the interest rate and the level of income that are consistent with equilibrium in the market for real money balances. The LM curve is drawn for a given supply of real money balances. Decreases in the supply of real money balances shift the LM curve upward. Increases in the supply of real money balances shift the LM curve downward.

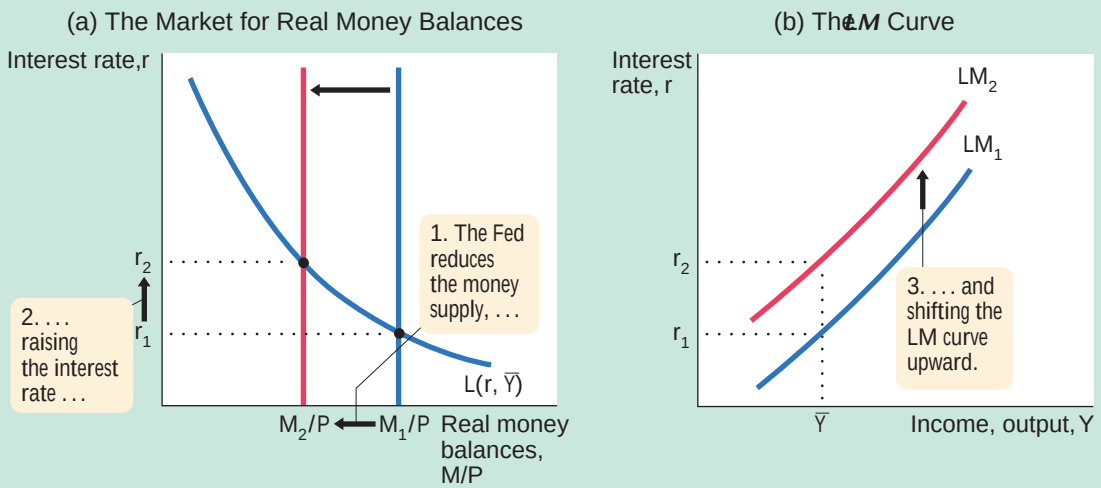
A Quantity-Equation Interpretation of the *LM* Curve

When we first discussed aggregate demand and the short-run determination of income in Chapter 9, we derived the aggregate demand curve from the quantity theory of money. We described the money market with the quantity equation,

$$MV = PY,$$

and assumed that velocity V is constant. This assumption implies that, for any given price level P, the supply of money M by itself determines the level of

figure 10-13



A Reduction in the Money Supply Shifts the *LM* Curve Upward Panel (a) shows that for any given level of income, $\bar{Y}$, a reduction in the money supply raises the interest rate that equilibrates the money market. Therefore, the *LM* curve in panel (b) shifts upward.

income Y. Because the level of income does not depend on the interest rate, the quantity theory is equivalent to a vertical *LM* curve.

We can derive the more realistic upward-sloping *LM* curve from the quantity equation by relaxing the assumption that velocity is constant. The assumption of constant velocity is based on the assumption that the demand for real money balances depends only on the level of income. Yet, as we have noted in our discussion of the liquidity-preference model, the demand for real money balances also depends on the interest rate: a higher interest rate raises the cost of holding money and reduces money demand. When people respond to a higher interest rate by holding less money, each dollar they do hold must be used more often to support a given volume of transactions. That is, the velocity of money must increase. We can write this as

$$MV(r) = PY.$$

The velocity function $V(r)$ indicates that velocity is positively related to the interest rate.

This form of the quantity equation yields an *LM* curve that slopes upward. Because an increase in the interest rate raises the velocity of money, it raises the level of income for any given money supply and price level. The *LM* curve expresses this positive relationship between the interest rate and income.

This equation also shows why changes in the money supply shift the *LM* curve. For any given interest rate and price level, the money supply and the level of income must move together. Thus, increases in the money supply shift the *LM* curve to the right, and decreases in the money supply shift the *LM* curve to the left.

Keep in mind that the quantity equation is merely another way to express the theory behind the *LM* curve. This quantity-theory interpretation of the *LM* curve is substantively the same as that provided by the theory of liquidity preference. In both cases, the *LM* curve represents a positive relationship between income and the interest rate that arises from the money market.

Finally, remember that the *LM* curve by itself does not determine either income Y or the interest rate r that will prevail in the economy. Like the *IS* curve, the *LM* curve is only a relationship between these two endogenous variables. The *IS* and *LM* curves together determine the economy's equilibrium.

