For example, in some union contracts in the United States, nominal wages are set in advance for three years. Unions and firms have to decide what nominal wages will be over the following three years based on what they expect the price level to be over those three years. Even when wages are set by firms, or by bargaining between the firm and each worker, nominal wages are typically set for a year. If the price level goes up unexpectedly during the year, nominal wages are typically not readjusted. (How workers and firms form expectations of the price level will occupy us for much of the next three chapters; we will leave this issue aside for the moment.)

**The Unemployment Rate**

Also affecting the aggregate wage in equation (6.1) is the unemployment rate $u$. The minus sign under $u$ indicates that an increase in the unemployment rate decreases wages.

The fact that wages depend on the unemployment rate was one of the main conclusions of our earlier discussion. If we think of wages as being determined by bargaining, then higher unemployment weakens workers’ bargaining power, forcing them to accept lower wages. If we think of wages as being determined by efficiency wage considerations, then higher unemployment allows firms to pay lower wages and still keep workers willing to work.

**The Other Factors**

The third variable in equation (6.1), $z$, is a catchall variable that stands for all the factors that affect wages given the expected price level and the unemployment rate. By convention, we will define $z$ so that an increase in $z$ implies an increase in the wage (thus, the positive sign under $z$ in the equation). Our earlier discussion suggests a long list of potential factors here.

Take, for example, unemployment insurance—the payment of unemployment benefits to workers who lose their jobs. There are very good reasons why society should provide some insurance to workers who lose their job and have a hard time finding another. But there is little question that, by making the prospects of unemployment less distressing, more generous unemployment benefits do increase wages at a given unemployment rate. To take an extreme example, suppose unemployment insurance did not exist. Some workers would have little to live on and would be willing to accept very low wages to avoid remaining unemployed. But unemployment insurance does exist, and it allows unemployed workers to hold out for higher wages. In this case, we can think of $z$ as representing the level of unemployment benefits: At a given unemployment rate, higher unemployment benefits increase the wage.

It is easy to think of other factors. An increase in the minimum wage may increase not only the minimum wage itself, but also wages just above the minimum wage, leading to an increase in the average wage, $W$, at a given unemployment rate. Or take an increase in employment protection, which makes it more expensive for firms to lay off workers. Such a change is likely to increase the bargaining power of workers covered by this protection (laying them off and hiring other workers is now more costly for firms), increasing the wage for a given unemployment rate.

We will explore some of these factors as we go along.
6-4 Price Determination

Having looked at wage determination, let’s now turn to price determination.

The prices set by firms depend on the costs they face. These costs depend, in turn, on the nature of the production function—the relation between the inputs used in production and the quantity of output produced—and on the prices of these inputs.

For the moment, we will assume firms produce goods using labor as the only factor of production. We will write the production function as follows:

\[ Y = AN \]

where \( Y \) is output, \( N \) is employment, and \( A \) is labor productivity. This way of writing the production function implies that labor productivity—output per worker—is constant and equal to \( A \).

It should be clear that this is a strong simplification. In reality, firms use other factors of production in addition to labor. They use capital—machines and factories. They use raw materials—oil, for example. Moreover, there is technological progress, so that labor productivity \((A)\) is not constant but steadily increases over time. We shall introduce these complications later. We will introduce raw materials in Chapter 7 when we discuss changes in the price of oil. We will focus on the role of capital and technological progress when we turn to the determination of output in the long run in Chapters 10 through 13. For the moment, though, this simple relation between output and employment will make our lives easier and still serve our purposes.

Given the assumption that labor productivity, \( A \), is constant, we can make one further simplification. We can choose the units of output so that one worker produces one unit of output—in other words, so that \( A = 1 \). (This way we do not have to carry the letter \( A \) around, and this will simplify notation.) With this assumption, the production function becomes

\[ Y = N \] (6.2)

The production function \( Y = N \) implies that the cost of producing one more unit of output is the cost of employing one more worker, at wage \( W \). Using the terminology introduced in your microeconomics course: The marginal cost of production—the cost of producing one more unit of output—is equal to \( W \).

If there were perfect competition in the goods market, the price of a unit of output would be equal to marginal cost: \( P \) would be equal to \( W \). But many goods markets are not competitive, and firms charge a price higher than their marginal cost. A simple way of capturing this fact is to assume that firms set their price according to

\[ P = (1 + m)W \] (6.3)

where \( m \) is the markup of the price over the cost. If goods markets were perfectly competitive, \( m \) would be equal to zero, and the price \( P \) would simply equal the cost \( W \). To the extent they are not competitive and firms have market power, \( m \) is positive, and the price \( P \) will exceed the cost \( W \) by a factor equal to \((1 + m)\).

6-5 The Natural Rate of Unemployment

Let’s now look at the implications of wage and price determination for unemployment.

For the rest of this chapter, we shall do so under the assumption that nominal wages depend on the actual price level, \( P \), rather than on the expected price level, \( P^e \) (why we make this assumption will become clear soon). Under this additional
assumption, wage setting and price setting determine the equilibrium (also called “natural”) rate of unemployment. Let’s see how.

The Wage-Setting Relation

Given the assumption that nominal wages depend on the actual price level \(P\) rather than on the expected price level \(P^e\), equation (6.1), which characterizes wage determination, becomes:

\[
W = P F( u, z )
\]

Dividing both sides by the price level,

\[
\frac{W}{P} = F( u, z )
\]

Wage determination implies a negative relation between the real wage, \(W/P\), and the unemployment rate, \(u\): The higher the unemployment rate, the lower the real wage chosen by wage setters. The intuition is straightforward: The higher the unemployment rate, the weaker the workers’ bargaining position, and the lower the real wage will be.

This relation between the real wage and the rate of unemployment—let’s call it the wage-setting relation—is drawn in Figure 6-6. The real wage is measured on the vertical axis. The unemployment rate is measured on the horizontal axis. The wage-setting relation is drawn as the downward–sloping curve \(WS\) (for wage setting): The higher the unemployment rate, the lower the real wage.

The Price–Setting Relation

Let’s now look at the implications of price determination. If we divide both sides of the price–determination equation, (6.3), by the nominal wage, we get

\[
\frac{P}{W} = 1 + m
\]

The ratio of the price level to the wage implied by the price-setting behavior of firms equals 1 plus the markup. Now invert both sides of this equation to get the implied real wage:

\[
\frac{W}{P} = \frac{1}{1 + m}
\]

Note what this equation says: Price-setting decisions determine the real wage paid by firms. An increase in the markup leads firms to increase their prices given the wage they have to pay; equivalently, it leads to a decrease in the real wage.

The step from equation (6.5) to equation (6.6) is algebraically straightforward. But how price setting actually determines the real wage paid by firms may not be intuitively obvious. Think of it this way: Suppose the firm you work for increases its markup and therefore increases the price of its product. Your real wage does not change very much: You are still paid the same nominal wage, and the product produced by the firm is at most a small part of your consumption basket.

Now suppose that not only the firm you work for, but all the firms in the economy increase their markup. All the prices go up. Even if you are paid the same nominal wage,

“Wage setters”: Unions and firms if wages are set by collective bargaining; individual workers and firms if wages are set on a case-by-case basis; firms if wages are set on a take-it-or-leave-it basis.
wage, your real wage goes down. So, the higher the markup set by firms, the lower your (and everyone else’s) real wage will be. This is what equation (6.6) says.

The price-setting relation in equation (6.6) is drawn as the horizontal line PS (for price setting) in Figure 6-6. The real wage implied by price setting is \( \frac{1}{1 + m} \); it does not depend on the unemployment rate.

**Equilibrium Real Wages and Unemployment**

Equilibrium in the labor market requires that the real wage chosen in wage setting be equal to the real wage implied by price setting. (This way of stating equilibrium may sound strange if you learned to think in terms of labor supply and labor demand in your microeconomics course. The relation between wage setting and price setting, on the one hand, and labor supply and labor demand, on the other, is closer than it looks at first and is explored further in the appendix at the end of this chapter.) In Figure 6-6, equilibrium is therefore given by point A, and the equilibrium unemployment rate is given by \( u_n \).

We can also characterize the equilibrium unemployment rate algebraically; eliminating \( W/P \) between equations (6.4) and (6.6) gives

\[
F(u_n, z) = \frac{1}{1 + m}
\]  

(6.7)

The equilibrium unemployment rate, \( u_n \), is such that the real wage chosen in wage setting—the left side of equation (6.7)—is equal to the real wage implied by price setting—the right side of equation (6.7).

The equilibrium unemployment rate, \( u_n \), is called the natural rate of unemployment (which is why we have used the subscript \( n \) to denote it). The terminology has become standard, so we shall adopt it, but this is actually a bad choice of words. The word “natural” suggests a constant of nature, one that is unaffected by institutions and policy. As its derivation makes clear, however, the “natural” rate of unemployment is anything but natural. The positions of the wage-setting and price-setting curves, and thus the equilibrium unemployment rate, depend on both \( z \) and \( m \). Consider two examples:

- An increase in unemployment benefits. An increase in unemployment benefits can be represented by an increase in \( z \): Since an increase in benefits makes the
prospect of unemployment less painful, it increases the wage set by wage setters at a given unemployment rate. So it shifts the wage-setting relation up, from $WS$ to $WS'$ in Figure 6-7. The economy moves along the $PS$ line, from $A$ to $A'$. The natural rate of unemployment increases from $u_n$ to $u'_n$.

In words: At a given unemployment rate, higher unemployment benefits lead to a higher real wage. A higher unemployment rate is needed to bring the real wage back to what firms are willing to pay.

- A less stringent enforcement of existing antitrust legislation. To the extent that this allows firms to collude more easily and increase their market power, it will lead to an increase in their markup—an increase in $m$. The increase in $m$ implies a decrease in the real wage paid by firms, and so it shifts the price-setting relation down, from $PS$ to $PS'$ in Figure 6-8. The economy moves along $WS$. The equilibrium moves from $A$ to $A'$, and the natural rate of unemployment increases from $u_n$ to $u'_n$.

In words: By letting firms increase their prices given the wage, less stringent enforcement of antitrust legislation leads to a decrease in the real wage. Higher unemployment is required to make workers accept this lower real wage, leading to an increase in the natural rate of unemployment.

Factors like the generosity of unemployment benefits or antitrust legislation can hardly be thought of as the result of nature. Rather, they reflect various characteristics of the structure of the economy. For that reason, a better name for the equilibrium rate of unemployment would be the **structural rate of unemployment**, but so far the name has not caught on.

### From Unemployment to Employment

Associated with the natural rate of unemployment is a **natural level of employment**, the level of employment that prevails when unemployment is equal to its natural rate.
Let’s review the relation among unemployment, employment, and the labor force. Let $U$ denote unemployment, $N$ denote employment, and $L$ the labor force. Then:

$$u = \frac{U}{L} = \frac{L - N}{L} = 1 - \frac{N}{L}$$

The first step follows from the definition of the unemployment rate ($u$). The second follows from the fact that, from the definition of the labor force, the level of unemployment ($U$) equals the labor force ($L$) minus employment ($N$). The third step follows from simplifying the fraction. Putting all three steps together: The unemployment rate $u$ equals 1 minus the ratio of employment $N$ to the labor force $L$.

Rearranging to get employment in terms of the labor force and the unemployment rate gives:

$$N = L(1 - u)$$

Employment $N$ is equal to the labor force $L$, times 1 minus the unemployment rate $u$. So, if the natural rate of unemployment is $u_n$ and the labor force is equal to $L$, the natural level of employment $N_n$ is given by

$$N_n = L(1 - u_n)$$

For example, if the labor force is 150 million and the natural rate of unemployment is, say, 5%, then the natural level of employment is $150 \times (1 - 0.05) = 142.5$ million.

**From Employment to Output**

Finally, associated with the natural level of employment is the natural level of output, the level of production when employment is equal to the natural level of employment. Given the production function we have used in this chapter ($Y = N$), the natural level of output $Y_n$ is easy to derive. It is given by

$$Y_n = N_n = L(1 - u_n)$$
Using equation (6.7) and the relations among the unemployment rate, employment, and the output we just derived, the natural level of output satisfies the following equation:

\[
F \left( 1 - \frac{Y_n}{L} \right) z = \frac{1}{1 + m} \tag{6.8}
\]

The natural level of output \( (Y_n) \) is such that, at the associated rate of unemployment \( (u_n = 1 - Y_n/L) \), the real wage chosen in wage setting—the left side of equation (6.8)—is equal to the real wage implied by price setting—the right side of equation (6.8). As you will see, equation (6.8) will turn out to be very useful in the next chapter. Make sure you understand it.

We have gone through many steps in this section. Let’s summarize:

- The real wage chosen in wage setting is a decreasing function of the unemployment rate.
- The real wage implied by price setting is constant.
- Equilibrium in the labor market requires that the real wage chosen in wage setting be equal to the real wage implied by price setting. This determines the equilibrium unemployment rate.
- This equilibrium unemployment rate is known as the natural rate of unemployment.
- Associated with the natural rate of unemployment is a natural level of employment and a natural level of output.

## 6-6 Where We Go from Here

We have just seen how equilibrium in the labor market determines the equilibrium unemployment rate (we have called it the natural rate of unemployment), which in turn determines the level of output (we have called it the natural level of output).

So, you may ask, what did we do in the previous four chapters? If equilibrium in the labor market determines the unemployment rate and, by implication, the level of output, why did we spend so much time looking at the goods and financial markets? What about our earlier conclusions that the level of output was determined by factors such as monetary policy, fiscal policy, consumer confidence, and so on—all factors that do not enter equation (6.8) and therefore do not affect the natural level of output?

The key to the answer lies in the difference between the short run and the medium run:

- We have derived the natural rate of unemployment and the associated levels of employment and output under two assumptions. First, we have assumed equilibrium in the labor market. Second, we have assumed that the price level was equal to the expected price level.
- However, there is no reason for the second assumption to be true in the short run. The price level may well turn out to be different from what was expected when nominal wages were set. Hence, in the short run, there is no reason for unemployment to be equal to the natural rate or for output to be equal to its natural level.

As we will see in the next chapter, the factors that determine movements in output in the short run are indeed the factors we focused on in the preceding three chapters: monetary policy, fiscal policy, and so on. Your time (and mine) was not wasted.
In the medium run, output tends to return to the natural level, and the factors that determine output are the factors we have focused on in this chapter.

But expectations are unlikely to be systematically wrong (say, too high or too low) forever. That is why, in the medium run, unemployment tends to return to the natural rate, and output tends to return to the natural level. *In the medium run*, the factors that determine unemployment and output are the factors that appear in equations (6.7) and (6.8).

These, in short, are the answers to the questions asked in the first paragraph of this chapter. Developing these answers in detail will be our task in the next three chapters.

**Summary**

- The labor force consists of those who are working (employed) or looking for work (unemployed). The unemployment rate is equal to the ratio of the number of unemployed to the number in the labor force. The participation rate is equal to the ratio of the labor force to the working-age population.
- The U.S. labor market is characterized by large flows between employment, unemployment, and "out of the labor force." On average, each month, about 47% of the unemployed move out of unemployment, either to take a job or to drop out of the labor force.
- Unemployment is high in recessions and low in expansions. During periods of high unemployment, the probability of losing a job increases and the probability of finding a job decreases.
- Wages are set unilaterally by firms or by bargaining between workers and firms. They depend negatively on the unemployment rate and positively on the expected price level. The reason why wages depend on the expected price level is that they are typically set in nominal terms for some period of time. During that time, even if the price level turns out to be different from what was expected, wages are typically not readjusted.
- The price set by firms depends on the wage and on the markup of prices over wages. The higher the markup chosen by firms, the higher the price given the wage, and thus the lower the real wage implied by price-setting decisions.
- Equilibrium in the labor market requires that the real wage chosen in wage setting be equal to the real wage implied by price setting. Under the additional assumption that the expected price level is equal to the actual price level, equilibrium in the labor market determines the unemployment rate. This unemployment rate is known as the **natural rate of unemployment**.
- In general, the actual price level may turn out to be different from the price level expected by wage setters. Therefore, the unemployment rate need not be equal to the natural rate.
- The coming chapters will show that:
  - In the short run, unemployment and output are determined by the factors we focused on in the previous chapters.
  - But, in the medium run, unemployment tends to return to the natural rate, and output tends to return to its natural level.

**Key Terms**

- noninstitutional civilian population, 112
- labor force; out of the labor force, 112
- participation rate, 112
- unemployment rate, 112
- separations, hires, 113
- Current Population Survey (CPS), 113
- quits, layoffs, 113
- duration of unemployment, 114
- discouraged workers, 115
- employment rate, 115
- collective bargaining, 117
- reservation wage, 118
- bargaining power, 118
- efficiency wage theories, 119
- unemployment insurance, 121
- employment protection, 121
- production function, 122
- labor productivity, 122
- markup, 122
- wage-setting relation, 123
- price-setting relation, 124
- natural rate of unemployment, 124
- structural rate of unemployment, 125
- natural level of employment, 125
- natural level of output, 126
Questions and Problems

QUICK CHECK
All Quick Check questions and problems are available on MyEconLab.
1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. Since 1950, the participation rate in the United States has remained roughly constant at 60%.
   b. Each month, the flows into and out of employment are very small compared to the size of the labor force.
   c. Fewer than 10% of all unemployed workers exit the unemployment pool each year.
   d. The unemployment rate tends to be high in recessions and low in expansions.
   e. Most workers are typically paid their reservation wage.
   f. Workers who do not belong to unions have no bargaining power.
   g. It may be in the best interest of employers to pay wages higher than their workers’ reservation wage.
   h. The natural rate of unemployment is unaffected by policy changes.

2. Answer the following questions using the information provided in this chapter.
   a. As a percentage of the employed workers, what is the size of the flows into and out of employment (i.e., hires and separations) each month?
   b. As a percentage of the unemployed workers, what is the size of the flows from unemployment into employment each month?
   c. As a percentage of the unemployed, what is the size of the total flows out of unemployment each month? What is the average duration of unemployment?
   d. As a percentage of the labor force, what is the size of the total flows into and out of the labor force each month?
   e. In the text we say that there is an average of 400,000 new workers entering the labor force each month. What percentage of total flows into the labor force do new workers entering the labor force constitute?

3. The natural rate of unemployment
   Suppose that the markup of goods prices over marginal cost is 5%, and that the wage-setting equation is
   \[ W = P (1 - u), \]
   where \( u \) is the unemployment rate.
   a. What is the real wage, as determined by the price-setting equation?
   b. What is the natural rate of unemployment?
   c. Suppose that the markup of prices over costs increases to 10%. What happens to the natural rate of unemployment? Explain the logic behind your answer.

DIG DEEPER
All Dig Deeper questions and problems are available on MyEconLab.
4. Reservation wages
   In the mid-1980s, a famous supermodel once said that she would not get out of bed for less than $10,000 (presumably per day).
In the first economy, EatIn, the 25 food-preparation workers (one per household) cook for their families and do not work outside the home. All meals are prepared and eaten at home. The 25 food-preparation workers in this economy do not seek work in the formal labor market (and when asked, they say they are not looking for work). In the second economy, EatOut, the 25 food-preparation workers are employed by restaurants. All meals are purchased in restaurants.

a. Calculate measured employment and unemployment and the measured labor force for each economy. Calculate the measured unemployment rate and participation rate for each economy. In which economy is measured GDP higher?

b. Suppose now that EatIn’s economy changes. A few restaurants open, and the food preparation workers in 10 households take jobs restaurants. The members of these 10 households now eat all of their meals in restaurants. The food-preparation workers in the remaining 15 households continue to work at home and do not seek jobs in the formal sector. The members of these 15 households continue to eat all of their meals at home. Without calculating the numbers, what will happen to measured employment and unemployment and to the measured labor force, unemployment rate, and participation rate in EatIn? What will happen to measured GDP in EatIn?

c. Suppose that you want to include work at home in GDP and the employment statistics. How would you measure the value of work at home in GDP? How would you alter the definitions of employment, unemployment, and out of the labor force?

d. Given your new definitions in part (c), would the labor-market statistics differ for EatIn and EatOut? Assuming that the food produced by these economies has the same value, would measured GDP in these economies differ? Under your new definitions, would the experiment in part (b) have any effect on the labor market or GDP statistics for EatIn?

EXPLORE FURTHER

8. Unemployment spells and long-term unemployment

According to the data presented in this chapter, about 47% of unemployed workers leave unemployment each month.

a. What is the probability that an unemployed worker will still be unemployed after one month? two months? six months?

Now consider the composition of the unemployment pool. We will use a simple experiment to determine the proportion of the unemployed who have been unemployed six months or more. Suppose the number of unemployed workers is constant and equal to $x$ (where $x$ is some constant). Each month, 47% of the unemployed find jobs, and an equivalent number of previously employed workers become unemployed.

b. Consider the group of $x$ workers who are unemployed this month. After a month, what percentage of this group will still be unemployed? (Hint: If 47% of unemployed workers find jobs every month, what percentage of the original $x$ unemployed workers did not find jobs in the first month?)

c. After a second month, what percentage of the original $x$ unemployed workers has been unemployed for at least two months? (Hint: Given your answer to part (b), what percentage of those unemployed for at least one month do not find jobs in the second month?) After the sixth month, what percentage of the original $x$ unemployed workers has been unemployed for at least six months?

d. Using Table B-44 of the Economic Report of the President (www.gpoaccess.gov/eop/), compute the proportion of unemployed who have been unemployed six months or more (27 weeks or more) for each year between 1996 and 2010. How do these numbers compare with the answer you obtained in part (c)? Can you guess what may account for the difference between the actual numbers and the answer you obtained in this problem? (Hint: Suppose that the probability of exiting unemployment goes down the longer you are unemployed.)

e. Part of the policy response to the crisis was an extension of the length of time that an unemployed worker could receive unemployment benefits. How would you predict this change would affect the proportion of those unemployed more than six months?


a. What are the latest monthly data on the size of the U.S. civilian labor force, on the number of unemployed, and on the unemployment rate?

b. How many people are employed?

c. Compute the change in the number of unemployed from the first number in the table to the most recent month in the table. Do the same for the number of employed workers. Is the decline in unemployment equal to the increase in employment? Explain in words.

10. The typical dynamics of unemployment over a recession.

The table below shows the behavior of annual real GDP growth during three recessions. These data are from Table B-4 of the Economic Report of the President:

<table>
<thead>
<tr>
<th>Year</th>
<th>Real GDP Growth</th>
<th>Unemployment Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td>–1.9</td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>–0.2</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>–2.6</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>2.9</td>
<td></td>
</tr>
</tbody>
</table>

Use Table B-35 from the Economic Report of the President to fill in the annual values of the unemployment rate in the table above and consider these questions.

a. When is the unemployment rate in a recession higher, the year of declining output or the following year?
b. Explain the pattern of the unemployment rate after a recession if the production function is not linear in the workforce.
c. Explain the pattern of the unemployment rate after a recession if discouraged workers return to the labor force as the economy recovers.
d. The rate of unemployment remains substantially higher after the crisis-induced recession in 2009. In that recession, unemployment benefits were extended in length from 6 months to 12 months. What does the model predict the effect of this policy will be on the natural rate of unemployment? Do the data support this prediction in any way?

Further Reading

A further discussion of unemployment along the lines of this chapter is given by Richard Layard, Stephen Nickell, and Richard Jackman in The Unemployment Crisis (Oxford: Oxford University Press, 1994).

APPENDIX: Wage- and Price-Setting Relations versus Labor Supply and Labor Demand

If you have taken a microeconomics course, you probably saw a representation of labor-market equilibrium in terms of labor supply and labor demand. You may therefore be asking yourself: How does the representation in terms of wage setting and price setting relate to the representation of the labor market I saw in that course?

In an important sense, the two representations are similar:

To see why, let’s redraw Figure 6-6 in terms of the real wage on the vertical axis, and the level of employment (rather than the unemployment rate) on the horizontal axis. We do this in Figure 1.

Employment, \( N \), is measured on the horizontal axis. The level of employment must be somewhere between zero and \( L \), the labor force: Employment cannot exceed the number of people available for work, (i.e., the labor force). For any employment level \( N \), unemployment is given by \( U = L - N \). Knowing this, we can measure unemployment by starting from \( L \) and moving to the left on the horizontal axis: Unemployment is given by the distance between \( L \) and \( N \). The lower is employment, \( N \), the higher is unemployment, and by implication the higher is the unemployment rate, \( u \).

Let’s now draw the wage-setting and price-setting relations and characterize the equilibrium:

- An increase in employment (a movement to the right along the horizontal axis) implies a decrease in unemployment and therefore an increase in the real wage chosen in wage setting. Thus, the wage-setting relation is now upward sloping: Higher employment implies a higher real wage.
- The price-setting relation is still a horizontal line at \( W/P = 1/(1 + m) \).

The equilibrium is given by point \( A \), with “natural” employment level \( N_n \) (and an implied natural unemployment rate equal to \( u_n = (L - N_n)/L \)).

In this figure the wage-setting relation looks like a labor-supply relation. As the level of employment increases, the real wage paid to workers increases as well. For that reason, the wage-setting relation is sometimes called the “labor-supply” relation (in quotes).

Figure 1

Wage and Price Setting and the Natural Level of Employment
What we have called the price-setting relation looks like a flat labor-demand relation. The reason it is flat rather than downward sloping has to do with our simplifying assumption of constant returns to labor in production. Had we assumed, more conventionally, that there were decreasing returns to labor in production, our price-setting curve would, like the standard labor-demand curve, be downward sloping: As employment increased, the marginal cost of production would increase, forcing firms to increase their prices given the wages they pay. In other words, the real wage implied by price setting would decrease as employment increased.