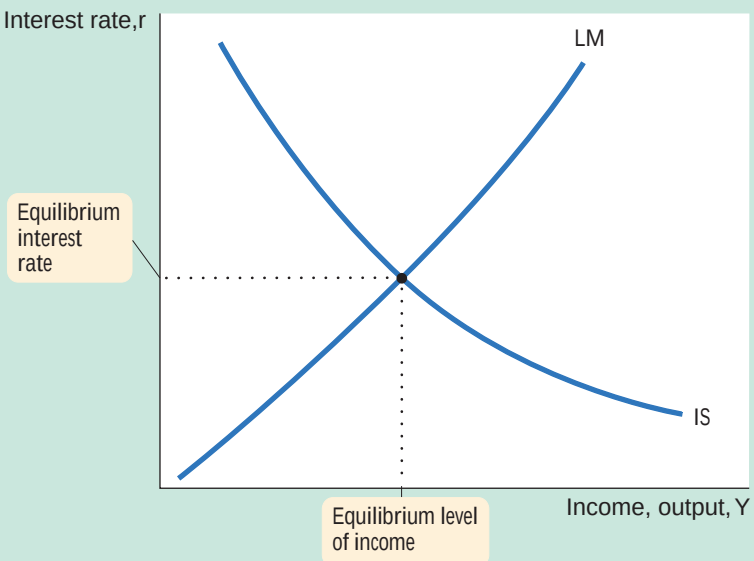
10-3 Conclusion: The Short-Run Equilibrium

We now have all the pieces of the *IS*–*LM* model. The two equations of this model are

$$Y = C(Y - T) + I(r) + G \qquad \text{IS},$$
$$M/P = L(r, Y) \qquad \text{LM}.$$

The model takes fiscal policy, G and T, monetary policy, M, and the price level P as exogenous. Given these exogenous variables, the *IS* curve provides the combinations of r and Y that satisfy the equation representing the goods market, and the *LM* curve provides the combinations of r and Y that satisfy the equation representing the money market. These two curves are shown together in Figure 10-14.

figure 10-14

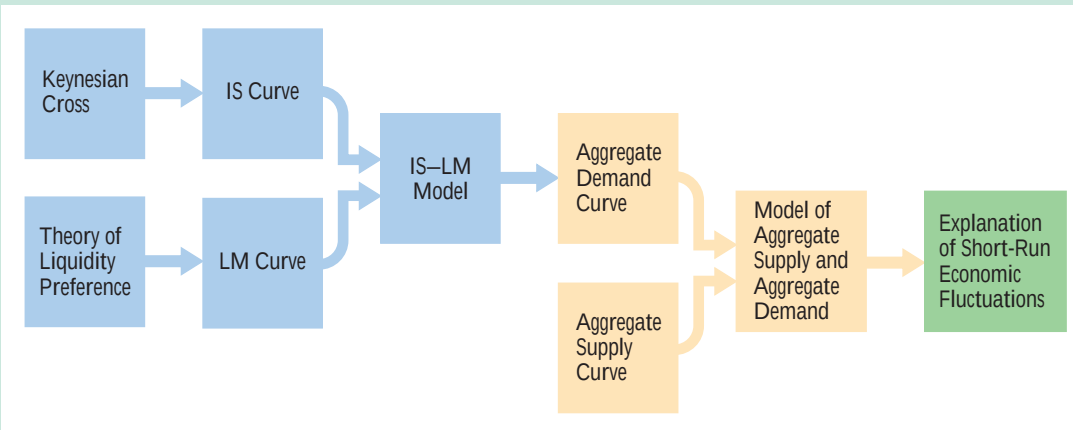


Equilibrium in the ***IS*–*LM*** Model The intersection of the *IS* and *LM* curves represents simultaneous equilibrium in the market for goods and services and in the market for real money balances for given values of government spending, taxes, the money supply, and the price level.

The equilibrium of the economy is the point at which the IS curve and the LM curve cross. This point gives the interest rate r and the level of income Y that satisfy conditions for equilibrium in both the goods market and the money market. In other words, at this intersection, actual expenditure equals planned expenditure, and the demand for real money balances equals the supply.

As we conclude this chapter, let us recall that our ultimate goal in developing the IS–LM model is to analyze short-run fluctuations in economic activity. Figure 10-15 illustrates how the different pieces of our theory fit together. In this chapter we developed the Keynesian cross and the theory of liquidity preference as building blocks for the IS–LM model. As we see more fully in the next chapter, the IS–LM model helps explain the position and slope of the aggregate demand curve. The aggregate demand curve, in turn, is a piece of the model of aggregate supply and aggregate demand, which economists use to explain the short-run effects of policy changes and other events on national income.

figure 10-15



The Theory of Short-Run Fluctuations This schematic diagram shows how the different pieces of the theory of short-run fluctuations fit together. The Keynesian cross explains the IS curve, and the theory of liquidity preference explains the LM curve. The IS and LM curves together yield the IS–LM model, which explains the aggregate demand curve. The aggregate demand curve is part of the model of aggregate supply and aggregate demand, which economists use to explain short-run fluctuations in economic activity.

Summary

1. The Keynesian cross is a basic model of income determination. It takes fiscal policy and planned investment as exogenous and then shows that there is one level of national income at which actual expenditure equals planned expenditure. It shows that changes in fiscal policy have a multiplied impact on income.

2. Once we allow planned investment to depend on the interest rate, the Keynesian cross yields a relationship between the interest rate and national income. A higher interest rate lowers planned investment, and this in turn lowers national income. The downward-sloping *IS* curve summarizes this negative relationship between the interest rate and income.
3. The theory of liquidity preference is a basic model of the determination of the interest rate. It takes the money supply and the price level as exogenous and assumes that the interest rate adjusts to equilibrate the supply and demand for real money balances. The theory implies that increases in the money supply lower the interest rate.
4. Once we allow the demand for real money balances to depend on national income, the theory of liquidity preference yields a relationship between income and the interest rate. A higher level of income raises the demand for real money balances, and this in turn raises the interest rate. The upward-sloping *LM* curve summarizes this positive relationship between income and the interest rate.
5. The *IS*–*LM* model combines the elements of the Keynesian cross and the elements of the theory of liquidity preference. The *IS* curve shows the points that satisfy equilibrium in the goods market, and the *LM* curve shows the points that satisfy equilibrium in the money market. The intersection of the *IS* and *LM* curves shows the interest rate and income that satisfy equilibrium in both markets.

KEY CONCEPTS

IS–*LM* model
IS curve
LM curve
Keynesian cross
Government-purchases multiplier
Tax multiplier
Theory of liquidity preference

QUESTIONS FOR REVIEW

1. Use the Keynesian cross to explain why fiscal policy has a multiplied effect on national income.
2. Use the theory of liquidity preference to explain why an increase in the money supply lowers the interest rate. What does this explanation assume about the price level?
3. Why does the *IS* curve slope downward?
4. Why does the *LM* curve slope upward?

PROBLEMS AND APPLICATIONS

1. Use the Keynesian cross to predict the impact of
   a. An increase in government purchases.
   b. An increase in taxes.
   c. An equal increase in government purchases and taxes.
2. In the Keynesian cross, assume that the consumption function is given by

$$C = 200 + 0.75 (Y - T).$$

Planned investment is 100; government purchases and taxes are both 100.

a. Graph planned expenditure as a function of income.

b. What is the equilibrium level of income?

c. If government purchases increase to 125, what is the new equilibrium income?

d. What level of government purchases is needed to achieve an income of 1,600?

3. Although our development of the Keynesian cross in this chapter assumes that taxes are fixed by a amount, in many countries (including the United States) taxes depend on income. Let us represent the tax system by writing tax revenue as

$$T = \bar{T} + tY,$$

where $\bar{T}$ and t are parameters of the tax code. The parameter t is the marginal tax rate: if income rises by \$1, taxes rise by t by \$1.

a. How does this tax system change the way consumption responds to changes in GDP?

b. In the Keynesian cross, how does this tax system alter the government-purchases multiplier?

c. In the IS–LM model, how does this tax system alter the slope of the IS curve?

4. Consider the impact of an increase in thriftiness in the Keynesian cross. Suppose the consumption function is

$$C = \bar{C} + c(Y - T),$$

where $\bar{C}$ is a parameter called autonomous consumption and c is the marginal propensity to consume.

a. What happens to equilibrium income when the society becomes more thrifty, as represented by a decline in $\bar{C}$?

b. What happens to equilibrium saving?

c. Why do you suppose this result is called the paradox of thrift?

d. Does this paradox arise in the classical model of Chapter 3? Why or why not?

5. Suppose that the money demand function is

$$(M/P)^d = 1{,}000 - 100r,$$

where r is the interest rate in percent. The money supply M is 1,000 and the price level P is 2.

a. Graph the supply and demand for real money balances.

b. What is the equilibrium interest rate?

c. Assume that the price level is fixed. What happens to the equilibrium interest rate if the supply of money is raised from 1,000 to 1,200?

d. If the Fed wishes to raise the interest rate to 7 percent, what money supply should it set?