But, in a number of ways, the two approaches are different:

- The standard labor-supply relation gives the wage at which a given number of workers are willing to work: The higher the wage, the larger the number of workers who are willing to work.

In contrast, the wage corresponding to a given level of employment in the wage-setting relation is the result of a process of bargaining between workers and firms, or unilateral wage setting by firms. Factors like the structure of collective bargaining or the use of wages to deter quits affect the wage-setting relation. In the real world, they seem to play an important role. Yet they play no role in the standard labor-supply relation.

- The standard labor-demand relation gives the level of employment chosen by firms at a given real wage. It is derived under the assumption that firms operate in competitive goods and labor markets and therefore take wages and prices—and by implication the real wage—as given.

In contrast, the price-setting relation takes into account the fact that in most markets firms actually set prices. Factors such as the degree of competition in the goods market affect the price-setting relation by affecting the markup. But these factors aren’t considered in the standard labor-demand relation.

- In the labor supply–labor demand framework, those unemployed are willingly unemployed: At the equilibrium real wage, they prefer to be unemployed rather than work.

In contrast, in the wage setting–price setting framework, unemployment is likely to be involuntary. For example, if firms pay an efficiency wage—a wage above the reservation wage—workers would rather be employed than unemployed. Yet, in equilibrium, there is still involuntary unemployment. This also seems to capture reality better than does the labor supply–labor demand framework.

These are the three reasons why we have relied on the wage-setting and the price-setting relations rather than on the labor supply–labor demand approach to characterize equilibrium in this chapter.
In Chapter 5, we looked at the determination of output in the short run. In Chapter 6, we looked at the determination of output in the medium run. We are now ready to put the two together and look at the determination of output in both the short run and the medium run.

To do so, we use the equilibrium conditions for all the markets we have looked at so far—the goods and financial markets in Chapter 5, the labor market in Chapter 6.

Then, using these equilibrium conditions, we derive two relations:

The first relation, which we call the aggregate supply relation, captures the implications of equilibrium in the labor market; it builds on what you saw in Chapter 6.

The second relation, which we call the aggregate demand relation, captures the implications of equilibrium in both the goods market and financial markets; it builds on what you saw in Chapter 5.

Combining these two relations gives us the AS–AD model (for aggregate supply–aggregate demand). This chapter presents the basic version of the model. When confronted with a macroeconomic question, this is the version we typically use to organize our thoughts. For some questions—if we want to focus on the behavior of inflation, for example, or understand the role of the financial system in the current crisis—the basic AS–AD model must be extended. But it provides a base on which one can build, and this is what we shall do in the next two chapters.

This chapter is organized as follows:

Section 7-1 derives the aggregate supply relation, and Section 7-2 derives the aggregate demand relation.

Section 7-3 combines the two to characterize equilibrium output in the short run and in the medium run.

Section 7-4 looks at the dynamic effects of monetary policy.

Section 7-5 looks at the dynamic effects of fiscal policy.

Section 7-6 looks at the effects of an increase in the price of oil.

Section 7-7 summarizes.
7-1 Aggregate Supply

The aggregate supply relation captures the effects of output on the price level. It is derived from the behavior of wages and prices we described in Chapter 6.

In Chapter 6, we derived the following equation for wage determination (equation (6.1)):

\[ W = P^e F(u, z) \]

The nominal wage \( W \), set by wage setters, depends on the expected price level \( P^e \), on the unemployment rate \( u \), and on the catchall variable \( z \) for all the other factors that affect wage determination, from unemployment benefits to the form of collective bargaining.

Also in Chapter 6, we derived the following equation for price determination (equation (6.3)):

\[ P = (1 + m)W \]

The price \( P \) set by firms (equivalently, the price level) is equal to the nominal wage \( W \), times 1 plus the markup \( m \).

We then used these two relations together with the additional assumption that the actual price level was equal to the expected price level. Under this additional assumption, we derived the natural rate of unemployment and, by implication, the natural level of output.

The difference in this chapter is that we will not impose this additional assumption. (It will turn out that the price level is equal to the expected price level in the medium run but will typically not be equal to the expected price level in the short run.) Without this additional assumption, the price-setting relation and the wage-setting relation give us a relation, which we now derive, among the price level, the output level, and the expected price level.

The first step is to eliminate the nominal wage \( W \) between the two equations. Replacing the nominal wage in the second equation above by its expression from the first gives

\[ P = P^e (1 + m) F(u, z) \]  

(7.1)

The price level \( P \) depends on the expected price level \( P^e \), on the unemployment rate \( u \) (as well as on the markup \( m \) and on the catchall variable \( z \); but we will assume both \( m \) and \( z \) are constant here).

The second step is to replace the unemployment rate \( u \) with its expression in terms of output. To replace \( u \), recall the relation between the unemployment rate, employment, and output we derived in Chapter 6:

\[ u = \frac{U}{L} = \frac{L - N}{L} = 1 - \frac{N}{L} = 1 - \frac{Y}{L} \]

The first equality follows from the definition of the unemployment rate. The second equality follows from the definition of unemployment \( (U = L - N) \). The third equality just simplifies the fraction. The fourth equality follows from the specification of the production function, which says that to produce one unit of output requires one worker, so that \( Y = N \). What we get then is

\[ u = 1 - \frac{Y}{L} \]

In words: For a given labor force, the higher the output, the lower the unemployment rate.

Replacing \( u \) by \( 1 - (Y/L) \) in equation (7.1) gives us the aggregate supply relation, or AS relation for short:

\[ P = P^e (1 + m) F\left(1 - \frac{Y}{L}, z\right) \]  

(7.2)

A better name would be “the labor market relation.” But because the relation looks graphically like a supply curve (there is a positive relation between output and the price), it is called “the aggregate supply relation.” we follow this tradition.
The price level $P$ depends on the expected price level $P_e$ and the level of output $Y$ (and also on the markup $m$, the catchall variable $z$, and the labor force $L$, which we all take as constant here). The AS relation has two important properties:

The first property is that, given the expected price level, an increase in output leads to an increase in the price level. This is the result of four underlying steps:

1. An increase in output leads to an increase in employment.
2. The increase in employment leads to a decrease in unemployment and therefore to a decrease in the unemployment rate.
3. The lower unemployment rate leads to an increase in the nominal wage.
4. The increase in the nominal wage leads to an increase in the prices set by firms and therefore to an increase in the price level.

The second property is that, given unemployment, an increase in the expected price level leads, one for one, to an increase in the actual price level. For example, if the expected price level doubles, then the price level will also double. This effect works through wages:

1. If wage setters expect the price level to be higher, they set a higher nominal wage.
2. The increase in the nominal wage leads to an increase in costs, which leads to an increase in the prices set by firms and a higher price level.

If output is equal to the natural level of output, the price level is equal to the expected price level.

The relation between the price level $P$ and output $Y$, for a given value of the expected price level $P_e$, is represented by the curve $AS$ in Figure 7-1. The AS curve has three properties that will prove useful in what follows:

- The aggregate supply curve is upward sloping. Put another way, an increase in output $Y$ leads to an increase in the price level $P$. You saw why earlier.
- The aggregate supply curve goes through point $A$, where $Y = Y_n$ and $P = P_e$. Put another way: When output $Y$ is equal to the natural level of output $Y_n$, the price level $P$ turns out to be exactly equal to the expected price level $P_e$.

How do we know this? From the definition of the natural level of output in Chapter 6. Recall that we defined the natural rate of unemployment (and by implication the natural level of output) as the rate of unemployment (and by implication the level of output) that prevails if the price level and the expected price level are equal.

**Figure 7-1**

*The Aggregate Supply Curve*

Given the expected price level, an increase in output leads to an increase in the price level.
This property—that the price level equals the expected price level when output is equal to the natural level of output—has two straightforward implications:

When output is above the natural level of output, the price level turns out to be higher than expected. In Figure 7-1: If $Y$ is to the right of $Y_n$, $P$ is higher than $P^e$.

Conversely: When output is below the natural level of output, the price level turns to be lower than expected. In Figure 7-1: If $Y$ is to the left of $Y_n$, $P$ is lower than $P^e$.

An increase in the expected price level $P^e$ shifts the aggregate supply curve up. Conversely: A decrease in the expected price level shifts the aggregate supply curve down.

This third property is shown in Figure 7-2. Suppose the expected price level increases from $P^e$ to $P^e'$. At a given level of output, and, correspondingly, at a given unemployment rate, the increase in the expected price level leads to an increase in wages, which leads in turn to an increase in prices. So, at any level of output, the price level is higher: The aggregate supply curve shifts up. In particular, instead of going through point $A$ (where $Y = Y_n$ and $P = P^e$), the aggregate supply curve now goes through point $A'$ (where $Y = Y_m$, $P = P^e'$).

Let’s summarize:

- Starting from wage determination and price determination in the labor market, we have derived the aggregate supply relation.
- This relation implies that for a given expected price level, the price level is an increasing function of the level of output. It is represented by an upward-sloping curve, called the aggregate supply curve.
- Increases in the expected price level shift the aggregate supply curve up; decreases in the expected price level shift the aggregate supply curve down.

## 7-2 Aggregate Demand

The aggregate demand relation captures the effect of the price level on output. It is derived from the equilibrium conditions in the goods and financial markets we described in Chapter 5.

In Chapter 5, we derived the following equation for goods-market equilibrium (equation (5.2)):

$$Y = C(Y - T) + I(Y, i) + G$$
Equilibrium in the goods market requires that output equal the demand for goods—the sum of consumption, investment, and government spending. This is the *IS* relation.

Also in Chapter 5, we derived the following equation for equilibrium in financial markets (equation (5.3)):

\[
\frac{M}{P} = Y L(i)
\]

Equilibrium in financial markets requires that the supply of money equal the demand for money. This is the *LM* relation.

Recall that what appears on the left side of the *LM* equation is the real money stock, \(M/P\). We focused in Chapters 5 and 6 on changes in the real money stock that came from changes in nominal money \(M\) by the Fed. But changes in the real money stock \(M/P\) can also come from changes in the price level \(P\). A 10% increase in the price level \(P\) has the same effect on the real money stock as a 10% decrease in the stock of nominal money \(M\): Either leads to a 10% decrease in the real money stock.

Using the *IS* and *LM* relations, we can derive the relation between the price level and the level of output implied by equilibrium in the goods and financial markets. We do this in Figure 7-3.

- Figure 7-3(a) draws the *IS* curve and the *LM* curve. The *IS* curve is drawn for given values of \(G\) and \(T\). It is downward sloping: An increase in the interest rate leads to a decrease in output. The *LM* curve is drawn for a given value of \(M/P\). It is upward sloping: An increase in output increases the demand for money, and the interest rate

---

**Figure 7-3**

*The Derivation of the Aggregate Demand Curve*

An increase in the price level leads to a decrease in output.
increases so as to maintain equality of money demand and the (unchanged) money supply. The point at which the goods market and the financial markets are both in equilibrium is at the intersection of the IS curve and the LM curve, at point A.

Now consider the effects of an increase in the price level from $P$ to $P'$. Given the stock of nominal money, $M$, the increase in the price level $P$ decreases the real money stock, $M/P$. This implies that the LM curve shifts up: At a given level of output, the lower real money stock leads to an increase in the interest rate. The economy moves along the IS curve, and the equilibrium moves from $A$ to $A'$. The interest rate increases from $i$ to $i'$, and output decreases from $Y$ to $Y'$. In short: The increase in the price level leads to a decrease in output.

In words: The increase in the price level leads to a decrease in the real money stock. This monetary contraction leads to an increase in the interest rate, which leads in turn to a lower demand for goods and lower output.

The negative relation between output and the price level is drawn as the downward-sloping AD curve in Figure 7-3(b). Points A and $A'$ in Figure 7-3(b) correspond to points $A$ and $A'$ in Figure 7-3(a). An increase in the price level from $P$ to $P'$ leads to a decrease in output from $Y$ to $Y'$. This curve is called the aggregate demand curve. The underlying negative relation between output and the price level is called the aggregate demand relation.

Any variable other than the price level that shifts either the IS curve or the LM curve also shifts the aggregate demand relation.

Take, for example, an increase in government spending $G$. At a given price level, the level of output implied by equilibrium in the goods and the financial markets is higher: In Figure 7-4, the aggregate demand curve shifts to the right, from $AD$ to $AD'$. Or take a contractionary, open market operation—a decrease in $M$. At a given price level, the level of output implied by equilibrium in the goods and the financial markets is lower. In Figure 7-4, the aggregate demand curve shifts to the left, from $AD$ to $AD''$.

Let’s represent what we have just derived by the following aggregate demand relation:

$$Y = Y \left( \frac{M}{P}, G, T \right)$$  \hspace{1cm} (7.3)

### Figure 7-4

**Shifts of the Aggregate Demand Curve**

At a given price level, an increase in government spending increases output, shifting the aggregate demand curve to the right. At a given price level, a decrease in nominal money decreases output, shifting the aggregate demand curve to the left.
Output $Y$ is an increasing function of the real money stock $M/P$, an increasing function of government spending $G$, and a decreasing function of taxes, $T$.

Given monetary and fiscal policy—that is, given $M$, $G$, and $T$—an increase in the price level $P$ leads to a decrease in the real money stock, $M/P$, which leads to a decrease in output. This is the relation captured by the $AD$ curve in Figure 7-3(b).

Let's summarize:

- Starting from the equilibrium conditions for the goods and financial markets, we have derived the aggregate demand relation.
- This relation implies that the level of output is a decreasing function of the price level. It is represented by a downward-sloping curve, called the aggregate demand curve.
- Changes in monetary or fiscal policy—or, more generally, in any variable, other than the price level, that shifts the $IS$ or the $LM$ curves—shift the aggregate demand curve.

### 7-3 Equilibrium in the Short Run and in the Medium Run

The next step is to put the $AS$ and the $AD$ relations together. From Sections 7-1 and 7-2, the two relations are given by

\[
\begin{align*}
\text{AS relation} & \quad P = P^e (1 + m) F \left( 1 - \frac{Y}{L}, z \right) \\
\text{AD relation} & \quad Y = Y \left( \frac{M}{P}, G, T \right)
\end{align*}
\]

For a given value of the expected price level, $P^e$ (which enters the aggregate supply relation), and for given values of the monetary and fiscal policy variables $M$, $G$, and $T$ (which enter the aggregate demand relation), these two relations determine the equilibrium values of output, $Y$, and the price level, $P$.

Note the equilibrium depends on the value of $P^e$. The value of $P^e$ determines the position of the aggregate supply curve (go back to Figure 7-2), and the position of the aggregate supply curve affects the equilibrium. In the short run, we can take $P^e$, the price level expected by wage setters when they last set wages, as given. But, over time, $P^e$ is likely to change, shifting the aggregate supply curve and changing the equilibrium. With this in mind, we first characterize equilibrium in the short run—that is, taking $P^e$ as given. We then look at how $P^e$ changes over time, and how that change affects the equilibrium.

### Equilibrium in the Short Run

The short-run equilibrium is characterized in Figure 7-5:

- The aggregate supply curve $AS$ is drawn for a given value of $P^e$. It is upward sloping: The higher the level of output, the higher the price level. The position of the curve depends on $P^e$. Recall from Section 7-1 that, when output is equal to the natural level of output, the price level is equal to the expected price level. This means that, in Figure 7-5, the aggregate supply curve goes through point $B$: If $Y = Y_n$, then $P = P^e$.
- The aggregate demand curve $AD$ is drawn for given values of $M$, $G$, and $T$. It is downward sloping: The higher the price level, the lower the level of output.

The equilibrium is given by the intersection of the $AS$ and $AD$ curves at point $A$. By construction, at point $A$, the goods market, the financial markets, and the labor market are all in equilibrium. The fact that the labor market is in equilibrium is because point
The equilibrium is given by the intersection of the aggregate supply curve and the aggregate demand curve. At point $A$, the labor market, the goods market, and financial markets are all in equilibrium.

$A$ is on the aggregate supply curve. That fact that the goods and financial markets are in equilibrium is because point $A$ is on the aggregate demand curve. The equilibrium level of output and price level are given by $Y$ and $P$.

There is no reason why, in general, equilibrium output $Y$ should be equal to the natural level of output $Y_n$. Equilibrium output depends both on the position of the aggregate supply curve (and therefore on the value of $P^e$) and on the position of the aggregate demand curve (and therefore on the values of $M$, $G$, and $T$). As we have drawn the two curves, $Y$ is greater than $Y_n$: In other words, the equilibrium level of output exceeds the natural level of output. But clearly we could have drawn the $AS$ and the $AD$ curves so equilibrium output $Y$ was smaller than the natural level of output $Y_n$.

Figure 7-5 gives us our first important conclusion: In the short run, there is no reason why output should equal the natural level of output. Whether it does depends on the specific values of the expected price level and the values of the variables affecting the position of aggregate demand.

We must now ask, What happens over time? More precisely: Suppose, in the short run, output is above the natural level of output—as is the case in Figure 7-5. What will happen over time? Will output eventually return to the natural level of output? If so, how? These are the questions we take up in the rest of this section.

**From the Short Run to the Medium Run**

To think about what happens over time, consider Figure 7-6. The curves denoted $AS$ and $AD$ are the same as in Figure 7-5, and so the short-run equilibrium is at point $A$—which corresponds to point $A$ in Figure 7-5. Output is equal to $Y$, and is higher than the natural level of output $Y_n$.

At point $A$, output exceeds the natural level of output. So we know from Section 7-1 that the price level is higher than the expected price level—higher than the price level wage setters expected when they set nominal wages.

The fact that the price level is higher than wage setters expected is likely to lead them to revise upward their expectations of what the price level will be in the future. So, next time they set nominal wages, they are likely to make that decision based on a higher expected price level, say based on $P^e$, where $Pe > Pe$.

This increase in the expected price level implies that in the next period, the aggregate supply curve shifts up, from $AS$ to $AS’$: At a given level of output, wage setters expect a higher price level. They set a higher nominal wage, which in turn leads firms to set a higher price. The price level therefore increases.
This upward shift in the $AS$ curve implies that the economy moves up along the $AD$ curve. The equilibrium moves from $A$ to $A'$. Equilibrium output decreases from $Y$ to $Y'$. The adjustment does not end at point $A'$. At $A'$, output $Y'$ still exceeds the natural level of output $Y_n$, so the price level is still higher than the expected price level. Because of this, wage setters are likely to continue to revise upwards their expectation of the price level.

This means that as long as equilibrium output exceeds the natural level of output $Y_n$, the expected price level increases, shifting the $AS$ curve upward. As the $AS$ curve shifts upward and the economy moves up along the $AD$ curve, equilibrium output continues to decrease.

Does this adjustment eventually come to an end? Yes. It ends when the $AS$ curve has shifted all the way to $AS''$, when the equilibrium has moved all the way to $A''$, and the equilibrium level of output is equal to $Y_n$. At $A''$, equilibrium output is equal to the natural level of output, so the price level is equal to the expected price level. At this point, wage setters have no reason to change their expectations; the $AS$ curve no longer shifts, and the economy stays at $A''$.

In words: So long as output exceeds the natural level of output, the price level turns out to be higher than expected. This leads wage setters to revise their expectations of the price level upward, leading to an increase in the price level. The increase in the price level leads to a decrease in the real money stock, which leads to an increase in the interest rate, which leads to a decrease in output. The adjustment stops when output is equal to the natural level of output. At that point, the price level is equal to the expected price level, expectations no longer change, and, output remains at the natural level of output. Put another way, in the medium run, output returns to the natural level of output.

We have looked at the dynamics of adjustment starting from a case in which initial output was higher than the natural level of output. Clearly, a symmetric argument holds when initial output is below the natural level of output. In this case, the price level is lower than the expected price level, leading wage setters to lower their expectations of the price level. Lower expectations of the price level cause the $AS$ curve to shift down and the economy to move down the $AD$ curve until output has increased back to the natural level of output.

Let’s summarize:

- In the short run, output can be above or below the natural level of output. Changes in any of the variables that enter either the aggregate supply relation or the aggregate demand relation lead to changes in output and to changes in the price level.  
  \[ \text{In the short run, } Y \neq Y_n. \]

- In the medium run, output eventually returns to the natural level of output. The adjustment works through changes in the price level. When output is above the natural level of output, the price level increases. The higher price level decreases demand and output. When output is below the natural level of output, the price level decreases, increasing demand and output.  
  \[ \text{In the medium run, } Y = Y_n. \]
The best way to understand more fully the AS–AD model is to use it to look at the dynamic effects of changes in policy or in the economic environment. In the next three sections, we focus on three such changes. The first two are our two favorite policy changes: a change in the stock of nominal money; and a change in the budget deficit. The third, which we could not examine before we had developed a theory of wage and price determination, is an increase in the oil price.

7-4 The Effects of a Monetary Expansion

What are the short-run and medium-run effects of an expansionary monetary policy, say of a one-time increase in the level of nominal money from \( M \) to \( M' \)?

The Dynamics of Adjustment

Look at Figure 7-7. Assume that before the change in nominal money, output is at its natural level. Aggregate demand and aggregate supply cross at point \( A \), the level of output at \( A \) equals \( Y_n \), and the price level equals \( P \).

Now consider an increase in nominal money. Recall the specification of aggregate demand from equation (7.3):

\[
Y = Y\left(\frac{M}{P}, G, T\right)
\]

For a given price level \( P \), the increase in nominal money \( M \) leads to an increase in the real money stock \( M/P \), leading to an increase in output. The aggregate demand curve shifts to the right, from \( AD \) to \( AD' \). In the short run, the economy goes from point \( A \) to \( A' \). Output increases from \( Y_n \) to \( Y' \), and the price level increases from \( P \) to \( P' \).

Over time, the adjustment of price level expectations comes into play. As output is higher than the natural level of output, the price level is higher than wage setters expected. They then revise their expectations, which causes the aggregate supply curve to shift up over time. The economy moves up along the aggregate demand curve \( AD' \). The adjustment process stops when output has returned to the natural level of output. At that point, the price level is equal to the expected price level. In the medium run, the aggregate supply curve is given by \( AS'' \), and the economy is at point \( A'' \): Output is back to \( Y'' \), and the price level is equal to \( P'' \).

Figure 7-7

The Dynamic Effects of a Monetary Expansion

A monetary expansion leads to an increase in output in the short run but has no effect on output in the medium run.
We can actually pin down the exact size of the eventual increase in the price level. If output is back to the natural level of output, the real money stock must also be back to its initial value. In other words, the proportional increase in prices must be equal to the proportional increase in the nominal money stock: If the initial increase in nominal money is equal to 10%, then the price level ends up 10% higher.

**Going Behind the Scenes**

To get a better sense of what is going on, it is useful to go behind the scenes to see what happens not only to output and to the price level, but also what happens to the interest rate. We can do this by looking at what happens in terms of the IS–LM model.

Figure 7-8(a) reproduces Figure 7-7 (leaving out the AS” curve to keep things simple) and shows the adjustment of output and the price level in response to the increase in nominal money. Figure 7-8(b) shows the adjustment of output and the interest rate by looking at the same adjustment process, but in terms of the IS–LM model.

Look first at Figure 7-8(b). Before the change in nominal money, the equilibrium is given by the intersection of the IS and LM curves; that is, at point A—which corresponds to point A in Figure 7-8(a). Output is equal to the natural level of output, $Y_n$, and the interest rate is given by $i$.

**Figure 7-8**

*The Dynamic Effects of a Monetary Expansion on Output and the Interest Rate*

The increase in nominal money initially shifts the LM curve down, decreasing the interest rate and increasing output. Over time, the price level increases, shifting the LM curve back up until output is back at the natural level of output.
The short-run effect of the monetary expansion is to shift the $LM$ curve down from $LM$ to $LM'$, moving the equilibrium from point $A$ to point $A'$—which corresponds to point $A'$ in Figure 7-8(a). The interest rate is lower, and output is higher.

Note that there are two effects at work behind the shift from $LM$ to $LM'$: One is due to the increase in nominal money. The other, which partly offsets the first, is due to the increase in the price level. Let’s look at these two effects more closely:

- If the price level did not change, the increase in nominal money would shift the $LM$ curve down to $LM''$. So, if the price level did not change—as was our assumption in Chapter 5—the equilibrium would be at the intersection of $IS$ and $LM''$, or point $B$.
- But even in the short run, the price level increases—from $P$ to $P'$ in Figure 7-8(a). This increase in the price level shifts the $LM$ curve upward from $LM''$ to $LM'$, partially offsetting the effect of the increase in nominal money.
- The net effect of these two shifts—down from $LM$ to $LM''$ in response to the increase in nominal money, and up from $LM''$ to $LM'$ in response to the increase in the price level—is a shift of the $LM$ curve from $LM$ to $LM'$, and the equilibrium is given by $A'$.

Over time, the fact that output is above its natural level implies that the price level continues to increase. As the price level increases, it further reduces the real money stock and shifts the $LM$ curve back up. The economy moves along the $IS$ curve: The interest rate increases and output declines. Eventually, the $LM$ curve returns to where it was before the increase in nominal money.

The economy ends up at point $A$, which corresponds to point $A''$ in Figure 7-8(a): The increase in nominal money is exactly offset by a proportional increase in the price level. The real money stock is therefore unchanged. With the real money stock unchanged, output is back to its initial value, $Y_n$, which is the natural level of output, and the interest rate is also back to its initial value, $i$.

**The Neutrality of Money**

Let’s summarize what we have just learned about the effects of monetary policy:

- In the short run, a monetary expansion leads to an increase in output, a decrease in the interest rate, and an increase in the price level.

  How much of the effect of a monetary expansion falls initially on output and how much on the price level depends on the slope of the aggregate supply curve. In Chapter 5, we assumed the price level did not respond at all to an increase in output—we assumed in effect that the aggregate supply curve was flat. Although we intended this as a simplification, empirical evidence does show that the initial effect of changes in output on the price level is indeed quite small. We saw this when we looked at estimated responses to changes in the Federal Funds rate in Figure 5-9: Despite the change in output, the price level remained practically unchanged for nearly a year.

- Over time, the price level increases, and the effects of the monetary expansion on output and on the interest rate disappear. In the medium run, the increase in nominal money is reflected entirely in a proportional increase in the price level. The increase in nominal money has no effect on output or on the interest rate. (How long it takes in reality for the effects of money on output to disappear is the topic of the Focus box “How Long Lasting Are the Real Effects of Money?”) Economists refer to the absence of a medium-run effect of money on output and on the interest rate by saying that money is neutral in the medium run.

The neutrality of money in the medium run does not mean that monetary policy cannot or should not be used to affect output. An expansionary monetary policy can, for example, help the economy move out of a recession and return more quickly to the natural level of output. As we saw in Chapter 5, this is exactly the way monetary policy was used to fight the 2001 recession. But it is a warning that monetary policy cannot sustain higher output forever.
How Long Lasting Are the Real Effects of Money?

To determine how long lasting the real effects of money are, economists use macroeconometric models. These models are larger-scale versions of the aggregate supply and aggregate demand model in this chapter.

The model we examine in this box was built in the early 1990s by John Taylor, at Stanford University. The Taylor model is substantially larger than the model we studied in this chapter. On the aggregate supply side, it has separate equations for price and for wage setting. On the demand side, it has separate equations for consumption, for investment, for exports, and for imports. (Recall that, so far, we have assumed the economy is closed, so we have ignored exports and imports altogether.) In addition, instead of looking at just one country as we have done here, it looks at eight countries (the United States, and seven major OECD countries) and solves for equilibrium in all eight countries simultaneously. Each equation, for each country, is estimated using econometrics, and allows for a richer dynamic structure than the equations we have relied on in this chapter.

The implications of the model for the effects of money on output are shown in Figure 1. The simulation looks at the effects of an increase in nominal money of 3% over the initial year, taking place over four quarters—0.1% in the first quarter, 0.6% in the second, 1.2% in the third, and 1.1% in the fourth. After these four step increases, nominal money remains at its new higher level forever.

The effects of money on output reach a maximum after three quarters. By then, output is 1.8% higher than it would have been without the increase in nominal money. Over time, however, the price level increases and output returns to the natural level of output. In year 4, the price level is up by 2.5%, while output is up by only 0.3%. Therefore, the Taylor model suggests, it takes roughly four years for output to return to its natural level, or, put another way, four years for changes in nominal money to become neutral.

Do all macroeconometric models give the same answer? No. Because they differ in the way they are constructed, in the way variables are chosen, and in the way equations are estimated, their answers are different. But most of them have the following implications in common: The effects of an increase in money on output build up for one to two years and then decline over time. (To get a sense of how the answers might differ across models, see the Focus box “Twelve Macroeconometric Models” in Chapter 22.)

Source: Figure 1 is reproduced from John Taylor, Macroeconomic Policy in a World Economy (W.W. Norton, 1993) Figure 5-1A, p. 138.

![Figure 1](image-url)
A Decrease in the Budget Deficit

The policy we just looked at—a monetary expansion—led to a shift in aggregate demand coming from a shift in the LM curve. Let’s now look at the effects of a shift in aggregate demand coming from a shift in the IS curve.

Suppose the government is running a budget deficit and decides to reduce it by decreasing its spending from $G$ to $G'$ while leaving taxes $T$ unchanged. How will this affect the economy in the short run and in the medium run?

Assume that output is initially at the natural level of output, so that the economy is at point $A$ in Figure 7-9: Output equals $Y_n$. The decrease in government spending from $G$ to $G'$ shifts the aggregate demand curve to the left, from $AD$ to $AD'$: For a given price level, output is lower. In the short run, the equilibrium moves from $A$ to $A'$; output decreases from $Y_n$ to $Y'$ and the price level decreases from $P$ to $P'$.

The initial effect of the deficit reduction triggers lower output. We first derived this result in Chapter 3, then in Chapter 5, and it holds here as well.

What happens over time? As long as output is below the natural level of output, we know that the aggregate supply curve keeps shifting down. The economy moves down along the aggregate demand curve $AD'$ until the aggregate supply curve is given by $AS''$ and the economy reaches point $A''$. By then, the recession is over, and output is back at $Y_n$.

Like an increase in nominal money, a reduction in the budget deficit does not affect output forever. Eventually, output returns to its natural level. But there is an important difference between the effects of a change in money and the effects of a change in the deficit. At point $A''$, not everything is the same as before: Output is back to the natural level of output, but the price level and the interest rate are lower than before the shift. (The fact that the price level decreases may feel strange given that, as we saw in Chapters 1 and 2, inflation is nearly always positive. This result comes from the fact that we are looking at an economy in which money growth is zero—we are assuming that $M$ is constant, not growing—and there is no sustained inflation. If we were to allow for money growth and thus for inflation, then the result would be that the price level decreases relative to what it would have been, or, in other words, that inflation goes down for a while. More on money growth and inflation in the next chapter.) The best way to see these specific effects is again to look at the adjustment in terms of the underlying IS–LM model.

Figure 7-9
The Dynamic Effects of a Decrease in the Budget Deficit

A decrease in the budget deficit leads initially to a decrease in output. Over time, however, output returns to the natural level of output.
Deficit Reduction, Output, and the Interest Rate

Figure 7-10(a) reproduces Figure 7-9, showing the adjustment of output and the price level in response to the increase in the budget deficit (but leaving out $AS''$ to keep things visually simple). Figure 7-10(b) shows the adjustment of output and the interest rate by looking at the same adjustment process, but in terms of the IS–LM model.

Look first at Figure 7-10(b). Before the change in fiscal policy, the equilibrium is given by the intersection of the $IS$ curve and the $LM$ curve, at point $A$—which corresponds to point $A$ in Figure 7-10(a). Output is equal to the natural level of output, $Y_n$, and the interest rate is given by $i$.

As the government reduces the budget deficit, the $IS$ curve shifts to the left, to $IS'$. If the price level did not change (the assumption we made in Chapter 5), the economy would move from point $A$ to point $B$. But, because the price level declines in response to the decrease in output, the real money stock increases, leading to a partially offsetting shift of the $LM$ curve, down to $LM'$. So, the initial effect of deficit reduction is to move the economy from point $A$ to point $A'$. (Point $A'$ in Figure 7-10(b) corresponds to point $A'$ in Figure 7-10(a).) Both output and the interest rate are lower than before the

Figure 7-10

The Dynamic Effects of a Decrease in the Budget Deficit on Output and the Interest Rate

A deficit reduction leads in the short run to a decrease in output and to a decrease in the interest rate. In the medium run, output returns to its natural level, while the interest rate declines further.
fiscal contraction. Note that, just as was the case in Chapter 5, we cannot tell whether investment increases or decreases in the short run: Lower output decreases investment, but the lower interest rate increases investment.

So long as output remains below the natural level of output, the price level continues to decline, leading to a further increase in the real money stock. The LM curve continues to shift down. In Figure 7-10(b), the economy moves down from point $A'$ along IS', and eventually reaches $A''$ (which corresponds to $A''$ in Figure 7-10(a)). At $A''$, the LM curve is given by $LM''$.

At $A''$, output is back at the natural level of output. But the interest rate is lower than it was before deficit reduction, down from $i$ to $i''$. The composition of output is also different: To see how and why, let’s rewrite the IS relation, taking into account that at $A''$, output is back at the natural level of output, so that $Y = Y_n$

$$Y_n = C(Y_n - T) + I(Y_n, i) + G$$

Because income $Y_n$ and taxes $T$ are unchanged, consumption $C$ is the same as before the deficit reduction. By assumption, government spending $G$ is lower than before. Therefore investment $I$ must be higher than before the deficit reduction—higher by an amount exactly equal to the decrease in $G$. Put another way, in the medium run, a reduction in the budget deficit unambiguously leads to a decrease in the interest rate and an increase in investment.

**Budget Deficits, Output, and Investment**

Let’s summarize what you have just learned about the effects of fiscal policy:

- **In the short run**, a budget deficit reduction, if implemented alone—that is, without an accompanying change in monetary policy—leads to a decrease in output and may lead to a decrease in investment. Note the qualification “without an accompanying change in monetary policy”: In principle, these adverse short-run effects on output can be avoided by using the right monetary–fiscal policy mix. What is needed is for the central bank to increase the money supply enough to offset the adverse effects of the decrease in government spending on aggregate demand. This is what happened in the United States in the 1990s. As the Clinton administration reduced budget deficits, the Fed made sure that, even in the short run, the deficit reduction did not lead to a recession and lower output.

- **In the medium run**, output returns to the natural level of output, and the interest rate is lower. In the medium run, a deficit reduction leads unambiguously to an increase in investment.

  We have not taken into account so far the effects of investment on capital accumulation and the effects of capital on production (we will do so in Chapters 10 to 13 when we look at the long run). But it is easy to see how our conclusions would be modified if we did take into account the effects on capital accumulation. In the long run, the level of output depends on the capital stock in the economy. So if a lower government budget deficit leads to more investment, it will lead to a higher capital stock, and the higher capital stock will lead to higher output.

  Everything we have just said about the effects of deficit reduction through a fiscal consolidation would apply equally to policy measures aimed at increasing private saving. An increase in the private saving rate—that is, lower consumption at the same level of disposable income—decreases demand and output in the short run, leaves output unchanged in the medium run, and, through increases in the capital stock from increased investment, increases output in the long run.
Disagreements among economists about the effects of measures aimed at increasing either public saving or private saving often come from differences in time frames. Those who are concerned with short-run effects worry that measures to increase saving, public or private, might create a recession and decrease saving and investment for some time. Those who look beyond the short run see the eventual increase in saving and investment and emphasize the favorable medium-run and long-run effects on output.

7-6 An Increase in the Price of Oil

So far we have looked at shocks that shift the aggregate demand curve: an increase in the money supply, or a reduction in the budget deficit. There are other shocks, however, that affect both aggregate demand and aggregate supply and play an important role in fluctuations. An obvious candidate is movements in the price of oil. To see why, turn to Figure 7-11.

Figure 7-11 plots two series. The first, represented by the red line, is the dollar price of oil—that is, the price of a barrel of oil in dollars—since 1970. It is measured on the vertical axis on the left. This is the series that is quoted in the newspapers, more or less every day. What matters, however, for economic decisions is not the dollar price, but the real price of oil; that is, the dollar price of oil divided by the price level. Thus, the second series in the figure, represented by the blue line, shows the real price of oil, constructed as the dollar price of oil divided by the U.S. consumer price index. Note that the real price is an index; it is normalized to equal 100 in 1970. It is measured on the vertical axis on the right.

What is perhaps most striking in the figure is the large increase in the real price of oil in the 2000s: In 10 years, from 1998 to 2008, the index for the real price went from about 100 to more than 500, a more than five-fold increase. As the figure shows, however, there were two similar increases in the price of oil in the 1970s, the first in 1973–1975 and the second in 1977–1981. Just as in the more recent episode, the real price of oil increased from 100 in 1970 (this is the normalization we have chosen) to more than 500 in 1981.

Figure 7-11
The Nominal and the Real Price of Oil, 1970–2010

Over the last 40 years, there have been three sharp increases in the real price of oil. The first two increases took place in the 1970s. The more recent one took place in the 2000s, until the crisis hit.

Source: Series OILPRICE, CPIAUSCL Federal Reserve Economic Data (FRED) http://research.stlouisfed.org/fred2/. The value of the index is set equal to 100 in 1970.)
What was behind these large increases? In the 1970s, the main factors were the formation of OPEC (the Organization of Petroleum Exporting Countries), a cartel of oil producers that was able to act as a monopoly and increase prices, and disruptions due to wars and revolutions in the Middle East. In the 2000s, the main factor was quite different, namely the fast growth of emerging economies, in particular China, which led to a rapid increase in the world demand for oil and, by implication, a steady increase in real oil prices. Whether coming from changes in supply in the 1970s or from changes in the demand from emerging countries in the 2000s, the implication for U.S. firms and consumers was the same: more expensive oil, more expensive energy.

In thinking about the macroeconomic effects of such increases, it is clear that we face a serious problem in using the model we have developed so far: The price of oil appears neither in our aggregate supply relation nor in our aggregate demand relation! The reason is that, until now, we have assumed that output was produced using only labor. One way to extend our model would be to recognize explicitly that output is produced using labor and other inputs (including energy), and then figure out what effect an increase in the price of oil has on the price set by firms and on the relation between output and employment. An easier way, and the way we shall go here, is simply to capture the increase in the price of oil by an increase in the markup of the price over the nominal wage. The justification is straightforward: Given wages, an increase in the price of oil increases the cost of production, forcing firms to increase prices.

Having made this assumption, we can then track the dynamic effects of an increase in the markup on output and the price level. It will be easiest here to work backward in time, first asking what happens in the medium run, and then working out the dynamics of adjustment from the short run to the medium run.

### Effects on the Natural Rate of Unemployment

Let’s start by asking what happens to the natural rate of unemployment when the real price of oil increases (for simplicity, we shall drop “real” in what follows). Figure 7-12 reproduces the characterization of labor-market equilibrium from Figure 6-8 in Chapter 6:

The wage-setting curve is downward sloping. The price-setting relation is represented by the horizontal line at \( \frac{W}{P} = \frac{1}{1 + m} \). The initial equilibrium is at point \( A \), and the initial natural unemployment rate is \( u_n \). An increase in the markup leads to a downward shift of the price-setting line, from \( PS \) to \( PS' \): The higher the markup, the lower the real wage implied by price setting. The equilibrium moves from \( A \) to \( A' \). The

**Figure 7-12**

*The Effects of an Increase in the Price of Oil on the Natural Rate of Unemployment*

An increase in the price of oil leads to a lower real wage and a higher natural rate of unemployment.
real wage is lower. The natural unemployment rate is higher: Getting workers to accept the lower real wage requires an increase in unemployment.

The increase in the natural rate of unemployment leads in turn to a decrease in the natural level of employment. If we assume that the relation between employment and output is unchanged—that is, that each unit of output still requires one worker in addition to the energy input—then the decrease in the natural level of employment leads to an identical decrease in the natural level of output. Putting things together: An increase in the price of oil leads to a decrease in the natural level of output.

The Dynamics of Adjustment

Let’s now turn to dynamics. Suppose that before the increase in the price of oil, the aggregate demand curve and the aggregate supply curve are given by $AD$ and $AS$, respectively, so the economy is at point $A$ in Figure 7-13, with output at the natural level of output, $Y_n$, and by implication $P = P^e$.

We have just established that the increase in the price of oil decreases the natural level of output. Call this lower level $Y_n'$. We now want to know what happens in the short run and how the economy moves from $Y_n$ to $Y_n'$.

To think about the short run, recall that the aggregate supply relation is given by

$$P = P^e (1 + m) F \left( 1 - \frac{Y}{L} z \right)$$

Recall that we capture the effect of an increase in the price of oil by an increase in the markup $m$. So, in the short run (given $P^e$), the increase in the price of oil shows up as an increase in the markup $m$. This increase in the markup leads firms to increase their prices, leading to an increase in the price level $P$ at any level of output $Y$. The aggregate supply curve shifts up.

We can be more specific about the size of the shift, and knowing the size of this shift will be useful in what follows. We know from Section 7-1 that the aggregate supply curve always goes through the point such that output equals the natural level of output and the price level equals the expected price level. Before the increase in the price of oil, the aggregate supply curve in Figure 7-13 goes through point $A$, where output equals $Y_n$ and the price level is equal to $P^e$. After the increase in the price of oil, the new aggregate supply curve goes through point $B$, where output equals the new lower

Figure 7-13

The Dynamic Effects of an Increase in the Price of Oil

An increase in the price of oil leads, in the short run, to a decrease in output and an increase in the price level. Over time, output decreases further and the price level increases further.
natural level of output $Y_n$ and the price level equals the expected price level, $P^e$. The aggregate supply curve shifts left from $AS$ to $AS'$.

Does the aggregate demand curve shift as a result of the increase in the price of oil? The answer is: maybe. There are many channels through which demand might be affected at a given price level: The higher price of oil may lead firms to change their investment plans, canceling some investment projects and/or shifting to less energy-intensive equipment. The increase in the price of oil also redistributes income from oil buyers to oil producers. Oil producers may spend less than oil buyers, leading to a decrease in consumption demand. Let’s take the easy way out: Because some of the effects shift the aggregate demand curve to the right and others shift the aggregate demand curve to the left, let’s simply assume that the effects cancel each other out and that aggregate demand does not shift.

Under this assumption, in the short run, only the $AS$ shifts. The economy therefore moves along the $AD$ curve, from $A$ to $A'$. Output decreases from $Y_n$ to $Y'$. The increase in the price of oil leads firms to increase their prices. This increase in the price level then decreases demand and output.

What happens over time? Although output has fallen, the natural level of output has fallen even more: At point $A'$, output $Y'$ is still above the new natural level of output $Y_n'$, so the aggregate supply curve continues to shift up. The economy therefore moves over time along the aggregate demand curve, from $A'$ to $A''$. At point $A''$, output $Y$ is equal to the new lower natural level of output $Y_n''$ and the price level is higher than before the oil shock: Shifts in aggregate supply affect output not only in the short run but in the medium run as well.

To summarize: Increases in the price of oil decrease output and increase prices in the short run. If the increase in the price of oil is permanent, then output is lower not only in the short run, but also in the medium run.

How do these implications fit what we observed in response to increases in the price of oil both in the 1970s and in the 2000s? The answers are given by Figure 7-14, which plots the evolution of the real price of oil and inflation—using the CPI—and
Oil Price Increases: Why Were the 2000s so Different from the 1970s?

The question raised by Figures 7-14 and 7-15 is an obvious one: Why is it that oil price increases were associated with stagflation in the 1970s, but have had so little apparent effect on the economy in the 2000s?

A first line of explanation is that shocks other than the increase in the price of oil were at work in the 1970s and in the 2000s.

In the 1970s, not only did the price of oil increase, but so did the price of many other raw materials. This implies that the aggregate supply relation shifted up by more than implied by just the increase in the price of oil, and so the adverse effect on output was stronger than in the 2000s.

In the 2000s, many economists believe that, partly because of globalization and foreign competition, workers bargaining power weakened. If true, this implies that, while the increase in oil prices shifted the aggregate supply curve up, the decrease in bargaining power of workers shifted it down, dampening or even eliminating the adverse effects of the oil price increase on output and the price level.

Econometric studies suggest, however, that more was at work, and that, even after controlling for the presence of these other factors, the effects of the price of oil have changed since the 1970s. Figure 1 shows the effects of a 100% increase in the price of oil on output and on the price level, estimated using data from two different periods. The black and blue lines show the effects of an increase in the price of oil on the CPI deflator and on GDP, based on data from 1970:1 to 1986:4; the green and red lines do the same, but based on data from 1987:1 to 2006:4 (the time scale on the horizontal axis is in quarters). The figure suggests two main conclusions. First, in both periods, as predicted by our model, the increase in the price of oil leads to an increase in the CPI and a decrease in GDP. Second, the effects of the increase in the price of oil on both the CPI and on GDP have become smaller, roughly half of what they were earlier.

Why have the adverse effects of the increase in the price of oil become smaller? This is still very much a topic of research. But, at this stage, two hypotheses appear plausible: The first hypothesis is that, today, U.S. workers have less bargaining power than they did in the 1970s. Thus, as the price of oil has increased, workers have been more willing to accept a reduction in wages, limiting the upward shift in the aggregate supply curve, and thus limiting the adverse effect on the price level and on output. (Make sure you understand this statement; use Figure 7-13 to increase your understanding.)

The second hypothesis concerns monetary policy. When the price of oil increased in the 1970s, people started expecting much higher prices in general, and $P_e$ increased dramatically. The result was further shifts of the aggregate supply curve, leading to a larger increase in the price level and a larger decrease in output. In the 2000s, monetary policy was conducted in a very different way than in the 1970s, and expectations were that the Fed would not let the increase in the price of oil lead to a much higher price level. Thus, $P_e$ barely increased over time, leading to a smaller shift of the aggregate supply curve, and thus a smaller effect on output and the price level than in the 1970s. (Again, use Figure 7-13 to make sure you understand this statement.)
The Medium Run

The Core

Figure 7-15, which plots the evolution of the real price of oil and the unemployment rate, in the United States since 1970.

First, the good news (for our model, although surely not for the U.S. economy at the time): Note how both of the increases in the price of oil of the 1970s were followed by major increases in inflation and in unemployment. This fits our conclusions very well. Now, the bad news (for our model, but not for the U.S. economy): Note how the increase in the price of oil in the 2000s was associated with neither an increase in inflation nor an increase in unemployment. In light of what happened in the 1970s, this lack of an effect has come as a surprise to macroeconomists. The state of research, and various hypotheses being explored, are discussed in the Focus box “Oil Price Increases: Why Were the 2000s so Different from the 1970s?”. A summary of the conclusions goes like this: Oil price increases still decrease output and increase inflation. Because of decreases in the use of oil in production, because of changes in the labor market, and because of improvements in the conduct of monetary policy, the effect of oil price increases on both output and inflation was smaller in the 2000s than it was in the 1970s. And the reason it is hard to see an adverse effect on output and on inflation in the 2000s in Figures 7-14 and 7-15 is that these oil price shocks were largely offset by other, favorable shocks.

7-7 Conclusions

This chapter has covered a lot of ground. Let us repeat some key ideas and develop some of the earlier conclusions.

The Short Run versus the Medium Run

One key message of this chapter is that changes in policy typically have different effects in the short run and in the medium run. The main results of this chapter are summarized in Table 7-1. A monetary expansion, for example, affects output in the short run but not in the medium run. In the short run, a reduction in the budget deficit decreases output and decreases the interest rate and may decrease investment. But in
the medium run, the interest rate decreases and output returns to the natural level of output, and investment unambiguously increases.

Disagreements among economists about the effects of various policies often come from differences in the time frame they have in mind. If you are worried about output and investment in the short run, you might be reluctant to proceed with fiscal consolidation. But if your focus is on the medium or the long run, you will see the consolidation as helping investment and eventually, through higher capital accumulation, increasing output. One implication is that where you stand depends in particular on how fast you think the economy adjusts to shocks. If you believe that it takes a long time for output to return to its natural level, you will naturally focus more on the short run and be willing to use policies that increase output in the short run, even if medium-run effects are nil or negative. If you believe instead that output returns to its natural level quickly, you will put more emphasis on the medium-run implications and will, by implication, be more reluctant to use those policies.

### Shocks and Propagation Mechanisms

This chapter also gives you a general way of thinking about output fluctuations (sometimes called business cycles)—movements in output around its trend (a trend that we have ignored so far but on which we will focus in Chapters 10 to 13):

You can think of the economy as being constantly hit by shocks. These shocks may be shifts in consumption coming from changes in consumer confidence, shifts in investment, shifts in the demand for money, and so on. Or they may come from changes in policy—from the introduction of a new tax law, to a new program of infrastructure investment, to a decision by the central bank to fight inflation by tightening the money supply.

Each shock has dynamic effects on output and its components. These dynamic effects are called the propagation mechanism of the shock. Propagation mechanisms are different for different shocks. The effects of a shock on activity may build up over time, affecting output in the medium run. Or the effects may build up for a while and then decrease and disappear. We saw, for example, that the effects of an increase in money on output reach a peak after six to nine months and then slowly decline afterward as the price level eventually increases in proportion to the increase in nominal money. At times, some shocks are sufficiently large or come in sufficiently bad combinations that they create a recession. The two recessions of the 1970s were due largely to increases in the price of oil; the recession of the early 1980s was due to a sharp contraction in money; the recession of the early 1990s was due primarily to a sudden decline in consumer confidence; the recession of 2001 was due to a sharp drop in investment spending. The current crisis and the sharp decrease in output in 2010 had its origins in

| Table 7-1 Short-Run Effects and Medium-Run Effects of a Monetary Expansion and a Budget Deficit Reduction on Output, the Interest Rate, and the Price Level |
|---|---|---|---|---|---|---|
| Monetary expansion | increase | decrease | increase (small) | no change | no change | increase |
| Deficit reduction | decrease | decrease | decrease (small) | no change | decrease | decrease |

How to define shocks is harder than it looks. Suppose a failed economic program in an Eastern European country leads to political chaos in that country, which leads to increased risk of nuclear war in the region, which leads to a fall in consumer confidence in the United States, which leads to a recession in the United States. What is the “shock”? The failed program? The fall of democracy? The increased risk of nuclear war? Or the decrease in consumer confidence? In practice, we have to cut the chain of causation somewhere. Thus, we may refer to the drop in consumer confidence as “the shock” and ignore its underlying causes.
the problems of the housing market, which then led to a major financial shock, and in turn to a sharp reduction in output. What we call economic fluctuations are the result of these shocks and their dynamic effects on output.