CHAPTER 11 ELEVEN

Aggregate Demand II

> Science is a parasite: the greater the patient population the better the advance in physiology and pathology; and out of pathology arises therapy. The year 1932 was the trough of the great depression, and from its rotten soil was belatedly begot a new subject that today we call macroeconomics.
>
> — Paul Samuelson

In Chapter 10 we assembled the pieces of the IS–LM model. We saw that the IS curve represents the equilibrium in the market for goods and services, that the LM curve represents the equilibrium in the market for real money balances, and that the IS and LM curves together determine the interest rate and national income in the short run when the price level is fixed. Now we turn our attention to applying the IS–LM model to analyze three issues.

First, we examine the potential causes of fluctuations in national income. We use the IS–LM model to see how changes in the exogenous variables (government purchases, taxes, and the money supply) influence the endogenous variables (the interest rate and national income). We also examine how various shocks to the goods markets (the IS curve) and the money market (the LM curve) affect the interest rate and national income in the short run.

Second, we discuss how the IS–LM model fits into the model of aggregate supply and aggregate demand we introduced in Chapter 9. In particular, we examine how the IS–LM model provides a theory of the slope and position of the aggregate demand curve. Here we relax the assumption that the price level is fixed, and we show that the IS–LM model implies a negative relationship between the price level and national income. The model can also tell us what events shift the aggregate demand curve and in what direction.

Third, we examine the Great Depression of the 1930s. As this chapter's opening quotation indicates, this episode gave birth to short-run macroeconomic theory, for it led Keynes and his many followers to think that aggregate demand was the key to understanding fluctuations in national income. With the benefit of hindsight, we can use the IS–LM model to discuss the various explanations of this traumatic economic downturn.

11-1 Explaining Fluctuations With the *IS–LM* Model

The intersection of the IS curve and the LM curve determines the level of national income. When one of these curves shifts, the short-run equilibrium of the economy changes, and national income fluctuates. In this section we examine how changes in policy and shocks to the economy can cause these curves to shift.

How Fiscal Policy Shifts the *IS* Curve and Changes the Short-Run Equilibrium

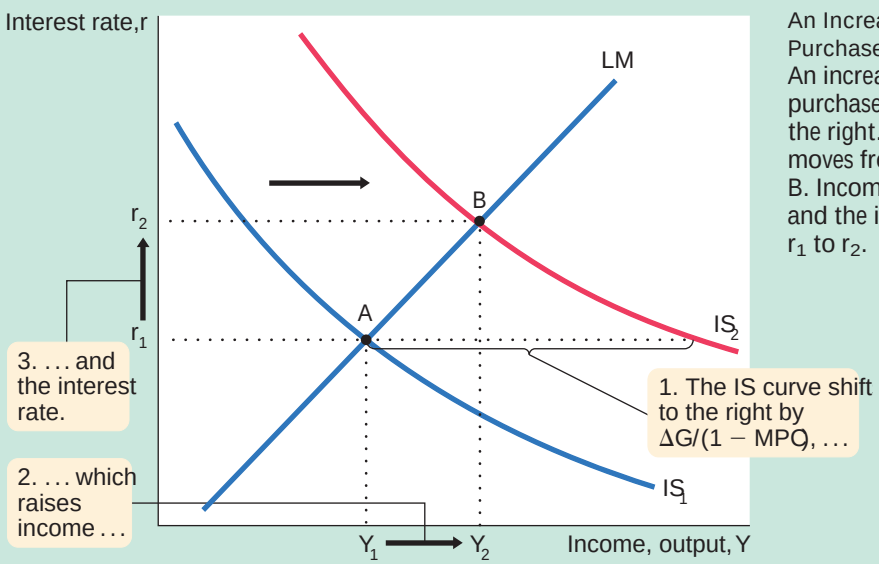
We begin by examining how changes in fiscal policy (government purchases and taxes) alter the economy's short-run equilibrium. Recall that changes in fiscal policy influence planned expenditure and thereby shift the IS curve. The IS–LM model shows how these shifts in the IS curve affect income and the interest rate.

Changes in Government Purchases

Consider an increase in government purchases of ΔG. The government-purchases multiplier in the Keynesian cross tells us that, at any given interest rate, this change in fiscal policy raises the level of income by $\Delta G/(1 - MPC)$. Therefore, as Figure 11-1 shows, the IS curve shifts to the right by this amount. The equilibrium of the economy moves from point A to point B. The increase in government purchases raises both income and the interest rate.

To understand fully what is happening in Figure 11-1, it helps to keep in mind the building blocks for the IS–LM model from the preceding chapter—the Keynesian cross and the theory of liquidity preference. Here is the story. When the government

figure 11-1



An Increase in Government Purchases in the *IS–LM* Model
An increase in government purchases shifts the IS curve to the right. The equilibrium moves from point A to point B. Income rises from Y_1 to Y_2, and the interest rate rises from r_1 to r_2.