Where We Go from Here

■ We assumed in this chapter that the nominal money stock was constant, that there was no nominal money growth. This led to a constant price level in the medium run. What we observe most of the time, however, is positive inflation, namely a steady increase in the price level. This in turn requires us to extend our analysis to the case where money growth is positive and to revisit the relation among output, unemployment, inflation, and money growth. We take this up in Chapter 8.

■ The AS–AD model we constructed in this chapter has a reassuring property. While shocks move output away from its natural level in the short run, there are forces that tend to take it back to its natural level over time. Output below its natural level leads to a decrease in the price level, which leads in turn to an increase in the real money stock, a decrease in the interest rate, and an increase in demand and in output. This process takes place until output has returned to its natural level and there is no longer pressure on the price level to adjust further. And, if this process is too slow, fiscal and monetary policies can help accelerate the return to the natural rate. The crisis and the very slow recovery (recall the facts presented in Chapter 1) we are experiencing force us to reconsider these conclusions. This is what we do in Chapter 9.

Summary

■ The model of aggregate supply and aggregate demand describes the movements in output and the price level when account is taken of equilibrium in the goods market, the financial markets, and the labor market.

■ The aggregate supply relation captures the effects of output on the price level. It is derived from equilibrium in the labor market. It is a relation among the price level, the expected price level, and the level of output. An increase in output decreases unemployment; the decrease in unemployment increases wages and, in turn, increases the price level. An increase in the expected price level leads, one for one, to an increase in the actual price level.

■ The aggregate demand relation captures the effects of the price level on output. It is derived from equilibrium in goods and financial markets. An increase in the price level decreases the real money stock, increasing the interest rate and decreasing output.

■ In the short run, movements in output come from shifts in either aggregate demand or aggregate supply. In the medium run, output returns to the natural level of output, which is determined by equilibrium in the labor market.

■ An expansionary monetary policy leads in the short run to an increase in the real money stock, a decrease in the interest rate, and an increase in output. Over time, the price level increases, and the real money stock decreases until output has returned to its natural level. In the medium run, money does not affect output, and changes in money are reflected in proportional increases in the price level. Economists refer to this fact by saying that, in the medium run, money is neutral.

■ A reduction in the budget deficit leads in the short run to a decrease in the demand for goods and therefore to a decrease in output. Over time, the price level decreases, leading to an increase in the real money stock and a decrease in the interest rate. In the medium run, output increases back to the natural level of output, but the interest rate is lower and investment is higher.

■ An increase in the price of oil leads, in both the short run and in the medium run, to a decrease in output. In the short run, it leads to an increase in the price level, which decreases the real money stock and leads to a contraction of demand and output. In the medium run, an increase in the price of oil decreases the real wage paid by firms, increases the natural rate of unemployment, and therefore decreases the natural level of output.

■ The difference between short-run effects and medium-run effects of policies is one of the reasons economists disagree in their policy recommendations. Some economists
believe the economy adjusts quickly to its medium-run equilibrium, so they emphasize medium-run implications of policy. Others believe the adjustment mechanism through which output returns to the natural level of output is a slow process at best, and so they put more emphasis on the short-run effects of policy.

Economic fluctuations are the result of a continual stream of shocks to aggregate supply or to aggregate demand and of the dynamic effects of each of these shocks on output. Sometimes the shocks are sufficiently adverse, alone or in combination, that they lead to a recession.

Key Terms

aggregate supply relation, 134
aggregate demand relation, 136
neutrality of money, 144
macroeconometric models, 145
Organization of Petroleum Exporting Countries (OPEC), 150
output fluctuations, 155
business cycles, 155
shocks, 155
propagation mechanism, 155

Questions and Problems

QUICK CHECK
All Quick Check questions and problems are available on MyEconLab.
1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. The aggregate supply relation implies that an increase in output leads to an increase in the price level.
   b. The natural level of output can be determined by looking solely at the aggregate supply relation.
   c. The aggregate demand relation is downward sloping because at a higher price level, consumers wish to purchase fewer goods.
   d. In the absence of changes in fiscal or monetary policy, the economy will always remain at the natural level of output.
   e. Expansionary monetary policy has no effect on the level of output in the medium run.
   f. Fiscal policy cannot affect investment in the medium run because output always returns to its natural level.
   g. In the medium run, output and the price level always return to the same value.

2. Aggregate demand shocks and the medium run
   Suppose the economy begins with output equal to its natural level. Then, there is a reduction in income taxes.
   a. Using the AS–AD model developed in this chapter, show the effects of a reduction in income taxes on the position of the AD, AS, IS, and LM curves in the medium run.
   b. What happens to output, the interest rate, and the price level in the medium run? What happens to consumption and investment in the medium run?

3. Aggregate supply shocks and the medium run
   Consider an economy with output equal to the natural level of output. Now suppose there is an increase in unemployment benefits.
   a. Using the model developed in this chapter, show the effects of an increase in unemployment benefits on the position of the AD and AS curves in the short run and in the medium run.
   b. How will the increase in unemployment benefits affect output and the price level in the short run and in the medium run?

4. The neutrality of money
   a. In what sense is money neutral? How is monetary policy useful if money is neutral?
   b. Fiscal policy, like monetary policy, cannot change the natural level of output. Why then is monetary policy considered neutral but fiscal policy is not?
   c. Discuss the statement “Because neither fiscal nor monetary policy can affect the natural level of output, it follows that, in the medium run, the natural level of output is independent of all government policies.”

DIG DEEPER
All DIG Deeper questions and problems are available on MyEconLab.
5. The paradox of saving, one more time
   In chapter problems at the end of Chapters 3 and 5, we examined the paradox of saving in the short run, under different assumptions about the response of investment to output and the interest rate. Here we consider the issue one last time in the context of the AS–AD model.
   Suppose the economy begins with output equal to its natural level. Then there is a decrease in consumer confidence as households attempt to increase their saving for a given level of disposable income.
   a. In AS–AD and IS–LM diagrams, show the effects of the decline in consumer confidence in the short run and in the medium run. Explain why the curves shift in your diagrams.
   b. What happens to output, the interest rate, and the price level in the short run? What happens to consumption,
investment, and private saving in the short run? Is it possible that the decline in consumer confidence will actually lead to a fall in private saving in the short run?

c. Repeat part (b) for the medium run. Is there any paradox of saving in the medium run?

6. Suppose that the interest rate has no effect on investment.
   a. Can you think of a situation in which this may happen?
   b. What does this imply for the slope of the IS curve?
   c. What does this imply for the slope of the LM curve?
   d. What does this imply for the slope of the AD curve?

   Continue to assume that the interest rate has no effect on investment. Assume that the economy starts at the natural level of output. Suppose there is a shock to the variable z, so that the AS curve shifts up.
   e. What is the short-run effect on output and the price level? Explain in words.
   f. What happens to output and the price level over time? Explain in words.

7. Demand shocks and demand management

   Assume that the economy starts at the natural level of output. Now suppose there is a decline in business confidence, so that investment demand falls for any interest rate.
   a. In an AS–AD diagram, show what happens to output and the price level in the short run and the medium run.
   b. What happens to the unemployment rate in the short run? in the medium run?

   Suppose that the Federal Reserve decides to respond immediately to the decline in business confidence in the short run. In particular, suppose that the Fed wants to prevent the unemployment rate from changing in the short run after the decline in business confidence.
   a. What should the Fed do? Show how the Fed’s action, combined with the decline in business confidence, affects the AS–AD diagram in the short run and the medium run.
   b. How do short-run output and the short-run price level compare to your answers from part (a)?
   c. How do the short-run and medium-run unemployment rates compare to your answers from part (b)?

8. Supply shocks and demand management

   Assume that the economy starts at the natural level of output. Now suppose there is an increase in the price of oil.
   a. In an AS–AD diagram, show what happens to output and the price level in the short run and the medium run.
   b. What happens to the unemployment rate in the short run? in the medium run?

   Suppose that the Federal Reserve decides to respond immediately to the increase in the price of oil. In particular, suppose that the Fed wants to prevent the unemployment rate from changing in the short run after the increase in the price of oil. Assume that the Fed changes the money supply once—immediately after the increase in the price of oil—and then does not change the money supply again.
   c. What should the Fed do to prevent the unemployment rate from changing in the short run? Show how the Fed’s action, combined with the decline in business confidence, affects the AS–AD diagram in the short run and the medium run.

9. Based on your answers to Problems 7 and 8 and the material from the chapter, comment on the following statement: “The Federal Reserve has the easiest job in the world. All it has to do is conduct expansionary monetary policy when the unemployment rate increases and contractionary monetary policy when the unemployment rate falls.”

10. Taxes, oil prices, and workers

   Everyone in the labor force is concerned with two things: whether they have a job and, if so, their after-tax income from that job (i.e., their after-tax real wage). An unemployed worker may also be concerned with the availability and amount of unemployment benefits, but we will leave that issue aside for this problem.
   a. Suppose there is an increase in oil prices. How will this affect the unemployment rate in the short run and the medium run? How will it affect the real wage (W/P)?
   b. Suppose there is a reduction in income taxes. How will this affect the unemployment rate in the short run and the medium run? How about the real wage? For a given worker, how will after-tax income be affected?
   c. According to our model, what policy tools does the government have available to increase the real wage?
   d. During 2003 and 2004, oil prices increased more or less at the same time that income taxes were reduced. A popular joke at the time was that people could use their tax refunds to pay for the higher gas prices. How do your answers to this problem make sense of this joke?

EXPLORE FURTHER

11. Adding energy prices to the AS curve

   In this problem, we incorporate the price of energy inputs (e.g., oil) explicitly into the AS curve.

   Suppose the price-setting equation is given by

   \[ P = (1 + m)W^a P_{E}^{1-a} \]

   where \( P_{E} \) is the price of energy resources and \( 0 < a < 1 \).

   Ignoring a multiplicative constant, \( W^a P_{E}^{1-a} \) is the marginal cost function that would result from the production technology, \( Y = N^a E^{1-a} \), where \( N \) is employed labor and \( E \) represents units of energy resources used in production.

   As in the text, the wage-setting relation is given by

   \[ W = P^e F(u, z) \]

   Make sure to distinguish between \( P_{E} \), the price of energy resources, and \( P^e \), the expected price level for the economy as a whole.

   a. Substitute the wage-setting relation into the price-setting relation to obtain the aggregate supply relation.
   b. Let \( x = P_{E}/P \), the real price of energy. Observe that \( X \times x = P_{E} \) and substitute for \( P_{E} \) in the AS relation you derived in part (a). Solve for \( P \) to obtain

   \[ P = P^e (1 + m)^{1/a} F(u, z) x^{(1-a)/a} \]
c. Graph the AS relation from part (b) for a given $P^e$ and a given $x$.

d. Suppose that $P = P^e$. How will the natural rate of unemployment change if $x$, the real price of energy, increases? [Hint: You can solve the AS equation for $x$ to obtain the answer, or you can use your intuition. If $P = P^e$, how must $F(u, z)$ change when $x$ increases to maintain the equality in part (b)? How must $u$ change to have the required effect on $F(u, z)$?]

e. Suppose that the economy begins with output equal to the natural level of output. Then the real price of energy increases. Show the short-run and medium-run effects of the increase in the real price of energy in an AS–AD diagram.

The text suggests that a change in expectations about monetary policy may help explain why increases in oil prices over the past few years have had less of an adverse effect on the economy than the oil price shocks of the 1970s. Let us examine how such a change in expectations would alter the effect of an oil price shock.

f. Suppose there is an increase in the real price of energy. In addition, despite the increase in the real price of energy, suppose that the expected price level (i.e., $P^e$) does not change. After the short-run effect of the increase in the real price of energy, will there be any further adjustment of the economy over the medium run? In order for the expected price level not to change, what monetary action must wage-setters be expecting after an increase in the real price of energy?

12. Growth and fluctuations: some economic history

When economists think about history, fluctuations often stand out—oil shocks and stagflation in the 1970s, a recession followed by a long expansion in the 1980s, a recession followed by an extraordinary low-unemployment, low-inflation boom in the 1990s. This question puts these fluctuations into some perspective.

Go to the Web site of the Bureau of Economic Analysis (www.bea.gov) and retrieve the quarterly version of NIPA Table 1.1.6, real GDP in chained (2005) dollars. Get real GDP for the fourth quarter of 1959, 1969, 1979, 1989, 1999, 2000, and 2010 as well as for the fourth quarter of the most recent year available.

a. Using the real GDP numbers for 1959 and 1969, calculate the decadal growth rate of real GDP for the 1960s. Do the same for the 1970s, 1980s, 1990s, and the 2000s and for the available years of the most recent decade.

b. How does growth in the 1970s compare to growth in the later decades? How does growth in the 1960s compare to the later decades? Which decade looks most unusual?
In 1958, A. W. Phillips drew a diagram plotting the rate of inflation against the rate of unemployment in the United Kingdom for each year from 1861 to 1957. He found clear evidence of a negative relation between inflation and unemployment: When unemployment was low, inflation was high, and when unemployment was high, inflation was low, often even negative.

Two years later, Paul Samuelson and Robert Solow replicated Phillips’s exercise for the United States, using data from 1900 to 1960. Figure 8-1 reproduces their findings using CPI inflation as a measure of the inflation rate. Apart from the period of very high unemployment during the 1930s (the years from 1931 to 1939 are denoted by triangles and are clearly to the right of the other points in the figure), there also appeared to be a negative relation between inflation and unemployment in the United States. This relation, which Samuelson and Solow labeled the Phillips curve, rapidly became central to macroeconomic thinking and policy. It appeared to imply that countries could choose between different combinations of unemployment and inflation. A country could achieve low unemployment if it were willing to tolerate higher inflation, or it could achieve price level stability—zero inflation—if it were willing to tolerate higher unemployment. Much of the discussion about macroeconomic policy became a discussion about which point to choose on the Phillips curve.

In the 1970s, however, this relation broke down. In the United States and most OECD countries, there was both high inflation and high unemployment, clearly contradicting the original Phillips curve. A relation reappeared, but it reappeared as a relation between the unemployment rate and the change in the inflation rate. Today in the United States, high unemployment typically...
The Medium Run

The Core

leads not to low inflation, but to a decrease in inflation over time. Conversely, low unemployment doesn’t lead to high inflation, but to an increase in inflation over time.

The purpose of this chapter is to explore the mutations of the Phillips curve and, more generally, to understand the relation between inflation and unemployment. You will see that what Phillips discovered was the aggregate supply relation, and that the mutations of the Phillips curve came from changes in the way people and firms formed expectations.

The chapter has three sections:

Section 8-1 shows how we can think of the aggregate supply relation as a relation between inflation, expected inflation, and unemployment.

Section 8-2 uses this relation to interpret the mutations in the Phillips curve over time. It shows the relation between the actual unemployment rate, the natural unemployment rate, and inflation. It then draws two central implications of the Phillips curve for the medium run: In the medium run, unemployment returns to the natural rate, independent of inflation. The inflation rate is determined by the rate of money growth.

Section 8-3 further discusses the relation between unemployment and inflation across countries and over time.

8-1 Inflation, Expected Inflation, and Unemployment

Our first step will be to show that the aggregate supply relation we derived in Chapter 7 can be rewritten as a relation between inflation, expected inflation, and the unemployment rate.

To rewrite, go back to the aggregate supply relation between the price level, the expected price level, and output. As our focus is on the relation between unemployment (rather than output) and inflation, we do not need to take this step here.

The function $F$ comes from the wage-setting relation, equation (6.1):

$$W = P^eF(u, z).$$

We then replaced the unemployment rate by its expression in terms of output to obtain a relation between the price level, the expected price level, and output. As our focus is on the relation between unemployment (rather than output) and inflation, we do not need to take this step here.

The function $F$ comes from the wage-setting relation, equation (6.1):

$$W = P^eF(u, z).$$

Recall that the function $F$ captures the effects on the wage of the unemployment rate, $u$, and of the other factors that affect wage setting represented by the catchall variable $z$. $m$ is the markup of prices over wages. It will be convenient here to assume a specific form for this function:

$$F(u, z) = 1 - au + z$$
This captures the notion that the higher the unemployment rate, the lower the wage; and the higher $z$ (for example, the more generous unemployment benefits are), the higher the wage. The parameter $\alpha$ (the Greek lowercase letter alpha) captures the strength of the effect of unemployment on the wage.

Replace the function $F$ by this specific form in the aggregate supply relation above:

$$P = P^e(1 + m)(1 - \alpha u + z)$$

(8.1)

Finally, let $\pi$ denote the inflation rate, and $\pi^e$ denote the expected inflation rate. Then equation (8.1) can be rewritten as

$$\pi = \pi^e + (m + z) - \alpha u$$

(8.2)

Deriving equation (8.2) from equation (8.1) is not difficult, but it is tedious, so it is left to an appendix at the end of this chapter. What is important is that you understand each of the effects at work in equation (8.2):

- **An increase in expected inflation, $\pi^e$, leads to an increase in actual inflation, $\pi$.**
  
  To see why, start from equation (8.1). An increase in the expected price level $P^e$ leads, one for one, to an increase in the actual price level $P$: If wage setters expect a higher price level, they set a higher nominal wage, which leads to an increase in the price level.

  Now note that, given last period's price level, a higher price level this period implies a higher rate of increase in the price level from last period to this period—that is, higher inflation. Similarly, given last period's price level, a higher expected price level this period implies a higher expected rate of increase in the price level from last period to this period—that is, higher expected inflation. So the fact that an increase in the expected price level leads to an increase in the actual price level can be restated as: An increase in expected inflation leads to an increase in inflation.

- **Given expected inflation, $\pi^e$, an increase in the markup $m$, or an increase in the factors that affect wage determination—an increase in $z$—leads to an increase in inflation, $\pi$.**
  
  From equation (8.1): Given the expected price level $P^e$, an increase in either $m$ or $z$ increases the price level $P$. Using the same argument as in the previous bullet to restate this proposition in terms of inflation and expected inflation: Given expected inflation $\pi^e$, an increase in either $m$ or $z$ leads to an increase in inflation $\pi$.

- **Given expected inflation, $\pi^e$, an increase in the unemployment rate $u$ leads to a decrease in inflation $\pi$.**
  
  From equation (8.1): Given the expected price level $P^e$, an increase in the unemployment rate $u$ leads to a lower nominal wage, which leads to a lower price level $P$. Restating this in terms of inflation and expected inflation: Given expected inflation $\pi^e$, an increase in the unemployment rate $u$ leads to a decrease in inflation $\pi$.

We need one more step before we return to a discussion of the Phillips curve: When we look at movements in inflation and unemployment in the rest of the chapter, it will often be convenient to use time indexes so that we can refer to variables like inflation, or expected inflation, or unemployment, in a specific year. So we rewrite equation (8.2) as:

$$\pi_t = \pi^e_t + (m + z) - \alpha u_t$$

(8.3)

The variables $\pi_t$, $\pi^e_t$, and $u_t$ refer to inflation, expected inflation, and unemployment in year $t$. Be sure you see that there are no time indexes on $m$ and $z$. This is because we
shall typically think of both $m$ and $z$ as constant while we look at movements in inflation, expected inflation, and unemployment over time.

### 8-2 The Phillips Curve

Let’s start with the relation between unemployment and inflation as it was first discovered by Phillips, Samuelson, and Solow, around 1960.

#### The Early Incarnation

Imagine an economy where inflation is positive in some years, negative in others, and is on average equal to zero. This is not the way things have been for some time: Since 1960, inflation has been positive in all years but one, 2009, when it was negative, but small. But as we shall see later in this chapter, average inflation was close to zero during much of the period Phillips, Samuelson, and Solow were studying.

In such an environment, how will wage setters choose nominal wages for the coming year? With the average inflation rate equal to zero in the past, it is reasonable for wage setters to expect that inflation will be equal to zero over the next year as well. So, let’s assume that expected inflation is equal to zero—that $\pi^e_t = 0$. Equation (8.3) then becomes

$$\pi_t = (m + z) - \alpha u_t$$

(8.4)

This is precisely the negative relation between unemployment and inflation that Phillips found for the United Kingdom and Solow and Samuelson found for the United States. The story behind it is simple: Given the expected price level, which workers simply take to be last year’s price level, lower unemployment leads to a higher nominal wage. A higher nominal wage leads to a higher price level. Putting the steps together, lower unemployment leads to a higher price level this year relative to last year’s price level—that is, to higher inflation. This mechanism has sometimes been called the **wage–price spiral**, an expression that captures well the basic mechanism at work:

- Low unemployment leads to a higher nominal wage.
- In response to the higher nominal wage, firms increase their prices. The price level increases.
- In response to the higher price level, workers ask for a higher nominal wage the next time the wage is set.
- The higher nominal wage leads firms to further increase their prices. As a result, the price level increases further.
- In response to this further increase in the price level, workers, when they set the wage again, ask for a further increase in the nominal wage.
- And so the race between prices and wages results in steady wage and price inflation.

#### Mutations

The combination of an apparently reliable empirical relation, together with a plausible story to explain it, led to the adoption of the Phillips curve by macroeconomists and policy makers. During the 1960s, U.S. macroeconomic policy was aimed at maintaining unemployment in the range that appeared consistent with moderate inflation. And, throughout the 1960s, the negative relation between unemployment and inflation provided a reliable guide to the joint movements in unemployment and inflation.
Figure 8-2 plots the combinations of the inflation rate and the unemployment rate in the United States for each year from 1948 to 1969. Note how well the Phillips relation held during the long economic expansion that lasted throughout most of the 1960s. During the years 1961 to 1969, denoted by black diamonds in the figure, the unemployment rate declined steadily from 6.8% to 3.4%, and the inflation rate steadily increased, from 1.0% to 5.5%. Put informally, from 1961 to 1969, the U.S. economy moved up along the Phillips curve.

Around 1970, however, the relation between the inflation rate and the unemployment rate, so visible in Figure 8-2, broke down. Figure 8-3 shows the combination of the inflation rate and the unemployment rate in the United States for each year since 1970. The points are scattered in a roughly symmetric cloud: There is no visible relation between the unemployment rate and the inflation rate.

Why did the original Phillips curve vanish? There are two main reasons:

- The United States was hit twice in the 1970s by a large increase in the price of oil (see Chapter 7). The effect of this increase in nonlabor costs was to force firms to increase their prices relative to the wages they were paying—in other words, to increase the markup $m$. As shown in equation (8.3), an increase in $m$ leads to an increase in inflation, even at a given rate of unemployment, and this happened twice in the 1970s. The main reason for the breakdown of the Phillips curve relation, however, lay elsewhere:

- Wage setters changed the way they formed their expectations. This change came, in turn, from a change in the behavior of inflation. Look at Figure 8-4, which shows the U.S. inflation rate since 1914. Starting in the 1960s (the decade shaded in the figure), you can see a clear change in the behavior of the rate of inflation. First, rather than being sometimes positive and sometimes negative, as it had for the first part of the century, the rate of inflation became consistently positive. Second, inflation became more persistent: High inflation in one year became more likely to be followed by high inflation the next year.

- The persistence of inflation led workers and firms to revise the way they formed their expectations. When inflation is consistently positive year after year, expecting...
Inflation rate (percent)


Inflation rate (percent)

Unemployment rate (percent)

Inflation versus Unemployment in the United States, 1970–2010

Beginning in 1970, the relation between the unemployment rate and the inflation rate disappeared in the United States.

Source: See Figure 8-2.

that the price level this year will be the same as the price level last year—which is the same as expecting zero inflation—becomes systematically incorrect; worse, it becomes foolish. People do not like to make the same mistake repeatedly. So, as inflation became consistently positive and more persistent, people, when forming expectations, started to take into account the presence and the persistence of inflation. This change in expectation formation changed the nature of the relation between unemployment and inflation.

U.S. Inflation, since 1914

Since the 1960s, the U.S. inflation rate has been consistently positive. Inflation has also become more persistent: A high inflation rate this year is more likely to be followed by a high inflation rate next year.

Let’s look at the argument in the previous paragraph more closely. First, suppose expectations of inflation are formed according to

\[ \pi_t^e = \theta \pi_{t-1} \]  

(8.5)

The value of the parameter \( \theta \) (the Greek lowercase theta) captures the effect of last year’s inflation rate, \( \pi_{t-1} \), on this year’s expected inflation rate, \( \pi_t^e \). The higher the value of \( \theta \), the more last year’s inflation leads workers and firms to revise their expectations of what inflation will be this year, and so the higher the expected inflation rate is. We can think of what happened in the 1970s as an increase in the value of \( \theta \) over time:

- As long as inflation was low and not very persistent, it was reasonable for workers and firms to ignore past inflation and to assume that the price level this year would be roughly the same as price level last year. For the period that Samuelson and Solow had looked at, \( \theta \) was close to zero, and expectations were roughly given by \( \pi_t^e = 0 \).
- But, as inflation became more persistent, workers and firms started changing the way they formed expectations. They started assuming that if inflation had been high last year, inflation was likely to be high this year as well. The parameter \( \theta \), the effect of last year’s inflation rate on this year’s expected inflation rate, increased. The evidence suggests that, by the mid-1970s, people expected this year’s inflation rate to be the same as last year’s inflation rate—in other words, that \( \theta \) was now equal to 1.

Now turn to the implications of different values of \( \theta \) for the relation between inflation and unemployment. To do so, substitute equation (8.5) for the value of \( \pi_t^e \) into equation (8.3):

\[ \pi_t = \theta \pi_{t-1} + (m + z) - \alpha u_t \]

- When \( \theta \) equals zero, we get the original Phillips curve, a relation between the inflation rate and the unemployment rate:
  \[ \pi_t = (m + z) - \alpha u_t \]
- When \( \theta \) is positive, the inflation rate depends not only on the unemployment rate but also on last year’s inflation rate:
  \[ \pi_t = \theta \pi_{t-1} + (m + z) - \alpha u_t \]
- When \( \theta \) equals 1, the relation becomes (moving last year’s inflation rate to the left side of the equation)
  \[ \pi_t - \pi_{t-1} = (m + z) - \alpha u_t \]  

(8.6)

So, when \( \theta = 1 \), the unemployment rate affects not the inflation rate, but rather the change in the inflation rate: High unemployment leads to decreasing inflation; low unemployment leads to increasing inflation.

This discussion is the key to what happened from 1970 onward. As \( \theta \) increased from 0 to 1, the simple relation between the unemployment rate and the inflation rate disappeared. This disappearance is what we saw in Figure 8-3. But a new relation emerged, this time between the unemployment rate and the change in the inflation rate.
rate—as predicted by equation (8.6). This relation is shown in Figure 8-5, which plots the change in the inflation rate versus the unemployment rate observed for each year since 1970. The figure shows a negative relation between the unemployment rate and the change in the inflation rate. The line that best fits the scatter of points for the period 1970–2010 is given by

$$\pi_t - \pi_{t-1} = 3.3\% - 0.55u_t$$  \hspace{1cm} (8.7)

The line is drawn in Figure 8-5. For low unemployment, the change in inflation is positive. For high unemployment, the change in inflation is negative. This is the form the Phillips curve relation between unemployment and inflation takes today.

To distinguish it from the original Phillips curve (equation (8.4)), equation (8.6) or its empirical counterpart, equation (8.7) is often called the **modified Phillips curve**, or the **expectations-augmented Phillips curve** (to indicate that $$\pi_{t-1}$$ stands for expected inflation), or the **accelerationist Phillips curve** (to indicate that a low unemployment rate leads to an increase in the inflation rate and thus an acceleration of the price level).

We shall simply call equation (8.6) the Phillips curve and refer to the earlier incarnation, equation (8.4), as the **original Phillips curve**.

---

**Figure 8-5**

*Change in Inflation versus Unemployment in the United States, 1970–2010*

Since 1970, there has been a negative relation between the unemployment rate and the change in the inflation rate in the United States.

Source: Series CPIUNSCL, UNRATE. Federal Reserve Economic Data (FRED) http://research.stlouisfed.org/fred2/
The Phillips Curve and the Natural Rate of Unemployment

The history of the Phillips curve is closely related to the discovery of the concept of the natural unemployment rate that we introduced in Chapter 6.

The original Phillips curve implied that there was no such thing as a natural unemployment rate: If policy makers were willing to tolerate a higher inflation rate, they could maintain a lower unemployment rate forever.

In the late 1960s, while the original Phillips curve still gave a good description of the data, two economists, Milton Friedman and Edmund Phelps, questioned the existence of such a trade-off between unemployment and inflation. They questioned it on logical grounds, arguing that such a trade-off could exist only if wage setters systematically underpredicted inflation, and that they were unlikely to make the same mistake forever. Friedman and Phelps also argued that if the government attempted to sustain lower unemployment by accepting higher inflation, the trade-off would ultimately disappear; the unemployment rate could not be sustained below a certain level, a level they called the “natural rate of unemployment.” Events proved them right, and the trade-off between the unemployment rate and the inflation rate indeed disappeared. (See the Focus box “Theory ahead of the Facts: Milton Friedman and Edmund Phelps.”) Today, most economists accept the notion of a natural rate of unemployment—subject to the many caveats we will see in the next section.

Let’s make explicit the connection between the Phillips curve and the natural rate of unemployment.

By definition (see Chapter 6), the natural rate of unemployment is the unemployment rate such that the actual price level is equal to the expected price level. Equivalently, and more conveniently here, the natural rate of unemployment is the unemployment rate such that the actual inflation rate is equal to the expected inflation rate. Denote the natural unemployment rate by $u_n$ (the index $n$ stands for “natural”). Then, imposing the condition that actual inflation and expected inflation be the same ($\pi_t = \pi^e_t$) in equation (8.3) gives

$$0 = (m + z) - \alpha u_n$$

Solving for the natural rate $u_n$,

$$u_n = \frac{m + z}{\alpha}$$

(8.8)

The higher the markup, $m$, or the higher the factors that affect wage setting, $z$, the higher the natural rate of unemployment.

Now rewrite equation (8.3) as

$$\pi_t - \pi^e_t = - \alpha \left( u_t - \frac{m + z}{\alpha} \right)$$

Note from equation (8.8) that the fraction on the right side is equal to $u_n$, so we can rewrite the equation as

$$\pi_t - \pi^e_t = - \alpha (u_t - u_n)$$

(8.9)

If—as is the case in the United States today—the expected rate of inflation, $\pi^e_t$, is well approximated by last year’s inflation rate, $\pi_{t-1}$, the equation finally becomes

$$\pi_t - \pi_{t-1} = -\alpha (u_t - u_n)$$

(8.10)
Equation (8.10) is an important relation, for two reasons:

- It gives us another way of thinking about the Phillips curve, as a relation between the actual unemployment rate $u_t$, the natural unemployment rate $u_n$, and the change in the inflation rate $\pi_t - \pi_{t-1}$:

  The change in the inflation rate depends on the difference between the actual and the natural unemployment rates. When the actual unemployment rate is higher than the natural unemployment rate, the inflation rate decreases; when the actual unemployment rate is lower than the natural unemployment rate, the inflation rate increases.

- It also gives us another way of thinking about the natural rate of unemployment.

The natural rate of unemployment is the rate of unemployment required to keep the inflation rate constant. This is why the natural rate is also called the non-accelerating inflation rate of unemployment, or NAIRU.

What has been the natural rate of unemployment in the United States since 1970? Put another way: What has been the unemployment rate that, on average, has led to constant inflation?

To answer this question, all we need to do is to return to equation (8.7), the estimated relation between the change in inflation and the unemployment rate since 1970. Setting the change in inflation equal to zero in that equation implies a value for the natural unemployment rate of $3.3%/0.55 = 6%$. In words: The evidence suggests that, since 1970 in the United States, the average rate of unemployment required to keep inflation constant has been equal to 6%.
The Neutrality of Money, Revisited

In Chapter 7, we looked at the effects of a change in the level of nominal money on output and on the price level in the medium run. We derived two propositions. First, output returned to its natural level, unaffected by the level of nominal money. Second, the price level moved in proportion to the nominal money stock, leaving the real money stock unchanged. We are now ready to extend these results and look at the effects of changes in the rate of growth of nominal money on unemployment and on inflation in the medium run.

Let us first look at unemployment and go back to equation (8.9). In the medium run, expected inflation must be equal to actual inflation. Thus the unemployment rate must be equal to the natural rate, which is clearly independent of the rate of growth of money.

Now turn to inflation, and go back to the aggregate demand relation we derived in Chapter 7, equation (7.3):

\[ Y = Y \left( \frac{M}{P}, G, T \right) \]

If unemployment returns to the natural rate, \( u_n \), it follows that output must return to its natural level, \( Y_n \). So the relation becomes:

\[ Y_n = Y \left( \frac{M}{P}, G, T \right) \]

If \( Y_n \) is constant, for this equality to hold (that is, for aggregate demand to be equal to the natural level of output), the right-hand side of the equation must be constant. If we assume unchanged fiscal policy (that is, constant \( G \) and constant \( T \)), this implies that the real money stock must also be constant. This implies in turn that the rate of inflation must be equal to the rate of money growth.

\[ \pi = g_M \]

This is an important result: In the medium run, the rate of inflation is determined by the rate of money growth. Milton Friedman put it this way: Inflation is always and everywhere a monetary phenomenon. As we have seen, factors such as the monopoly power of firms, strong unions, strikes, fiscal deficits, and increases in the price of oil do affect the price level and, by implication, do affect inflation in the short run. But, unless they affect the rate of money growth, they have no effect on inflation in the medium run.

8-3 A Summary and Many Warnings

Let’s take stock of what we have learned about the relation between inflation and unemployment:

- The aggregate supply relation is well captured in the United States today by a relation between the change in the inflation rate and the deviation of the unemployment rate from the natural rate of unemployment (equation (8.8)).

- When the unemployment rate exceeds the natural rate of unemployment, the inflation rate typically decreases. When the unemployment rate is below the natural rate of unemployment, the inflation rate typically increases.

This relation has held quite well since 1970. But evidence from its earlier history, as well as the evidence from other countries, points to the need for a number of warnings.
All of them are on the same theme: The relation between inflation and unemployment can and does vary across countries and time.

**Variations in the Natural Rate across Countries**

Recall from equation (8.8) that the natural rate of unemployment depends on: all the factors that affect wage setting, represented by the catchall variable $z$; the markup set by firms $m$; and the response of inflation to unemployment, represented by $\alpha$. If these factors differ across countries, there is no reason to expect all countries to have the same natural rate of unemployment. And natural rates indeed differ across countries, sometimes considerably.

Take, for example, the unemployment rate in the Euro area, which has averaged close to 9% since 1980. A high unemployment rate for a few years may well reflect a deviation of the unemployment rate from the natural rate. A high average unemployment rate for 30 years surely reflects a high natural rate. This tells us where we should look for explanations: in the factors determining the wage-setting and the price-setting relations.

Is it easy to identify the relevant factors? One often hears the statement that one of the main problems of Europe is its labor-market rigidities. These rigidities, the argument goes, are responsible for its high unemployment. While there is some truth to this statement, the reality is more complex. The Focus box, “What Explains European Unemployment?” discusses these issues further.

**Variations in the Natural Rate over Time**

In writing equation (8.6) and estimating equation (8.7), we treated $m + z$ as a constant. But there are good reasons to believe that $m$ and $z$ vary over time. The degree of monopoly power of firms, the structure of wage bargaining, the system of unemployment benefits, and so on are likely to change over time, leading to changes in either $m$ or $z$ and, by implication, changes in the natural rate of unemployment.

Changes in the natural unemployment rate over time are hard to measure. Again, the reason is that we do not observe the natural rate, only the actual rate. But broad evolutions can be established by comparing average unemployment rates, say across decades. Using this approach, the Focus box “What Explains European Unemployment?” discusses how and why the natural rate of unemployment had increased in Europe since the 1960s. The U.S. natural rate has moved much less than that in Europe. Nevertheless, it is also far from constant. Go back and look at Figure 6-3. You can see that, from the 1950s to the 1980s, the unemployment rate fluctuated around a slowly increasing trend: Average unemployment was 4.5% in the 1950s, and 7.3% in the 1980s. Then, from 1990 on, and until the crisis, the trend was reversed, with an average unemployment rate of 5.7% in the 1990s, and an average unemployment rate of 5.0% from 2000 to 2007. In 2007, the unemployment rate was 4.6%, and inflation was roughly constant, suggesting that unemployment was close to the natural rate. Why the U.S. natural rate of unemployment fell from the early 1990s on and what the effects of the crisis may be for the future are discussed in the Focus box “Changes in the U.S. Natural Rate of Unemployment since 1990.” We draw two conclusions from the behavior of the U.S. unemployment rate since 1990 and these conclusions parallel the conclusion from our look at European unemployment in the earlier box: The determinants of the natural rate are many. We can identify a number of them, but knowing their respective role and drawing policy lessons is not easy.

**Disinflation, Credibility, and Unemployment**

In 1979, U.S. unemployment was 5.8%, roughly equal to the natural rate at the time. But the inflation rate, measured using the CPI, was running above 13%. Some of this
What Explains European Unemployment?

What do critics have in mind when they talk about the “labor-market rigidities” afflicting Europe? They have in mind in particular:

- A generous system of unemployment insurance. The replacement rate—that is, the ratio of unemployment benefits to the after-tax wage—is often high in Europe, and the duration of benefits—the period of time for which the unemployed are entitled to receive benefits—often runs in years.

  Some unemployment insurance is clearly desirable. But generous benefits are likely to increase unemployment in at least two ways: They decrease the incentives the unemployed have to search for jobs. They may also increase the wage that firms have to pay. Recall our discussion of efficiency wages in Chapter 6. The higher unemployment benefits are, the higher the wages firms have to pay in order to motivate and keep workers.

- A high degree of employment protection. By employment protection, economists have in mind the set of rules that increase the cost of layoffs for firms. These range from high severance payments, to the need for firms to justify layoffs, to the possibility for workers to appeal the decision and have it reversed.

  The purpose of employment protection is to decrease layoffs, and thus to protect workers from the risk of unemployment. What it also does, however, is to increase the cost of labor for firms and thus to reduce hires and make it harder for the unemployed to get jobs. The evidence suggests that, while employment protection does not necessarily increase unemployment, it changes its nature: The flows in and out of unemployment decrease, but the average duration of unemployment increases. Such long duration increases the risk that the unemployed lose skills and morale, decreasing their employability.

- Minimum wages. Most European countries have national minimum wages. And in some countries, the ratio of the minimum wage to the median wage can be quite high. High minimum wages clearly run the risk of decreasing employment for the least-skilled workers, thus increasing their unemployment rate.

- Bargaining rules. In most European countries, labor contracts are subject to extension agreements. A contract agreed to by a subset of firms and unions can be automatically extended to all firms in the sector. This considerably reinforces the bargaining power of unions, as it reduces the scope for competition by nonunionized firms. As we saw in Chapter 6, stronger bargaining power on the part of the unions may result in higher unemployment: Higher unemployment is needed to reconcile the demands of workers with the wages paid by firms.

Do these labor-market institutions really explain high unemployment in Europe? Is the case open and shut? Not quite. Here it is important to recall two important facts:

  Fact 1: Unemployment was not always high in Europe. In the 1960s, the unemployment rate in the four major continental European countries was lower than that in the United States, around 2–3%. U.S. economists would cross the ocean to study the “European unemployment miracle!” The natural rate in these countries today is around 8–9%. How do we explain this increase?

  One hypothesis is that institutions were different then, and that labor-market rigidities have only appeared in the last 40 years. This turns out not to be the case, however. It is true that, in response to the adverse shocks of the 1970s (in particular the two recessions following the increases in the price of oil), many European governments increased the generosity of unemployment insurance and the degree of employment protection. But, even in the 1960s, European labor-market institutions looked nothing like U.S. labor-market institutions. Social protection was much higher in Europe; yet unemployment was lower.

  A more convincing line of explanation focuses on the interaction between institutions and shocks. Some labor-market institutions may be benign in some environments, yet very costly in others. Take employment protection. If competition between firms is limited, the need to adjust employment in each firm may be limited as well, and so the cost of employment protection may be low. But if competition, either from other domestic firms or from foreign firms, increases, the cost of employment protection may become very high. Firms that cannot adjust their labor force quickly may simply be unable to compete and go out of business.

  Fact 2: Until the current crisis started, a number of European countries actually had low unemployment. This is shown in Figure 1, which gives the unemployment rate for 15 European countries (the 15 members of the European Union before the increase in membership to 27) in 2006. We chose 2006 because, in all these countries, inflation was stable, suggesting that the unemployment rate was roughly equal to the natural rate.

  As you can see, the unemployment rate was indeed high in the large four continental countries: France, Spain, Germany, and Italy. But note how low the unemployment rate was in some of the other countries, in particular Denmark, Ireland, and the Netherlands.

  Is it the case that these low unemployment countries had low benefits, low employment protection, weak unions? Things are unfortunately not so simple: Countries such as Ireland or the United Kingdom indeed have labor-market institutions that resemble those of the United States: limited benefits, low employment protection does not necessarily increase unemployment, and the degree of employment protection is low.
protection, and weak unions. But countries like Denmark or the Netherlands have a high degree of social protection (in particular high unemployment benefits) and strong unions.

So what is one to conclude? An emerging consensus among economists is that the devil is in the details: Generous social protection is consistent with low unemployment. But it has to be provided efficiently. For example, unemployment benefits can be generous, so long as the unemployed are, at the same time, forced to take jobs if such jobs are available. Some employment protection (for example, in the form of generous severance payments) may be consistent with low unemployment, so long as firms do not face the prospect of long administrative or judicial uncertainty when they lay off workers. Countries such as Denmark appear to have been more successful in achieving these goals. Creating incentives for the unemployed to take jobs and simplifying the rules of employment protection are on the reform agenda of many European governments. One may hope they will lead to a decrease in the natural rate in the future.


was due to the large increase in oil prices, but, leaving this effect aside, underlying inflation was running at close to 10%. The question the Federal Reserve faced was no longer whether or not it should reduce inflation, but how fast it should reduce it. In August 1979, President Carter appointed Paul Volcker as Chairman of the Federal Reserve Board. Volcker, who had served in the Nixon administration, was considered an extremely qualified chairman who would and could lead the fight against inflation.

Fighting inflation implied tightening monetary policy, decreasing output growth, and thus accepting higher unemployment for some time. The question arose of how much unemployment, and for how long, would likely be needed to achieve a lower level of inflation, say 4%—which is the rate Volcker wanted to achieve.

Some economists argued that such a disinflation would likely be very costly. Their starting point was equation (8.10):

$$p_t - p_{t-1} = -\alpha(u_t - u_n)$$
Changes in the U.S. Natural Rate of Unemployment since 1990

From 2000 to 2007, the average unemployment rate was under 5%, and inflation was stable. It thus appeared that the natural rate was around 5%, so roughly 2% lower than it had been in the 1980s. Since the beginning of the crisis, the unemployment rate has increased to reach more than 9% and is forecast to remain high for many years. While most of this increase clearly reflects an increase in the actual unemployment rate over the natural rate, the long period of high actual unemployment raises the issue of whether we can hope to go back to the pre-crisis rate.

Start with the first issue, namely why the natural rate decreased from the 1980s on. Researchers have offered a number of explanations:

- Increased globalization and stronger competition between U.S. and foreign firms may have led to a decrease in monopoly power and a decrease in the markup. Also, the fact that firms can more easily move some of their operations abroad surely makes them stronger when bargaining with their workers. The evidence that unions in the U.S. economy are becoming weaker: The unionization rate in the United States, which stood at 25% in the mid 1970s, is below 15% today. As we saw, weaker bargaining power on the part of workers is likely to lead to lower unemployment.

- The nature of the labor market has changed. In 1980, employment by temporary help agencies accounted for less than 0.5% of total U.S. employment. Today, it accounts for more than 2%. This is also likely to have reduced the natural rate of unemployment. In effect, it allows many workers to look for jobs while being employed rather than unemployed. The increasing role of Internet-based job sites, such as Monster.com, has also made matching of jobs and workers easier, leading to lower unemployment.

Some of the other explanations may surprise you. For example, researchers have also pointed to:

- The aging of the U.S. population. The proportion of young workers (workers between the ages of 16 and 24) fell from 24% in 1980 to 14% in 2006. This reflects the end of the baby boom, which ended in the mid-1960s. Young workers tend to start their working life by going from job to job and typically have a higher unemployment rate. So, a decrease in the proportion of young workers leads to a decrease in the overall unemployment rate.

- An increase in the incarceration rate. The proportion of the population in prison or in jail has tripled in the last 20 years in the United States. In 1980, 0.3% of the U.S. population of working age was in prison. In 2006, the proportion had increased to 1.0%. Because many of those in prison would likely have been unemployed were they not incarcerated, this is likely to have had an effect on the unemployment rate.

- The increase in the number of workers on disability. A relaxation of eligibility criteria since 1984 has led to a steady increase in the number of workers receiving disability insurance, from 2.2% of the working age population in 1984 to 3.8% in 2006. It is again likely that, absent changes in the rules, some of the workers on disability insurance would have been unemployed instead.

Will the natural rate of unemployment remain low in the future? Globalization, aging, prisons, temporary help agencies, and the increasing role of the Internet are probably here to stay, suggesting that the natural rate should indeed remain low. Since the beginning of the crisis, however, there is an increasing worry, that the increase in the actual unemployment rate may eventually translate into an increase in the natural unemployment rate. The mechanism through which this may happen is known as hysteresis (in economics, hysteresis is used to mean that, after a shock, a variable does not return to its initial value, even when the shock has gone away): Workers who have been unemployed for a long time may lose their skills, or their morale, and become, in effect, unemployable, leading to a higher natural rate. This is a very relevant concern. As we saw in Chapter 6, in 2010, the average duration of unemployment was 33 weeks, an exceptionally high number by historical standards. Forty-three percent of the unemployed had been unemployed for more than six months, 28% for more than a year. This raises two important questions. First, when the economy picks up, how many of them will be scarred by their unemployment experience and hard to reemploy? Second, are there policies which should be put in place now to help the long-term unemployed get back to work?


According to this equation, the only way to bring down inflation is to accept unemployment above the natural rate for some time. We saw earlier that \( \alpha \) is estimated to be equal to 0.55. The equation therefore implies that, to decrease inflation by one percentage point, the unemployment rate has to be higher than the natural
unemployment rate by $1/0.55$, or about 1.8 percentage points for a year. Or to decrease inflation from 10% to 4%, the unemployment rate has to be higher than the natural rate by about $10 \left(\frac{10 - 4}{0.55}\right)$ percentage points for a year, or, more realistically, if inflation was decreased from 10% to 6% in five years, by about $2 \left(\frac{10}{5}\right)$ percentage points for five years. The natural conclusion was that it would make sense to go slowly, so as not to increase unemployment by too much in a given year.

Some economists argued that disinflation might in fact be much less costly. In what has become known as the **Lucas critique**, Lucas pointed out that when trying to predict the effects of a major policy change—like the change considered by too much the Fed at the time—it could be very misleading to take as given the relations estimated from past data. In the case of the Phillips curve, taking equation (8.10) as given was equivalent to assuming that wage setters would keep expecting inflation in the future to be the same as it was in the past, that the way wage setters formed their expectations would not change in response to the change in policy. This was an unwarranted assumption, Lucas argued: Why shouldn’t wage setters take policy changes directly into account? If wage setters believed that the Fed was committed to lower inflation, they might well expect inflation to be lower in the future than in the past. They argued that the relevant equation was not equation (8.10) but equation (8.9). And equation (8.9) implied that, if the Fed was fully credible, the decrease in inflation might not require any increase in the unemployment rate. If wage setters expected inflation to now be 4%, then actual inflation would decrease to 4%, with unemployment remaining at the natural rate:

$$\pi_t = \pi_e - \alpha(u_t - u_n)$$

$$4\% = 4\% - 0\%$$

Lucas did not believe that disinflation could really take place without some increase in unemployment. But Thomas Sargent, looking at the historical evidence on the end of several very high inflations, concluded that the increase in unemployment could be small. The essential ingredient of successful disinflation, he argued, was **credibility** of monetary policy—the belief by wage setters that the central bank was truly committed to reducing inflation. Only credibility would cause wage setters to change the way they formed their expectations. Furthermore, he argued, a clear and quick disinflation program was more likely to be credible than a protracted one that offered plenty of opportunities for reversal and political infighting along the way.

Who turned out to be right? In September 1979, Paul Volcker started increasing the interest rate so as to slow down the economy and reduce inflation. From 9% in 1979, the three-month Treasury bill rate was increased to 15% in August 1981. The effects on inflation, output growth, and unemployment are shown in Table 8-1. The table makes clear that there was no credibility miracle: Disinflation was associated with a sharp recession, with negative growth in both 1980 and 1982, and with a large and long-lasting increase in unemployment. The average unemployment rate was above 9% in both 1982 and 1983.

<table>
<thead>
<tr>
<th>Year</th>
<th>CPI inflation (%)</th>
<th>GDP growth (%)</th>
<th>Unemployment rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>13.3</td>
<td>2.5</td>
<td>5.8</td>
</tr>
<tr>
<td>1980</td>
<td>12.5</td>
<td>–0.5</td>
<td>7.1</td>
</tr>
<tr>
<td>1981</td>
<td>8.9</td>
<td>1.8</td>
<td>7.6</td>
</tr>
<tr>
<td>1982</td>
<td>3.8</td>
<td>–2.2</td>
<td>9.7</td>
</tr>
<tr>
<td>1983</td>
<td>3.8</td>
<td>3.9</td>
<td>9.6</td>
</tr>
<tr>
<td>1984</td>
<td>3.9</td>
<td>6.2</td>
<td>7.5</td>
</tr>
<tr>
<td>1985</td>
<td>3.8</td>
<td>3.2</td>
<td>7.2</td>
</tr>
</tbody>
</table>
peaking at 10.8% in the month of December 1982. If we assume a natural rate of 6% and add the excess of unemployment above the natural rate (so 7.1% − 6.0% = 1.1% for 1980, 7.6% − 6.0% = 1.6% for 1981, etc.), total excess unemployment from 1979 to 1985 was 12.7%, a number no better than what equation (8.10) predicted.

Does this settle the issue of how much credibility matters? Not really. Those who argued before the fact that credibility would help argued after the fact that Volcker had not been fully credible. After increasing the interest rate from September 1979 to April 1980 and inducing a sharp decrease in growth, he appeared to have second thoughts, reversing course and sharply decreasing the interest rate from April to September, only to increase it again in 1981. This lack of consistency, some argued, reduced his credibility and increased the unemployment cost of the disinflation. A larger lesson still stands: The behavior of inflation depends very much on how people and firms form expectations. The Lucas critique still stands: The past relation between unemployment and inflation may be a poor guide to what happens when policy changes.

**High Inflation and the Phillips Curve Relation**

Recall how, in the 1970s, the U.S. Phillips curve changed as inflation became more persistent and wage setters changed the way they formed inflation expectations. The lesson is a general one: The relation between unemployment and inflation is likely to change with the level and the persistence of inflation. Evidence from countries with high inflation confirms this lesson. Not only does the way workers and firms form their expectations change, but so do institutional arrangements:

When the inflation rate becomes high, inflation also tends to become more variable. As a result, workers and firms become more reluctant to enter into labor contracts that set nominal wages for a long period of time: If inflation turns out higher than expected, real wages may plunge and workers will suffer a large cut in their living standard. If inflation turns out lower than expected, real wages may go up sharply. Firms may not be able to pay their workers. Some may go bankrupt.

For this reason, the terms of wage agreements change with the level of inflation. Nominal wages are set for shorter periods of time, down from a year to a month or even less. **Wage indexation**, a provision that automatically increases wages in line with inflation, becomes more prevalent.

These changes lead in turn to a stronger response of inflation to unemployment. To see this, an example based on wage indexation will help. Imagine an economy that has two types of labor contracts. A proportion $\lambda$ (the Greek lowercase letter lambda) of labor contracts is indexed: Nominal wages in those contracts move one for one with variations in the actual price level. A proportion $1 - \lambda$ of labor contracts is not indexed: Nominal wages are set on the basis of expected inflation.

Under this assumption, equation (8.9) becomes

$$\pi_t = [\lambda \pi_t + (1 - \lambda) \pi_t^e] - \alpha (u_t - u_n)$$

The term in brackets on the right reflects the fact that a proportion $\lambda$ of contracts is indexed and thus responds to actual inflation, $\pi_t$, and a proportion $1 - \lambda$, responds to expected inflation, $\pi_t^e$. If we assume that this year’s expected inflation is equal to last year’s actual inflation, $\pi_t^e = \pi_{t-1}$, we get

$$\pi_t = [\lambda \pi_t + (1 - \lambda) \pi_{t-1}] - \alpha (u_t - u_n)$$ (8.11)

When $\lambda = 0$, all wages are set on the basis of expected inflation—which is equal to last year’s inflation, $\pi_{t-1}$—and the equation reduces to equation (8.10):

$$\pi_t - \pi_{t-1} = -\alpha (u_t - u_n)$$

More concretely: When inflation runs on average at 3% a year, wage setters can be reasonably confident inflation will be between 1% and 5%. When inflation runs on average at 30% a year, wage setters can be confident inflation will be between 20% and 40%. In the first case, the real wage may end up 2% higher or lower than they expected when they set the nominal wage. In the second case, it may end up 10% higher or lower than they expected. There is much more uncertainty in the second case.
When \( \lambda \) is positive, however, a proportion \( \lambda \) of wages is set on the basis of actual inflation rather than expected inflation. To see what this implies, reorganize equation (8.11): Move the term in brackets to the left, factor \((1 - \lambda)\) on the left of the equation, and divide both sides by \(1 - \lambda\) to get

\[
\pi_t - \pi_{t-1} = -\frac{\alpha}{(1 - \lambda)}(u_t - u_n)
\]

Wage indexation increases the effect of unemployment on inflation. The higher the proportion of wage contracts that are indexed—the higher \( \lambda \)—the larger the effect the unemployment rate has on the change in inflation—the higher the coefficient \( \alpha/(1 - \lambda) \).

The intuition is as follows: Without wage indexation, lower unemployment increases wages, which in turn increases prices. But because wages do not respond to prices right away, there is no further increase in prices within the year. With wage indexation, however, an increase in prices leads to a further increase in wages within the year, which leads to a further increase in prices, and so on, so that the effect of unemployment on inflation within the year is higher.

If, and when, \( \lambda \) gets close to 1—which is when most labor contracts allow for wage indexation—small changes in unemployment can lead to very large changes in inflation. Put another way, there can be large changes in inflation with nearly no change in unemployment. This is what happens in countries where inflation is very high: The relation between inflation and unemployment becomes more and more tenuous and eventually disappears altogether.

Deflation and the Phillips Curve Relation

We have just looked at what happens to the Phillips curve when inflation is very high. Another issue is what happens when inflation is low, and possibly negative—when there is deflation.

The motivation for asking this question is given by an aspect of Figure 8-1 we mentioned at the start of the chapter but then left aside. In that figure, note how the points corresponding to the 1930s (they are denoted by triangles) lie to the right of the others. Not only is unemployment unusually high—this is no surprise because we are looking at the years corresponding to the Great Depression—but, given the high unemployment rate, the inflation rate is surprisingly high. In other words, given the very high unemployment rate, we would have expected not merely deflation, but a large rate of deflation. In fact, deflation was limited, and from 1934 to 1937, despite still very high unemployment, inflation actually turned positive.

How do we interpret this fact? There are two potential explanations.

One is that the Great Depression was associated with an increase not only in the actual unemployment rate, but also in the natural unemployment rate. This seems unlikely. Most economic historians see the Great Depression primarily as the result of a large adverse shift in aggregate demand leading to an increase in the actual unemployment rate over the natural rate of unemployment, rather than an increase in the natural rate of unemployment itself.

The other is that, when the economy starts experiencing deflation, the Phillips curve relation breaks down. One possible reason: The reluctance of workers to accept decreases in their nominal wages. Workers will unwittingly accept a cut in their real wages that occurs when their nominal wages increase more slowly than inflation. However, they are likely to fight the same cut in their real wages if it results from an overt cut in their nominal wages. If this argument is correct, this implies that the Phillips curve
relation between the change in inflation and unemployment may disappear, or at least become weaker, when the economy is close to zero inflation.

This issue is not just of historical interest: In many countries today, unemployment is very high, and inflation is low. Whether inflation will turn into deflation is one of the developments closely watched by macroeconomists today.

Summary

- The aggregate supply relation can be expressed as a relation between inflation, expected inflation, and unemployment. Given unemployment, higher expected inflation leads to higher inflation. Given expected inflation, higher unemployment leads to lower inflation.
- When inflation is not very persistent, expected inflation does not depend very much on past inflation. Thus, the aggregate supply relation becomes a relation between inflation and unemployment. This is what Phillips in the United Kingdom and Solow and Samuelson in the United States discovered when they looked, in the late 1950s, at the joint behavior of unemployment and inflation.
- As inflation became more persistent in the 1970s and 1980s, expectations of inflation became based more and more on past inflation. In the United States today, the aggregate supply relation takes the form of a relation between unemployment and the change in inflation. High unemployment leads to decreasing inflation; low unemployment leads to increasing inflation.
- The natural unemployment rate is the unemployment rate at which the inflation rate remains constant. When the actual unemployment rate exceeds the natural rate of unemployment, the inflation rate typically decreases; when the actual unemployment rate is less than the natural unemployment rate, the inflation rate typically increases.
- The natural rate of unemployment depends on many factors that differ across countries and can change over time. This is why the natural rate of unemployment varies across countries: It is higher in Europe than in the United States. Also, the natural unemployment rate varies over time: In Europe, the natural unemployment rate has greatly increased since the 1960s. In the United States, the natural unemployment rate increased from the 1960s to the 1980s and appears to have decreased since.
- Changes in the way the inflation rate varies over time affect the way wage setters form expectations and also affects how much they use wage indexation. When wage indexation is widespread, small changes in unemployment can lead to very large changes in inflation. At high rates of inflation, the relation between inflation and unemployment disappears altogether.
- At very low or negative rates of inflation, the Phillips curve relation appears to become weaker. During the Great Depression even very high unemployment led only to limited deflation. The issue is important because many countries have both high unemployment and low inflation today.

Key Terms

- Phillips curve, 161
- wage-price spiral, 164
- modified, or expectations-augmented, or accelerationist Phillips curve, 168
- non-accelerating inflation rate of unemployment (NAIRU), 170
- labor-market rigidities, 172
- unemployment insurance, 173
- employment protection, 173
- extension agreements, 173
- disinflation, 174
- hysteresis, 175
- Lucas critique, 176
- credibility, 176
- wage indexation, 177

Why deflation is potentially dangerous, and why policymakers want to avoid it, will have to wait until we introduce the distinction between nominal and real interest rates. We shall do this and return to this issue in Chapter 14.
Questions and Problems

QUICK CHECK

All Quick Check questions and problems are available on MyEconLab.

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. The original Phillips curve is the negative relation between unemployment and inflation that was first observed in the United Kingdom.
   b. The original Phillips curve relation has proven to be very stable across countries and over time.
   c. The aggregate supply relation is consistent with the Phillips curve as observed before the 1970s, but not since.
   d. Policy makers can exploit the inflation–unemployment trade-off only temporarily.
   e. In the late 1960s, the economists Milton Friedman and Edmund Phelps said that policy makers could achieve as low a rate of unemployment as they wanted.
   f. The expectations-augmented Phillips curve is consistent with workers and firms adapting their expectations after the macroeconomic experience of the 1960s.
   g. The natural rate of unemployment is constant over time within a country.
   h. The natural rate of unemployment is the same in all countries.
   i. Disinflation means that the rate of inflation is negative.
   j. If Lucas was right, and if monetary policy was fully credible, there would be no relation between inflation and unemployment (i.e., no Phillips curve relation).

2. Discuss the following statements.
   a. The Phillips curve implies that when unemployment is high, inflation is low, and vice versa. Therefore, we may experience either high inflation or high unemployment, but we will never experience both together.
   b. As long as we do not mind having high inflation, we can achieve as low a level of unemployment as we want. All we have to do is increase the demand for goods and services by using, for example, expansionary fiscal policy.

3. Mutations of the Phillips curve
   Suppose that the Phillips curve is given by
   \[ \pi_t = \pi_t^e + 0.1 - 2u_t \]
   a. What is the natural rate of unemployment?
   Assume \( \pi_t^e = \theta \pi_{t-1} \)
   and suppose that \( \theta \) is initially equal to 0. Suppose that the rate of unemployment is initially equal to the natural rate. In year 5, the authorities decide to bring the unemployment rate down to 3% and hold it there forever.
   b. Determine the rate of inflation in years 1, 2, 3, and 4.
   c. Do you believe the answer given in (b)? Why or why not?

4. The neutrality of money revisited
   a. Fill in the empty spaces after Year 1 in the chart below:

<table>
<thead>
<tr>
<th>Year</th>
<th>M (billions)</th>
<th>gM</th>
<th>P (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>380.95</td>
<td>5%</td>
<td>95.2</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>420</td>
<td>105.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>441</td>
<td>110.25</td>
<td></td>
</tr>
</tbody>
</table>

   b. What is the growth rate of the nominal money supply between years 1 and 2, 2 and 3, and 3 and 4?
   c. What is the rate of inflation between years 1 and 2, 2 and 3, and 3 and 4?
   d. What is the change in the real money supply between years 1 and 2, 2 and 3, and 3 and 4?
   e. What assumption has been made about real output growth if this data describe the medium run?

5. The effects of a permanent decrease in the rate of nominal money growth
   Suppose that the economy can be described by the following three equations:
   \[ u_t - u_{t-1} = -0.4(g_{yt} - 3\%) \quad \text{Okun’s law} \]
   \[ \pi_t - \pi_{t-1} = -(u_t - 5\%) \quad \text{Phillips curve} \]
   \[ g_{yt} = g_{mt} - \pi_t \quad \text{Aggregate demand} \]
   a. Reduce the three equations to two by substituting \( g_{yt} \) from the aggregate demand equation into Okun’s law. (Okun’s law was presented in Chapter 2.)
   Assume initially that \( u_t = u_{t-1} = 5\%, g_{mt} = 13\%, \) and \( \pi_t = 10\% \).
   b. Explain why these values are consistent with the statement “Inflation is always and everywhere a monetary phenomenon.”
   Now suppose that money growth is permanently reduced from 13% to 3%, starting in year 1.
   c. Compute (using a calculator or a spreadsheet program) unemployment and inflation in years 1, 2, 3, and 4.
   d. Does inflation decline smoothly from 10% to 3%? Why or why not?
e. Compute the values of the unemployment rate and the inflation rate in the medium run.

f. Is the statement that “Inflation is always and everywhere a monetary phenomenon” a statement that refers to the medium run or the short run?

**DIG DEEPER**

All Dig Deeper questions and problems are available on MyEconLab.

6. *The macroeconomic effects of the indexation of wages*

Suppose that the Phillips curve is given by

\[ \pi_t - \pi_{t-1} = 0.1 - 2u_t \]

where

\[ \pi_{t-1} = \pi_{t-1} \]

Suppose that inflation in year \( t - 1 \) is zero. In year \( t \), the authorities decide to keep the unemployment rate at 4% forever.

a. Compute the rate of inflation for years \( t, t + 1, t + 2, \) and \( t + 3 \). Now suppose that half the workers have indexed labor contracts.

b. What is the new equation for the Phillips curve?

c. Based on your answer to part (b), recompute your answer to part (a).

d. What is the effect of wage indexation on the relation between \( \pi \) and \( u \)?

7. *Supply shocks and wage flexibility*

Suppose that the Phillips curve is given by

\[ \pi_t - \pi_{t-1} = -\alpha(u_t - u_n) \]

where

\[ u_n = (m + z)/\alpha. \]

Recall that this Phillips curve was derived in this chapter under the assumption that the wage-bargaining equation took the form

\[ W = P^e(1 - \alpha u_t + z) \]

We can think of \( \alpha \) as a measure of wage flexibility—the higher \( \alpha \) the greater the response of the wage to a change in the unemployment rate, \( u_t \).

a. Suppose \( m = 0.03 \) and \( z = 0.03 \). What is the natural rate of unemployment if \( \alpha = 1 \)? if \( \alpha = 2 \)? What is the relation between \( \alpha \) and the natural rate of unemployment? Interpret your answer.

In Chapter 7, the text suggested that a reduction in the bargaining power of workers may have something to do with the economy’s relatively mild response to the increases in oil prices in the past few years as compared to the economy’s response to increases in oil prices in the 1970s. One manifestation of a reduction in worker bargaining power could be an overall increase in wage flexibility (i.e., an increase in \( \alpha \)).

b. Suppose that as a result of an oil price increase, \( m \) increases to 0.06. What is the new natural rate of unemployment if \( \alpha = 1 \) if \( \alpha = 2 \)? Would an increase in wage flexibility tend to weaken the adverse effect of an oil price increase?

**EXPLORE FURTHER**

8. *Estimating the natural rate of unemployment*

To answer this question, you will need data on the annual U.S. unemployment and inflation rates since 1970, which can be obtained very easily from the Economic Report of the President Web site [http://www.gpoaccess.gov/eop/index.html](http://www.gpoaccess.gov/eop/index.html) Excel tables of the values can be downloaded.

Retrieve the annual data for the civilian unemployment rate from Table B-35. In addition, retrieve the annual percentage increase for the consumer price index (CPI), all urban consumers from Table B-63. You can access the same data at the Federal Reserve Bank of St. Louis FRED Web site.

a. Plot the data for all the years since 1970 on a diagram, with the change in inflation on the vertical axis and the rate of unemployment on the horizontal axis. Is your graph similar to Figure 8-5?

b. Using a ruler, draw the line that appears to fit best the points in the figure. Approximately what is the slope of the line? What is the intercept? Write down your equation.

c. According to your analysis in (b), what has been the natural rate of unemployment since 1970?

9. *Changes in the natural rate of unemployment*

a. Repeat Problem 8 but now draw separate graphs for the period 1970 to 1990 and the period since 1990.

b. Do you find that the relation between inflation and unemployment is different in the two periods? If so, how has the natural rate of unemployment changed?

10. *Money growth and the growth in real output over time*

a. Fill in the empty spaces after Year 1 in the chart. This economy is in medium-run equilibrium in every year:

<table>
<thead>
<tr>
<th>Year</th>
<th>( M ) Nominal Money Supply (billions)</th>
<th>( g_M ) Growth Rate of Nominal Money Supply (percent)</th>
<th>( P )</th>
<th>Real GDP (billions of Year 2 dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>367</td>
<td>95.2</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>105.0</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>436</td>
<td>1622.4</td>
<td>1622.4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>475.3</td>
<td>1687.3</td>
<td>1687.3</td>
<td></td>
</tr>
</tbody>
</table>
b. What is the growth rate of the nominal money supply between years 1 and 2, years 2 and 3, and years 3 and 4?
c. What is the rate of inflation between years 1 and 2, 2 and 3, and 3 and 4?
d. What is the change in the real money supply between years 1 and 2, 2 and 3, and 3 and 4?
e. Why does the equation that says that the rate of inflation and the rate of growth of money must be equal NOT hold in this case? (Hint: Think about what must happen to real money growth if real output, and thus the demand for real money is increasing over time? If real money growth must be positive, what does this imply about the relation between nominal money growth and inflation?)

APPENDIX: From the Aggregate Supply Relation to a Relation between Inflation, Expected Inflation, and Unemployment

This appendix shows how to go from the relation between the price level, the expected price level, and the unemployment rate given by equation (8.1),

\[ P = P^e (1 + m)(1 - au + z) \]

to the relation between inflation, expected inflation, and the unemployment rate given by equation (8.2),

\[ \pi = \pi^e + (m + z) - au \]

First, introduce time subscripts for the price level, the expected price level, and the unemployment rate, so \( P_t, P^e_t \) and \( u_t \) refer to the price level, the expected price level, and the unemployment rate in year \( t \). Equation (8.1) becomes

\[ P_t = P^e_t (1 + m)(1 - au_t + z) \]

Next, go from an expression in terms of price levels to an expression in terms of inflation rates. Divide both sides by last year’s price level, \( P_{t-1} \):

\[ \frac{P_t}{P_{t-1}} = \frac{P^e_t}{P_{t-1}} (1 + m)(1 - au_t + z) \quad (8A.1) \]

Take the fraction \( P_t/P_{t-1} \) on the left side and rewrite it as

\[ \frac{P_t}{P_{t-1}} = \frac{P_t - P_{t-1} + P_{t-1}}{P_{t-1}} = 1 + \frac{P_t - P_{t-1}}{P_{t-1}} = 1 + \pi_t \]

where the first equality follows from actually subtracting and adding \( P_{t-1} \) in the numerator of the fraction, the second equality follows from the fact that \( P_{t-1}/P_{t-1} = 1 \), and the third follows from the definition of the inflation rate \( \pi_t = (P_t - P_{t-1})/P_{t-1} \).

Do the same for the fraction \( P^e_t/P_{t-1} \) on the right side, using the definition of the expected inflation rate \( \pi^e_t = (P^e_t - P_{t-1})/P_{t-1} \):

\[ \frac{P^e_t}{P_{t-1}} = \frac{P^e_t - P_{t-1} + P_{t-1}}{P_{t-1}} = 1 + \frac{P^e_t - P_{t-1}}{P_{t-1}} = 1 + \pi^e_t \]

Replacing \( P_t/P_{t-1} \) and \( P^e_t/P_{t-1} \) in equation (8A.1) by the expressions we have just derived,

\[ (1 + \pi_t) = (1 + \pi^e_t)(1 + m)(1 - au_t + z) \]

This gives us a relation between inflation, \( \pi_t \), expected inflation, \( \pi^e_t \), and the unemployment rate, \( u_t \). The remaining steps make the relation look more friendly:

Divide both sides by \( (1 + \pi^e_t)(1 + m) \):

\[ \frac{(1 + \pi_t)}{(1 + \pi^e_t)(1 + m)} = 1 - au_t + z \]

So long as inflation, expected inflation, and the markup are not too large, a good approximation to the left side of this equation is given by \( 1 + \pi_t - \pi^e_t - m \) (see Propositions 3 and 6 in Appendix 2 at the end of the book). Replacing in the equation above and rearranging gives

\[ \pi_t = \pi^e_t + (m + z) - au_t \]

Dropping the time indexes, this is equation (8.2) in the text. With the time indexes kept, this is equation (8.3) in the text.

The inflation rate \( \pi_t \) depends on the expected inflation rate \( \pi^e_t \) and the unemployment rate \( u_t \). The relation also depends on the markup, \( m \), on the factors that affect wage setting, \( z \), and on the effect of the unemployment rate on wages, \( \alpha \).
When, in late 2006, U.S. housing prices started to decline, most economists forecast that this would affect housing investment and consumption adversely, and thus slow down growth. A few forecast that it might lead to a mild recession. Very few anticipated that it might lead to the largest economic crisis since the Great Depression. But it did.

What happened, and what few had anticipated, is that the decline in housing prices triggered a major financial crisis. The financial system was much more fragile than had been perceived, and within a few months, many banks and other financial institutions found themselves either bankrupt or near bankruptcy. As a result, banks became unable or unwilling to lend. The interest rates at which consumers and firms could borrow increased dramatically, leading to a fall in spending, and a fall in output.

As the extent of the economic crisis became clear, policy makers responded with financial, monetary, and fiscal measures: Central banks decreased the interest rates under their control. Governments embarked on major fiscal expansions. It is likely that these policies avoided what would have been an even larger decline in output.

Over time, however, both monetary and fiscal policies have run into sharp limits. The interest rates directly controlled by central banks are close to zero and cannot decline further: Many economies are in a “liquidity trap.” The fiscal expansions, and the drop in government revenues from lower output, have led to large and worrisome increases in public debt. These limits make it harder to use policy to help the economy recover. While growth has turned positive since 2010, the recovery is slow, and unemployment is forecast to remain high for a long time.

We had a first look at the sequence of events in Chapter 1. Now that we have developed some of the basic tools, we can look at the events in more detail in this chapter. We focus on the United States in this chapter. Later on, when we have developed tools to look at the open economy, we shall look at the crisis in the rest of the world.

This chapter has three sections.

Section 9-1 looks at the start of the crisis, the decline in housing prices, and its effects on the financial system.

Section 9-2 examines the macroeconomic effects of the housing and financial crises, the evolution of output, and the policy responses.

Section 9-3 turns to the recovery.

In the process of analyzing the crisis, you will see how we use the IS–LM and the AS–AD models we developed in the previous chapters. We shall need to extend both, but you will see how we can build on them, and how they help organize both facts and thoughts.
9-1 From a Housing Problem to a Financial Crisis

When, in 2006, housing prices started declining in the United States, most economists forecast that this would lead to a decrease in aggregate demand and a slowdown in growth. Only a few economists anticipated that it would lead to a major macroeconomic crisis. What most had not anticipated was the effect of the decline of housing prices on the financial system. This is the focus of this section.

Housing Prices and Subprime Mortgages

Figure 9-1 shows the evolution of an index of U.S. housing prices from 2000 on. The index is known as the Case-Shiller index, named for the two economists who have constructed it. The index is normalized to equal 100 in January 2000. You can see the large increase in prices the early 2000s, followed by a large decrease since then. From a value of 100 in 2000, the index increased to 226 in mid 2006. Starting in 2006, however, the index first stabilized and declined slightly in 2006, then, from 2007, starting declining rapidly. By the end of 2008, at the start of the financial crisis, the index was down to 162. It continued to decline and, at the time of this writing, it is roughly stable, at around 150.

Was the sharp price increase from 2000 to 2006 justified? In retrospect, and given the ensuing collapse, surely not. But, at the time, when prices were increasing, economists were not so sure. Some increase in prices was clearly justified:

- The 2000s were a period of unusually low interest rates. As a result, mortgage rates were also low, increasing the demand for housing and thus pushing up the price.
- Other factors were also at work. Mortgage lenders became increasingly willing to make loans to more risky borrowers. These mortgages, known as subprime mortgages, or subprimes for short, had existed since the mid-1990s but became more prevalent in the 2000s. By 2006, about 20% of all U.S. mortgages were subprimes. Was it necessarily bad? Again, at the time, this was seen by most economists as a positive development: It allowed more people to buy homes, and, under the assumption that housing prices would continue to increase, so the value of the mortgage would decrease over time relative to the price of the house, it looked safe both for lenders and...
for borrowers. Judging from the past, the assumption that housing prices would not decrease also seemed reasonable: As you can see from Figure 9-1, housing prices had not decreased during the 2000–2001 recession.

In retrospect, again, these developments were much less benign than most economists thought. First, housing prices could go down, as became evident from 2006 on. When this happened, many borrowers found themselves in a situation where the mortgage they owed now exceeded the value of their house (when the value of the mortgage exceeds the value of the house, the mortgage is said to be **underwater**). Second, it became clear that, in many cases, the mortgages were in fact much riskier than either the lender pretended or the borrower understood. In many cases, borrowers had taken mortgages with low initial interest rates and thus low initial interest payments, probably not fully realizing that payments would increase sharply over time. Even if house prices had not declined, many of these borrowers would have been unable to meet their mortgage payments.

Thus, as house prices turned around and many borrowers defaulted, many banks found themselves faced with large losses. In mid-2008, losses on mortgages were estimated to be around 300 billion dollars. This is obviously a large number, but, relative to the size of the U.S. economy, it is not a very large one: 300 billion dollars is only about 2% of U.S. GDP. One might have thought that the U.S. financial system could absorb the shock and that the adverse effect on output would be limited.

This was not to be. While the trigger of the crisis was indeed the decline in housing prices, its effects were enormously amplified. Even those economists who had anticipated the housing price decline did not realize how strong the amplification mechanisms would be. To understand them, we must return to the role of banks.

### The Role of Banks

In Chapter 4, we looked at the role of banks in the determination of the money supply. Their important characteristic in that context was that they issued money, or, more precisely, that they had checkable deposits as liabilities. Here, we shall focus on their more general role as **financial intermediaries**, institutions that receive funds from those who wish to save and use those funds to make loans to those who wish to borrow.

Figure 9-2 shows a (much simplified) bank balance sheet. The bank has assets of 100, liabilities of 80, and capital of 20. You can think of the owners of the bank as having directly invested 20 of their own funds, borrowed 80, and bought various assets for 100. As we saw in Chapter 4, the liabilities may be checkable deposits, or borrowing from investors and other banks. The assets may be reserves (central bank money), loans to consumers, loans to firms, loans to other banks, mortgages, government bonds, or other forms of securities. In Chapter 4, we ignored capital. But, for our purposes, introducing capital is important here. Suppose that a bank did not hold any capital. Then, if, for any reason, the assets it held went down in value and the liabilities remained the same, liabilities would exceed assets, and the bank would be bankrupt. It is thus essential for the bank to hold enough capital to limit the risk of bankruptcy.

How can things go wrong even if the bank holds some capital, as in our example? First, the assets may decline in value by so much that the capital the bank holds is not

---

Some economists were worried even as prices were going up. Robert Shiller, one of the two economists behind the Case-Shiller index, was among them, warning that the price increase was a bubble that would most likely crash.

Some of these loans became known as NINJA loans (for no income, no job, no assets).

Some mortgages offered very low interest rates at the beginning. The low rates were known as “teaser rates.” The rates then increased sharply after a few months or a few years.

---

See Section 4-3.

One wishes that the balance sheets of banks were this simple and transparent. Had it been the case, the crisis would have been much more limited.

---

**Figure 9-2**

Bank Assets, Capital, and Liabilities
enough to cover its losses. In our example, this will happen if the value of the assets decreases below 80. The bank will become insolvent. This is not, however, the only way the bank can get in trouble. Suppose that some of the investors that have loaned to the bank (made a deposit in the bank) want their funds back right away. If the bank can sell some of its assets, it can get the funds and pay the depositors. But it may be difficult for the bank to sell the assets quickly: Calling back loans is difficult; some securities may be hard to sell. The problem of the bank in this case is not solvency, but illiquidity. The bank is still solvent, but it is illiquid. The more liquid its liabilities, or the less liquid its assets, the more likely the bank is to find itself in trouble.

What happened in this crisis is a combination of all these factors: Banks had too little capital. Liabilities, both deposits and other securities issued by banks, were very liquid. Assets were often very illiquid. The outcome was a combination of both solvency and liquidity problems, which quickly paralyzed the financial system. We now look at three specific aspects of the crisis that affected banks (and other financial intermediaries) in more detail.

**Leverage**

Consider two banks. As in Figure 9-2, bank A has assets of 100, liabilities of 80, and capital of 20. Its capital ratio is defined as the ratio of capital to assets and is thus equal to 20%. Its leverage ratio is defined as the ratio of assets to capital (the inverse of the capital ratio) and is thus equal to 5. Bank B has assets of 100, liabilities of 95, and capital of 5. Thus, its capital ratio is equal to 5%, and its leverage ratio to 20.

Now suppose that some of the assets in each of the two banks go bad. For example, some borrowers cannot repay their loans. Suppose, as a result, that for both banks, the value of the assets decreases from 100 to 90. Bank A now has assets of 90, liabilities of 80, and capital of 90 – 80 = 10. Bank B has assets of 90, liabilities of 95, and thus negative capital of 90 – 95 = −5. Its liabilities exceed its assets: In other words, it is bankrupt. This is indeed what happened during the crisis: Many banks had such a high leverage ratio that even limited losses on assets greatly increased the risk of bankruptcy.

Why was leverage so high? The example suggests a simple answer: Higher leverage means higher expected profit. Suppose, for example, that assets pay an expected rate of return of 5%, and liabilities pay an expected rate of return of 4%. Then the owners of bank A have an expected rate of return on their capital of \((100 \times 5\% - 80 \times 4\%)/20 = 9\%\), and the owners of bank B have an expected rate of return of \((100 \times 5\% - 95 \times 4\%)/5 = 24\%\), so more than twice as high. But, as the example we just saw also makes clear, leverage also increases risk: The higher the leverage, the more likely the bank is to go bankrupt. What happened throughout the 2000s is that banks decided to get a higher return and thus to take on more risk as well.

Why did banks opt to take on more risk? This is the subject of much discussion. There appears to be a number of reasons: First, banks probably underestimated the risk they were taking: Times were good, and, in good times, banks, just like people, tend to underestimate the risk of bad times. Second, the compensation and bonus system also gave incentives to managers to go for high expected returns without fully taking the risk of bankruptcy into account. Third, while financial regulation required banks to keep their capital ratio above some minimum, banks found new ways of avoiding the regulation, by creating new financial structures such as SIVs. What these are and how banks used them is explained in the Focus box “Increasing Leverage and Alphabet Soup”.

**Complexity**

Another important development of the 1990s and the 2000s was the growth of securitization. Traditionally, the financial intermediaries that made loans or issued mortgages
kept them on their own balance sheet. This had obvious drawbacks. A local bank, with local loans and mortgages on its books, was very much exposed to the local economic situation. When, for example, oil prices had come down sharply in the mid-1980s and Texas was in recession, many local banks went bankrupt. Had they had a more diversified portfolio of mortgages, say mortgages from many parts of the country, these banks might have avoided bankruptcy.

This is the idea behind securitization. Securitization is the creation of securities based on a bundle of assets (for example, a bundle of loans, or a bundle of mortgages). For instance, a mortgage-based security, or MBS for short, is a title to the returns from a bundle of mortgages, with the number of underlying mortgages often in the tens of thousands. The advantage is that many investors, who would not want to hold individual mortgages, will be willing to buy and hold these securities. This increase in the supply of funds from investors is, in turn, likely to decrease the cost of borrowing.

One can think of further forms of securitization. For example, instead of issuing identical claims to the returns on the underlying bundle of assets, one can issue different types of securities. For example, one can issue two types of securities: senior securities, which have first claims on the returns from the bundle, and junior securities, which come after and pay only if something is left after the senior securities have been paid. Senior securities will appeal to investors who want little risk; junior securities will appeal to investors who are willing to take more risk. Such securities, known as collateralized debt obligations, or CDOs, were first issued in the late 1980s but, again, grew in importance in the 1990s and 2000s. Securitization went even further, with the creation of CDOs using previously created CDOs, or CDO². This could go on and on!

Securitization would seem like a good idea, a way of diversifying risk and getting a larger group of investors involved in lending to households or firms. And, indeed, it is. But it also came with a large cost, which became clear only during the crisis. It was a risk that rating agencies, those firms that assess the risk of various securities, had largely missed: When underlying mortgages went bad, assessing the value of the underlying bundles in the MBSs, or, even more so, of the underlying MBSs in the CDOs, was extremely hard to do. These assets came to be known as toxic assets. It led investors to assume the worst and be very reluctant either to hold them or to continue lending to those institutions that did hold them.

**Liquidity**

Yet another development of the 1990s and 2000s was the development of other sources of finance than checkable deposits by banks (the 80 dollars they borrowed in our example above). Increasingly, they relied on borrowing from other banks or other investors, in the form of short-term debt, to finance the purchase of their assets, a process known as wholesale funding. SIVs, the financial entities set up by banks, which we saw earlier, were entirely funded through such wholesale funding.

Wholesale funding again would seem like a good idea, giving banks more flexibility in the amount of funds they can use to make loans or buy assets. But it has a cost, which again became clear during the crisis. If investors or other banks, worried about the value of the assets held by the bank, decide to stop lending to the bank, the bank may find itself short of funds and be forced to sell some of its assets. If these assets are complex and hard to sell, it may have to sell them at very low prices, often referred to as fire sale prices.

We now have all the elements we need to explain what happened when housing prices declined, and why this led to a major financial crisis.
The Medium Run

The Core

Amplification Mechanisms

As the crisis worsened, solvency and liquidity concerns increased sharply, each reinforcing the other.

- When housing prices declined, and some mortgages went bad, high leverage implied a sharp decline in the capital of banks. This in turn forced them to sell some of their assets. Because these assets were often hard to value, they had to sell them at fire sale prices. This, in turn, decreased the value of similar assets remaining on their balance sheet, or on the balance sheet of other banks, leading to a further decline in capital ratio and forcing further sales of assets and further declines in prices.

- The complexity of the securities (MBSs, CDOs) and of the true balance sheets of banks (banks and their SIVs) made it very difficult to assess the solvency of banks and their risk of bankruptcy. Thus, investors became very reluctant to continue to lend to them, and wholesale funding came to a stop, forcing further asset sales.

Increasing Leverage and Alphabet Soup: SIVs, AIG, and CDSs

SIV stands for structured investment vehicle. Think of it as a virtual bank, created by an actual bank. On the liability side, it borrows from investors, typically in the form of short-term debt. On the asset side, it holds various forms of securities. To reassure the investors that they will get repaid, the SIV typically had a guarantee from the actual bank that, if needed, the bank will provide funds to the SIV.

While the first SIV was set up by Citigroup in 1988, SIVs rapidly grew in size in the 2000s. You may ask why banks did not simply do all these things on their own balance sheet rather than create a separate vehicle. The main reason was to be able to increase leverage. If the banks had done these operations themselves, the operations would have appeared on their balance sheet and been subject to regulatory capital requirements, forcing them to hold enough capital to limit the risk of bankruptcy. But, it turns out, doing these operations through an SIV did not require banks to put capital down. For that reason, through setting up an SIV, banks could increase leverage and increase expected profits, and they did.

When housing prices started declining, and many mortgages turned bad, the securities held by the SIVs decreased in value. Investors became reluctant to lend to the SIVs, out of fear that they may be insolvent. The banks that had created the SIVs had to honor their obligations by paying investors, but had limited capital to do so. It became clear that banks had in effect created a shadow banking system, and that leverage of the banking system as a whole (i.e., including the shadow banking part) was much higher than had been perceived. Small losses could lead to bankruptcies. As of October 2008, no SIVs were left; they had either closed, or all their assets and liabilities had been transferred to the banks that had created them.

AIG stands for American International Group. It is an insurance company that, in the 2000s, had what looked like a good idea at the time. It would sell not only regular insurance, but also insurance against default risk, through the sale of credit default swaps, or CDSs for short. If a bank was worried about default on a security it held in its portfolio, it could buy a CDS from AIG that promised to pay the bank in case of default on the security. For this, AIG charged the bank a price supposed to reflect the probability of such a default. For banks, it was an attractive deal, because by buying insurance, the securities they held became riskless and thus decreased the capital that banks had to hold (the less risky the asset, the smaller the amount of capital required by regulation). AIG, being an insurance company rather than a bank, did not have to hold capital against the promises it was making.

When housing prices started declining and mortgages began to default, AIG had to make good on many of its promises. AIG, however, did not have the funds to make the payments on the CDSs they had issued. Thus, suddenly, banks realized that, without the insurance payout, their assets were much riskier than they had assumed, and that they did not have the capital needed to sustain losses. Again, leverage of the financial system (including now the banks, the SIVs, and CDS issuers such as AIG) was much higher than had been perceived. As we shall see below, the U.S. government decided it had to provide funds to AIG to make payments on the CDSs. The alternative would have led to default of AIG, but also the potential default of many banks holding CDSs. As of the end of 2009, the government had advanced more than $180 billion to AIG, which AIG used to pay the banks as promised. Since then, AIG has been steadily reimbursing the U.S. government and is expected to fully repay the loan.
and price declines. Even banks became very reluctant to lend to each other. This is shown in Figure 9-3, which shows the difference between the riskless rate (measured by the rate of three-month government bonds), which you can think of as the rate determined by monetary policy, and the rate at which banks are willing to lend to each other (known as the **Libor rate**). This difference is known as the **Ted spread**.

If banks perceived no risk in lending to each other, the Ted spread would be equal to zero. And, indeed, until mid-2007, it was very close to zero. Note, however, how it became larger in the second half of 2007 and then increased sharply in September 2008. Why then? Because, on September 15, 2008, Lehman Brothers, a major bank with more than $600 billion in assets, declared bankruptcy, leading financial participants to conclude that many, if not most, other banks and financial institutions were indeed at risk.

By mid-September 2008, both mechanisms were in full force. The financial system had become paralyzed: Banks had basically stopped lending to each other or to anyone else. Quickly, what had been largely a financial crisis turned into a macroeconomic crisis.

### 9-2 The Use and Limits of Policy

The immediate effects of the financial crisis on the macro economy were two-fold: first, a large increase in the interest rates at which people and firms could borrow; second, a dramatic decrease in confidence.

- Figure 9-4 shows the effect of the financial crisis on different interest rates. The first interest rate is the rate on 10-year U.S. government bonds. The second and third are the rates charged by the bond markets to two different types of firms, corresponding to different risk ratings. Firms with a **AAA** (triple A) rating are considered the safest, firms with a **BBB** (triple B) are considered less safe. In normal times, AAA firms can borrow at a rate close to the rate on government bonds; BBB firms borrow at a higher rate, but the difference is typically small, on the order of 1%. You can see that this was indeed the case at the start of 2007. But, as you can also see, the difference increased from mid-2007 on, and, while the rate on government bonds remained very low, the rates on both AAA and BBB bonds jumped
The Medium Run

In September 2008, the financial crisis led to a sharp increase in the rates at which firms could borrow. Suddenly, borrowing became extremely expensive for most firms. And for the many firms too small to issue bonds and thus depending on bank credit, it became nearly impossible to borrow at all.

In short, the interest rate charged to borrowers became very high (in some cases borrowers were completely shut out from borrowing) relative to the interest rate controlled by monetary policy.

Figure 9-5 shows the effects of the financial crisis on expectations. The events of September 2008 triggered wide anxiety among consumers and firms. Thoughts of another Great Depression and, more generally, confusion and fear about what was happening in the financial system, led to a large drop in confidence. The evolution of consumer confidence and business confidence indexes for the United States are shown in Figure 9-5. Both indexes are normalized to equal 100 in January 2007. Note how consumer confidence, which had started declining in mid-2007, took a sharp drop in the fall of 2008 and reached a low of 22 in early 2009, a level far below previous historical lows. The result of lower confidence and lower housing and stock prices was a sharp decrease in consumption.

See the Focus box “The Lehman Bankruptcy, Fears of Another Great Depression, and Shifts in the Consumption Function” in Chapter 3.

Figure 9-5

U.S. Consumer and Business Confidence, since 2007

The financial crisis led to a sharp drop in confidence, which bottomed in early 2009.

Source: Bloomberg L.P.
Initial Policy Responses

The high cost of borrowing, lower stock prices, and lower confidence all combined to decrease the demand for goods. In terms of the IS–LM model, there was a sharp adverse shift of the IS curve. In the face of this large decrease in demand, policy makers did not remain passive.

The most urgent measures were aimed at strengthening the financial system:

- In order to prevent a run by depositors, federal deposit insurance was increased from $100,000 to $250,000 per account. Recall, however, that much of banks’ funding came not from deposits but from the issuance of short-term debt to investors. In order to allow the banks to continue to fund themselves through wholesale funding, the Federal government offered a program guaranteeing new debt issues by banks.

  - The Federal Reserve provided widespread liquidity to the financial system. We have seen that, if investors wanted to take their funds back, the banks had no alternative than to sell some of their assets, often at fire sale prices. In many cases, this would have meant bankruptcy. To avoid this, the Fed put in place a number of liquidity facilities to make it easier to borrow from the Fed. It allowed not only banks, but also other financial institutions to borrow from the Fed. Finally, it increased the set of assets that financial institutions could use as collateral when borrowing from the Fed (collateral refers to the asset a borrower pledges when borrowing from a lender. If the borrower defaults, the asset then goes to the lender). Together, these facilities allowed banks and financial institutions to pay back investors without having to sell their assets. It also decreased the incentives of investors to ask for their funds, as these facilities decreased the risk that banks would go bankrupt.

  - The government introduced a program, called the Troubled Asset Relief Program, or TARP, aimed at cleaning up banks. The initial goal of the $700 billion program, introduced in October 2008, was to remove the complex assets from the balance sheet of banks, thus decreasing uncertainty, reassuring investors, and making it easier to assess the health of each bank. The Treasury, however, faced the same problems as private investors. If these complex assets were going to be exchanged for, say, Treasury bills, at what price should the exchange be done? Within a few weeks, it became clear that the task of assessing the value of each of these assets was extremely hard and would take a long time, and the initial goal was abandoned. The new goal became to increase the capital of banks. This was done by the government acquiring shares and thus providing funds to most of the largest U.S. banks. By increasing their capital ratio, and thus decreasing leverage, the goal of the program was to allow the banks to avoid bankruptcy and, over time, return to normal. As of the end of September 2009, total spending under the TARP was $360 billion, of which $200 billion was spent through the purchase of shares in banks. At the time of writing, most banks have bought back their shares and have reimbursed the government. The final cost of TARP is expected to be small, perhaps even zero.

  - All these measures were aimed at providing liquidity to financial institutions, avoiding unnecessary bankruptcies, and allowing the financial system to function again. Worried, however, that some markets were slow to recover, the Fed directly intervened by purchasing private securities in these markets. In particular, given the importance of the housing sector in the crisis, it bought mortgage-backed securities. At the time of writing, the Fed is still the main buyer of these securities. Fiscal and monetary policies were used aggressively as well.

- Figure 9-6 shows the evolution of the T-bill rate from January 2006 on. Starting in the summer of 2007, the Fed began to worry about a slowdown in growth and

The specific interest rate used in the figure is the rate on T-bills with a maturity of three months, called the three-month T-bill rate. The interest rate is expressed as an annual rate.
The Medium Run

The Core

For technical reasons, the T-bill rate has not been quite equal to zero, but has typically been slightly positive. For all practical purposes, this has the same effect as a zero rate. By September 2008, the rate stood at 1.7%, down from about 5% in July 2007. And, when it became clear in the fall of 2008 that demand was falling quickly, the Fed decreased the rate further. By December 2008, the rate was down to zero, and still is equal to zero at the time of this writing.

When the size of the adverse shock became clear, the U.S. government turned to fiscal policy, using a combination of reductions in taxes and increases in spending. When the Obama administration assumed office in 2009, its first priority was to design a fiscal program that would increase demand and reduce the size of the recession. Such a fiscal program, called the American Recovery and Reinvestment Act, was passed in February 2009. It called for $780 billion in new measures, in the form of both tax reductions and spending increases, over 2009 and 2010. The U.S. budget deficit increased from 1.7% of GDP in 2007 to a very high 9.0% in 2010. The increase was largely the mechanical effect of the crisis, as the decrease in output led automatically to a decrease in tax revenues and to an increase in transfer programs such as unemployment benefits. But it was also the result of the specific measures in the fiscal program aimed at increasing either private or public spending.

Still, this combination of financial, fiscal, and monetary measures was not enough to avoid a large decrease in output, with U.S. GDP falling by 3.5% in 2009 and recovering only slowly thereafter. One would hope that fiscal and monetary policies could help strengthen the recovery. But, as we shall see now, both face sharp limits.

The Limits of Monetary Policy: The Liquidity Trap

Since December 2008, the Fed has kept the T-bill rate at zero. Could it do more? More generally, what happens if the interest rate is equal to zero and the central bank further increases the supply of money?

To answer this question, we must first go back first to our characterization of the demand and the supply of money in Chapter 4. There we drew the demand for money, for a given level of income, as a decreasing function of the interest rate. The lower the interest rate, the larger the demand for money—equivalently, the smaller the demand for bonds. What we did not ask in Chapter 4 is what happens to the demand for money
when the interest rate becomes equal to zero. The answer: Once people hold enough money for transaction purposes, they are then indifferent between holding the rest of their financial wealth in the form of money or in the form of bonds. The reason they are indifferent is that both money and bonds pay the same interest rate, namely zero. Thus, the demand for money is as shown in Figure 9-7:

- As the interest rate decreases, people want to hold more money (and thus less bonds): The demand for money increases.
- As the interest rate becomes equal to zero, people want to hold an amount of money at least equal to the distance $OB$: this is what they need for transaction purposes. But they are willing to hold even more money (and therefore hold less bonds) because they are indifferent between money and bonds. Therefore, the demand for money becomes horizontal beyond point $B$.

Now consider the effects of an increase in the money supply.

- Consider the case where the money supply is $M'$, so the interest rate consistent with financial market equilibrium is positive and equal to $i$. (This is the case we considered in Chapter 4.) Starting from that equilibrium, an increase in the money supply—a shift of the $M'$ line to the right—leads to a decrease in the interest rate.
- Now consider the case where the money supply is $M''$, so the equilibrium is at point $B$; or the case where the money supply is $M'''$, so the equilibrium is given at point $C$. In either case, the initial interest rate is zero. And, in either case, an increase in the money supply has no effect on the interest rate. Think of it this way:
  Suppose the central bank increases the money supply. It does so through an open market operation in which it buys bonds and pays for them by creating money. As the interest rate is zero, people are indifferent to how much money or bonds they hold, so they are willing to hold less bonds and more money at the same interest rate, namely zero. The money supply increases, but with no effect on the interest rate—which remains equal to zero.

If you look at Figure 4-1, you will see that we avoided the issue by not drawing the demand for money for interest rates close to zero.
In short: Once the interest rate is equal to zero, expansionary monetary policy becomes powerless. Or to use the words of Keynes, who was the first to point out the problem, the increase in money falls into a **liquidity trap**: People are willing to hold more money (*more liquidity*) at the same interest rate.

The derivation of the *LM* curve when one takes into account the possibility of a liquidity trap is shown in the two panels of Figure 9-8. Recall that the *LM* curve gives, for a given real money stock, the relation between the interest rate and the level of income implied by equilibrium in financial markets. To derive the *LM* curve, Figure 9-8(a) looks at equilibrium in the financial markets for a given value of the real money stock and draws three money demand curves, each corresponding to a different level of income:

- $M^d$ shows the demand for money for a given level of income $Y$. The equilibrium is given by point $A$, with interest rate equal to $i$. This combination of income $Y$ and interest rate $i$ gives us the first point on the *LM* curve, point $A$ in Figure 9-8(b).
- $M^d'$ shows the demand for money for a lower level of income, $Y' < Y$. Lower income means fewer transactions and, therefore, a lower demand for money at any interest rate. In this case, the equilibrium is given by point $A'$, with interest rate equal to $i'$. This combination of income $Y'$ and interest rate $i'$ gives us the second point on the *LM* curve, point $A'$ in Figure 9-8(b).
- $M^d''$ gives the demand for money for a still lower level of income $Y'' < Y'$. In this case, the equilibrium is given by point $A''$ in Figure 9-8(a), with interest rate equal to zero. Point $A''$ in Figure 9-8(b) corresponds to $A''$ in Figure 9-9(a).

What happens if income decreases below $Y''$, shifting the demand for money further to the left in Figure 9-8(a)? The intersection between the money supply curve and the money demand curve takes place on the horizontal portion of the money demand curve. The interest rate remains equal to zero.

Let’s summarize: In the presence of a liquidity trap, the *LM* curve is given by Figure 9-8(b). For values of income greater than $Y''$, it is upward sloping—just as it was in Chapter 5 when we first characterized the *LM* curve. For values of income less than $Y''$, it is flat at $i = 0$. Intuitively: The interest rate cannot go below zero.

**Figure 9-8**

*The Derivation of the LM Curve in the Presence of a Liquidity Trap*

For low levels of output, the *LM* curve is a flat segment, with an interest rate equal to zero. For higher levels of output, it is upward sloping: An increase in income leads to an increase in the interest rate.
Having derived the $LM$ curve in the presence of a liquidity trap, we can look at the properties of the $IS$–$LM$ model modified in this way. Suppose the economy is initially at point $A$ in Figure 9-9. Equilibrium is at point $A$, at the intersection of the $IS$ curve and the $LM$ curve, with output $Y$ and interest rate $i$. And suppose that this level of output is very low. The question is: Can monetary policy help the economy return to a higher level of output, say to $Y_n$?

Suppose the central bank increases the money supply, shifting the $LM$ curve from $LM$ to $LM'$. The equilibrium moves from point $A$ down to point $B$. The interest rate decreases from $i$ to zero, and output increases from $Y$ to $Y'$. Thus, to this extent, expansionary monetary policy can indeed increase output.

What happens, however, if starting from point $B$, the central bank increases the money supply further, shifting the $LM$ curve from $LM'$ to, say, $LM''$? The intersection of $IS$ and $LM''$ remains at point $B$, and output remains equal to $Y'$. Expansionary monetary policy no longer has an effect on output; it cannot therefore help output increase to $Y_n$.

In words: When the interest rate is equal to zero, the economy falls into a liquidity trap: The central bank can increase liquidity—that is, increase the money supply. But this liquidity falls into a trap: The additional money is willingly held by people at an unchanged interest rate, namely zero. If, at this zero interest rate, the demand for goods is still too low, then there is nothing further conventional monetary policy can do to increase output.

You will note that, in the previous paragraph, we referred to the limits of conventional monetary policy; that is, monetary policy using open market operations aimed at decreasing the interest rate typically controlled by the Fed—in the United States, policy aimed at decreasing the interest rate on T-bills. The question is whether some unconventional measures may still be used. This is what the Fed (and other central banks around the world) have explored since 2008.

In the simple $IS$–$LM$ model presented in Chapter 5, there was only one type of bond and one interest rate, and thus, once this rate was down to zero, there was nothing more monetary policy could do. But, in reality, there are many types of bonds and many interest rates. Some of these interest rates are higher than the interest rate on T-bills. This suggests the following unconventional monetary policy: Rather than

Figure 9-9

The IS–LM Model and the Liquidity Trap

In the presence of a liquidity trap, there is a limit to how much monetary policy can increase output.
buying Treasury bills through open market operations, the Fed could buy other bonds; for example, mortgages—loans made by banks to households, or Treasury bonds—government bonds which promise payment over, say, 10 or 20 years. By doing so, it may be able to decrease the interest rate on those bonds or on those mortgages. These lower interest rates can help increase demand.

Such a policy goes under the name of credit easing or quantitative easing, and this is indeed what the Fed has done at various times during this crisis. How helpful is quantitative easing? We shall look at the evidence in Chapter 17 and again in Chapter 24. But the conclusion can be stated simply. These unconventional measures have some effect, but the effect is often small. When the economy is in the liquidity trap, the scope for monetary policy to affect demand and output is sharply limited.

The Limits of Fiscal Policy: High Debt

A recurrent theme of this book is that both monetary policy and fiscal policy can be used to affect demand and, in turn, output. So, even if monetary policy has reached sharp limits, isn’t fiscal policy the solution? The answer is that fiscal policy also has limits. The problem is that, if the demand for goods does not recover over time by itself, if people or firms do not eventually become more optimistic and increase spending, the government must continue to run deficits to sustain higher demand and output. Continuing large deficits lead, however, to steadily higher public debt. In advanced countries, the ratio of government debt to GDP has increased from 46% in 2006 to 70% in 2011; in the United States, the ratio has increased from 42% in 2006 to 72% in 2011. High debt implies that, sooner or later, either taxes will have to increase, or spending will have to decrease, or the government will be unable to repay the debt. And when investors become worried about repayment of the debt, they start asking for higher interest rates on government bonds, making it even harder for the government to repay the debt. These worries are already leading to higher interest rates on government bonds in a number of European countries. They have not yet led to higher interest rates on government bonds in the United States. But the risk that interest rates might rise in the future is forcing the U.S. government to look for ways to begin to reduce its budget deficit now. This limits the contribution of fiscal policy to demand and to the recovery.

9-3 The Slow Recovery

While output growth is now positive in the United States, the recovery is very slow. Under current forecasts, unemployment is predicted to remain high for many years. There are increasing worries of a “lost decade.” Looking at what has happened in Japan since the 1990s, these worries are justified: For nearly two decades, Japan has been in an economic slump. As the Focus box “Japan, the Liquidity Trap, and Fiscal Policy” shows, zero interest rates and large budget deficits have not succeeded in getting the Japanese economy back to normal.

Why has the recovery from the crisis so slow in the United States? Some economists point to the aggregate supply side. They argue that the banking crisis has decreased the natural level of output, so that it would be wrong to think that we can go back to the pre-crisis level of output. More accurately, taking into account that output typically grows over time, it would be wrong to think that output can return to its old pre-crisis trend line. The weak recovery that we observe may be the best the economy can deliver. Indeed, the evidence from a large number of past banking crises, summarized in the Focus box “Do Banking Crises Affect Output in the
Japan, the Liquidity Trap, and Fiscal Policy

In the early 1990s, the Japanese stock market, which had boomed earlier, suddenly crashed. The Nikkei index, a broad index of Japanese stock prices, had gone up from 7,000 in 1980 to 35,000 at the beginning of 1990. Then, within two years, it went down to 16,000 and continued to decline after that, reaching a trough of 7,000 in 2003 (as we write, the Nikkei index is around 9,000). This decline in stock prices was followed by a decline in spending, and, in response to the decline in spending, the Japanese central bank cut the interest rate. As you can see from Figure 1, by the mid-1990s, the interest rate was down to less than 1%, and it has remained below 1% since.

With little room left for using monetary policy, fiscal policy was used to sustain demand. Figure 2 shows the evolution of government spending and revenues as a percentage of GDP since 1990. You can see the dramatic increase in spending from the early 1990s on. Much of the increased spending has taken the form of public works projects, and a joke circulating in Japan is that, by the time the Japanese economy has recovered, the entire shoreline of the Japanese archipelago will be covered in concrete. The result of this strong fiscal expansion, however, has been a sharp increase in debt. The ratio of government debt to GDP, which stood at 13% of GDP in 1991, is now above 120%. Meanwhile, the Japanese economy is still in a slump: GDP growth, which averaged 4.4% in the 1980s, was down to 1.4% in the 1990s, and 0.9% in the 2000s. What has happened in Japan since 1990 is a tough warning to other advanced countries that it may take a long time to recover.

Figure 1  The Interest Rate in Japan since 1990. Japan has been in a liquidity trap since the mid-1990s.
Source: One-year government bond rate, DLX, International Monetary Fund database

Figure 2  Government Spending and Revenues (as a percentage of GDP), Japan, since 1990. Increasing government spending and decreasing revenues have led to steadily larger deficits.
Source: IMF World Economic Outlook database
Medium Run?” suggests that indeed output remains below its old pre-crisis trend line for many years. One can think of a number of reasons why this may be the case. The banking crisis may affect the efficiency of the banking system for a long time, leading to lower productivity (again relative to trend): Some of the new regulations introduced to decrease the risk of another financial crisis, such as increases in the capital ratio that banks must maintain, may indeed decrease risk; but they may also make intermediation between borrowers and lenders more costly, thus decreasing the natural level of output.

It may indeed be that the economy cannot return to its pre-crisis trend line. But in the context of the United States, this does not appear sufficient to explain the slow recovery from the crisis. In 2011, unemployment was around 9%. Pre-crisis, most estimates of the natural rate of unemployment were about 6%. It is unlikely that such a large increase in unemployment is entirely due to an increase in the natural rate of unemployment. In other words, what we are observing seems to be a rate of unemployment far above the underlying natural rate and, by implication, a level of output far below its natural level. So, most economists point also to the aggregate demand side. For the time being, insufficient aggregate demand, they argue, is the issue.

They point first to the limits of policy we have examined earlier. In a typical recovery, monetary and fiscal policy can be used to hasten the return of output to its natural level. In the current crisis, they can play a limited role at best. There is no room left for conventional monetary policy, and the effects of unconventional monetary policy are limited and uncertain. Worries about debt are putting strong pressure on the government to reduce the deficit, to pursue fiscal consolidation rather than fiscal expansion.

They also point out that, in the presence of the liquidity trap, not only does conventional monetary policy not work, but the process of adjustment that typically takes output back to its natural level in the medium run also fails. Recall from Chapter 7 how the mechanism typically works:

A decrease in output below its natural level leads to a decrease in the price level (at least relative to its trend). This leads to an increase in the real money stock, which in turn leads to a decrease in the interest rate. The decrease in the interest rate leads then to an increase in spending, which in turn leads to an increase in output. The process goes on until output has returned to its natural level. The process can be made faster by using either monetary policy (that is, by increasing the money stock, which leads to a larger decrease in the interest rate) or fiscal policy, which increases demand directly. At the core of the adjustment is the aggregate demand relation (equation (8.3) in Chapter 8):

\[ Y = Y\left(\frac{M}{P}, G, T\right) \]

Now think about what happens when the economy is in the liquidity trap, with the interest rate equal to zero. In this case, an increase in the real money stock, \( M/P \), whether it comes from an increase in \( M \) or from a decrease in \( P \), has no effect on the interest rate, which remains equal to zero. So not only does monetary policy not affect spending, but the adjustment mechanism that returns output to its natural level in the AS–AD model also does not work: The decrease in the price level leads to a higher real money stock but does not lead to a lower interest rate and does not lead to higher spending.

Let’s formally introduce this in our AS–AD model. If the economy is in the liquidity trap, the aggregate demand relation takes the following form:

\[ Y = Y(G, T) \]
As before, increases in government spending or decreases in taxes increase demand. But in the liquidity trap, aggregate demand no longer depends on the real money stock.

What may then happen to the economy is represented in Figure 9-10, using the AS–AD model. Aggregate supply is still represented by an upward sloping curve in the figure: The higher the level of output, the higher the price level, given the expected price level. Conversely, and more relevant for our case, the lower the output, the lower the price level. The aggregate demand relation is now vertical. For given values of $G$, $T$, aggregate demand does not depend on the real money stock and thus does not depend on the price level. Suppose that the initial aggregate supply and demand curves are given by $AS$ and $AD$, respectively, so the initial equilibrium is at point $A$, with output $Y$ below the natural level $Y_n$. In other words, output is low, and the economy is in the liquidity trap. As output is below its natural level, the aggregate supply curve shifts down over time. (Recall the mechanism: Low output implies high unemployment, which puts downward pressure on wages, and in turn on prices.) The equilibrium moves over time from $A$ to $B$ to $C$: The price level keeps decreasing, but this does not lead to an increase in output.

So is there hope that the U.S. economy will eventually return to normal? Yes. There are a number of reasons to think that aggregate demand will eventually recover. Eventually, the damage done to the banking system should be repaired. Very low housing investment and thus a decreasing housing stock, together with a growing population, should eventually lead to an increase in prices and higher housing investment in the future. Also, some types of consumption and investment cannot be deferred forever. Low purchases of consumer durables and of equipment now imply higher purchases later: Eventually, cars and machines break down and must be replaced. Economists sometimes refer to this mechanism as pent-up demand: Demand that does not take place today is pent up and increases demand in the future. Still, this may all take time, and, at the time of writing, a strong recovery appears to be far in the future.
Do Banking Crises Affect the Natural Level of Output?

Leaving aside the current crisis, there is a lot of evidence that banking crises lead to large decreases in output in the short run. But do they have an effect on output in the medium run? Or, put in terms of our model, do they affect the natural level of output?

To answer this question, researchers at the IMF looked at a number of banking crises across many countries from 1970 to 2002. They defined banking crises as episodes where there were either bank runs or a large number of bank failures. They identified 88 such crises. In each case, they looked at the behavior of GDP in the years following each crisis.

Using econometrics, they reached two conclusions: First, financial crises typically lead to a decrease in output relative to trend, even in the medium run. Second, while this conclusion holds on average, there is a lot of variation across countries. Some countries go back to trend, while others suffer large decreases.

The flavor of their results is given in Figure 1, which shows what happened in four countries following a banking crisis: Mexico after 1994, Korea after 1997, Sweden after 1991, and Thailand after 1997. In all four cases, there were major bank failures. For each country, the figure shows the evolution of GDP (the blue line) relative to the pre-crisis trend (dashed red line). You can see that, in three of the cases—Korea, Sweden, and Thailand—there was a large decrease in output relative to the pre-crisis trend, and this decrease was still largely present five years after the crisis. In other words, five years after the crisis, the rate of growth of GDP was roughly the same as before the crisis, but the level of GDP was lower than it would have been absent the crisis.

Can we tell why banking crises affect output, even in the medium run? The same researchers also looked at what happened to employment and what happened to productivity as a result of the crisis. They concluded that, on average, the decline in output could be broken down as follows: one-third related to a decrease in employment; two-thirds related to a decrease in productivity (both relative to trend). This suggests that the banking system plays an important role in the economy. Banking crises weaken the ability of the banking system to allocate funds to the right borrowers. This, in turn, makes the economy less productive.

Source: International Monetary Fund World Economic Outlook, October 2009, Chapter 4.

Figure 1  The Evolution of Output after Four Banking Crises
The trigger of the crisis was a decrease in housing prices. The effect of lower housing prices was considerably amplified by the effects on the banking system. Because they had very low capital ratios, some banks became insolvent. Because the assets they held were highly complex, their value in the face of a decrease in housing prices and defaults on mortgages was highly uncertain, investors became reluctant to lend to banks, and many banks became illiquid. Banks became unwilling to lend to each other or to anyone else.

Much higher interest rates for borrowers and, in some cases, the inability to borrow at all, led to a large decrease in spending. Worries about another Great Depression led to sharp declines in confidence and a further decrease in spending. The financial crisis led to a macroeconomic crisis and a large decline in output.

Policies—fiscal, monetary, and financial—were used. They probably prevented an even larger decline in output but did not prevent the recession. Both fiscal and monetary policies now face sharp limits. Conventional monetary policy no longer works. The interest rate on T-bills has been decreased to zero, and the U.S. economy is in a liquidity trap. Large budget deficits have led to a large increase in debt, and there is strong pressure on the U.S. government to start reducing deficits now.

The recovery is slow, and unemployment is expected to remain high for some time. It may be that the financial crisis has done lasting damage to the banking system, and the natural level of output may have decreased relative to trend. At this stage, however, the problem is on the demand side. The limits of policy, and the failure of the standard adjustment mechanism to return the economy to its natural level, imply that demand is likely to remain weak, and the recovery is likely to remain slow for some time to come.

Key Terms

- mortgage lenders, 184
- subprime mortgages (subprimes), 184
- underwater mortgage, 185
- financial intermediaries, 185
- solvency, 186
- illiquidity, 186
- capital ratio, 186
- leverage ratio, 186
- securitization, 186
- mortgage-based security (MBS), 187
- senior securities, 187
- junior securities, 187
- collateralized debt obligation (CDO), 187
- rating agencies, 187
- toxic assets, 187
- wholesale funding, 187
- fire sale prices, 187
- structured investment vehicle (SIV), 188
- shadow banking system, 188
- American International Group (AIG), 188
- credit default swap (CDS), 188
- Libor rate, 189
- TED spread, 189
- AAA (triple A), BBB (triple B), 189
- liquidity facilities, 191
- collateral, 191
- Troubled Asset Relief Program (TARP), 191
- American Recovery and Reinvestment Act, 192
- liquidity trap, 194
- credit easing, 196
- quantitative easing, 196
- Nikkei index, 197
- pent-up demand, 199

Questions and Problems

QUICK CHECK

All Quick Check questions and problems are available on MyEconLab.

1. Using information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. The loss in output that resulted from the financial crisis is many times larger than the losses on mortgages held by U.S. financial institutions.
   b. An increase in a bank’s leverage ratio tends to increase both the expected profit of the bank and the risk of the bank going bankrupt.
   c. The high degree of securitization in the U.S. financial system helped to diversify risk and probably lessened the economic effect of the fall in housing prices.
   d. Since the financial crisis ultimately led to a global recession, the policy measures (adopted in many countries)
that provided substantial liquidity to financial institutions and that recapitalized banks (through the purchase of shares by governments) failed.

e. The fiscal stimulus programs adopted by many countries in response to the financial crisis helped offset the decline in aggregate demand and reduce the size of the recession.

f. The fiscal stimulus program adopted by many countries in response to the financial crisis did not lead to a large increase in the debt-to-GDP ratio.

g. Fiscal and monetary policy successfully saved Japan from a decade of slow growth following its financial crisis in the early 1990s.

2. Traditional monetary and fiscal policy—the IS–LM view

Consider an economy described by Figure 9-9, with output lower than the natural level of output and the nominal interest rate at zero.

a. Draw Figure 9-9 using the LM curve passing through Point A.

b. If the Federal Reserve increases the money supply, what will happen to the IS–LM diagram you drew in part (a)? Will equilibrium output move closer to the natural level?

c. Given your answer to part (b), what policy options are available to the government to try to increase output? Consider traditional policy options only, and not financial policies. How does your answer relate to the policy decisions of the Obama administration and the U.S. Congress in February 2009?

3. Traditional monetary and fiscal policy—the AS–AD view

Consider an economy described by Figure 9-10, with output lower than the natural level of output and the nominal interest rate at zero.

a. Draw Figure 9-10 and explain why the AD curve has a vertical portion.

b. If the Federal Reserve increases the money supply, what will happen to the AS–AD diagram you drew in part (a)? Will equilibrium output move closer to the natural level?

c. Given your answers to part (b), what policy options are available to the government to try to increase output? Consider traditional policy options only, and not financial policies. How does your answer relate to the policy decisions of the Obama administration and the U.S. Congress in February 2009?

4. Nontraditional macroeconomic policy: financial policy and quantitative easing

Consider again the economy described in Figure 9-9, and suppose that the IS and LM relations are

\[ IS: \quad Y = C(Y - T, \text{confidence}) + I(Y, \text{confidence}, \ i + \text{premium}) + G \]

\[ LM: \quad M/P = YL(i) \]

Interpret the interest rate as the federal funds rate, the policy interest rate of the Federal Reserve. Assume that the rate at which firms can borrow is much higher than the federal funds rate, equivalently that the premium in the IS equation is high.

a. Suppose that the government takes action to improve the solvency of the financial system. If the government’s action is successful and banks become more willing to lend—both to one another and to nonfinancial firms—what is likely to happen to the premium? What will happen to the IS–LM diagram? Can we consider financial policy as a kind of macroeconomic policy?

b. Faced with a zero nominal interest rate, suppose the Fed decides to purchase securities directly to facilitate the flow of credit in the financial markets. This policy is called quantitative easing. If quantitative easing is successful, so that it becomes easier for financial and nonfinancial firms to obtain credit, what is likely to happen to the premium? What effect will this have on the IS–LM diagram? If quantitative easing has some effect, is it true that the Fed has no policy options to stimulate the economy when the federal funds rate is zero?

5. Modern bank runs

Consider a simple bank that has assets of 100, capital of 20, and checking deposits of 80. Recall from Chapter 4 that checking deposits are liabilities of a bank.

a. Set up the bank’s balance sheet.

b. Now suppose that the perceived value of the bank’s assets falls by 10. What is the new value of the bank’s capital?

c. Suppose the deposits are insured by the government. Despite the decline in the value of bank capital, is there any immediate reason for depositors to withdraw their funds from the bank? Would your answer change if the perceived value of the bank’s assets fell by 15? 20? 25? Explain.

Now consider a different sort of bank, still with assets of 100 and capital of 20, but now with short-term credit of 80 instead of checkable deposits. Short-term credit must be repaid or rolled over (borrowed again) when it comes due.

d. Set up this bank’s balance sheet.

e. Again suppose the perceived value of the bank’s assets falls. If lenders are nervous about the solvency of the bank, will they be willing to continue to provide short-term credit to the bank at low interest rates?

f. Assuming that the bank cannot raise additional capital, how can it raise the funds necessary to repay its debt coming due? If many banks are in this position at the same time (and if banks hold similar kinds of assets), what will likely happen to the value of the assets of these banks? How will this affect the willingness of lenders to provide short-term credit?

6. The Troubled Asset Relief Program (TARP)

Consider a bank that has assets of 100, capital of 20, and short-term credit of 80. Among the bank’s assets are securitized assets whose value depends on the price of houses. These assets have a value of 50.

a. Set up the bank’s balance sheet.

Suppose that as a result of a housing price decline, the value of the bank’s securitized assets falls by an uncertain amount, so that these assets are now worth somewhere between 25
that the firm now has three assets: 50 of untroubled assets, 25 of troubled assets, and 25 of Treasury bonds.)

What is the total value of the bank’s capital? Will the bank be insolvent?

e. Given your answers and the material in the text, why might recapitalization be a better policy than buying the troubled assets?

EXPLORE FURTHER

7. The TED spread

The text described the fluctuations in the Ted spread that occurred during the financial crisis. Do an internet search and find the recent history of the Ted spread. You can find this information easily from various sources.

a. Consult Figure 9-3 to compare the current value of the Ted spread to its value before and during the financial crisis. How does the current value of the Ted spread compare to its highest values during the crisis? How does the current value of the Ted spread compare to its value at the beginning of 2007? (Note that the Ted spread is often quoted in basis points. One hundred basis points equals one percentage point.)

b. Has the Ted spread been relatively stable in recent months? In what range of values has the spread fluctuated?

c. What do you conclude about the willingness of banks to lend to one another now as compared to the beginning of 2007? as compared to the fall of 2008? Explain.

Further Readings

There are already many good books on the crisis: among them Michael Lewis’s The Big Short (W.W. Norton, 2010) and Gillian Tett’s Fool’s Gold (Free Press, 2009). Both books show how the financial system became increasingly risky until it finally collapsed. Both read like detective novels, with a lot of action and fascinating characters.

In Fed We Trust (Crown Business, 2009), written in 2009 by David Wessel, the economics editor of the Wall Street Journal, describes how the Fed reacted to the crisis. It also makes for fascinating reading.
The Long Run

The next four chapters focus on the long run. In the long run, what dominates is not fluctuations, but growth. So now we need to ask: What determines growth?

Chapter 10

Chapter 10 looks at the facts of growth. It first documents the large increase in output that has taken place in rich countries over the past fifty years. Then, taking a wider look, it shows that on the scale of human history, such growth is a recent phenomenon. And it is not a universal phenomenon: Some countries are catching up, but many poor countries are suffering from no or low growth.

Chapter 11

Chapter 11 focuses on the role of capital accumulation in growth. It shows that capital accumulation cannot by itself sustain growth, but that it does affect the level of output. A higher saving rate typically leads to lower consumption initially, but to more consumption in the long run.

Chapter 12

Chapter 12 turns to technological progress. It shows how, in the long run, the growth rate of an economy is determined by the rate of technological progress. It then looks at the role of R&D in generating such progress. It returns to the facts of growth presented in Chapter 10, and shows how to interpret these facts in the light of the theories developed in Chapters 11 and 12.

Chapter 13

Chapter 13 looks at various issues raised by technological progress in the short, the medium, and the long run. Focusing on the short and the medium run, it discusses the relation between technological progress, unemployment, and wage inequality. Focusing on the long run, it discusses the role of institutions in sustaining technological progress and growth.
Our perceptions of how the economy is doing are often dominated by year-to-year fluctuations in economic activity. A recession leads to gloom, and an expansion to optimism. But if we step back to get a look at activity over longer periods—say over many decades—the picture changes. Fluctuations fade. **Growth**—the steady increase in aggregate output over time—dominates the picture.

Figure 10-1, panels (a) and (b), shows the evolution of U.S. GDP and the evolution of U.S. GDP per person (both in 2000 dollars), respectively, since 1890. (The scale used to measure GDP on the vertical axis in Figure 10-1 is called a logarithmic scale. The defining characteristic of a logarithmic scale is that the same proportional increase in a variable is represented by the same distance on the vertical axis.)

The shaded years from 1929 to 1933 correspond to the large decrease in output during the Great Depression, and the other two shaded ranges correspond to the 1980–1982 recession—the largest postwar recession before the current crisis—and 2008–2010, the most recent crisis, the subject of much of the analysis in the rest of this book. Note how small these three episodes appear compared to the steady increase in output per person over the last 100 years. The cartoon makes the same point about growth and fluctuations, in an even more obvious way.

With this in mind, we now shift our focus from fluctuations to growth. Put another way, we turn from the study of the determination of output in the short and medium run—where fluctuations dominate—to the determination of output in the long run—where growth dominates. Our goal is to understand what determines growth, why some countries are growing while others are not, and why some countries are rich while many others are still poor.

**Section 10-1** discusses a central measurement issue; namely how to measure the standard of living.

**Section 10-2** looks at growth in the United States and other rich countries over the last fifty years.

**Section 10-3** takes a broader look, across both time and space.

**Section 10-4** then gives a primer on growth and introduces the framework that will be developed in the next three chapters.
The reason we care about growth is that we care about the *standard of living*. Looking across time, we want to know by how much the standard of living has increased. Looking across countries, we want to know how much higher the standard of living is in one country relative to another. Thus, the variable we want to focus on, and compare either over time or across countries, is *output per person*, rather than *output* itself.

A practical problem then arises: How do we compare output per person across countries? Countries use different currencies; thus output in each country is expressed in terms of its own currency. A natural solution is to use exchange rates: When

![Figure 10-1](http://hsus.cambridge.org/HSUSWeb/toc/hsusHome.do)

**Panel (a): U.S. GDP since 1890. Panel (b): U.S. GDP per person since 1890**

Panel (a) shows the enormous increase in U.S. output since 1890, by a factor of 43. Panel (b) shows that the increase in output is not simply due to the large increase in U.S. population from 63 million to more than 300 million over this period. Output per person has risen by a factor of 9.


Output per person is also called output per capita ("capita" means "head" in Latin). And given that output and income are always equal, it is also called income per person, or income per capita.
comparing, say, the output per person of India to the output per person of the United States, we can compute Indian GDP per person in rupees, use the exchange rate to get Indian GDP per person in dollars, and compare it to the U.S. GDP per person in dollars. This simple approach will not do, however, for two reasons:

- First, exchange rates can vary a lot (more on this in Chapters 18 to 21). For example, the dollar increased and then decreased in the 1980s by roughly 50% vis-à-vis the currencies of the trading partners of the United States. But, surely, the standard of living in the United States did not increase by 50% and then decrease by 50% compared to the standard of living of its trading partners during the decade. Yet this is the conclusion we would reach if we were to compare GDP per person using exchange rates.

- The second reason goes beyond fluctuations in exchange rates. In 2010, GDP per person in India, using the current exchange rate, was $1,300 compared to $47,300 in the United States. Surely no one could live on $1,300 a year in the United States. But people live on it—admittedly, not very well—in India, where the prices of basic goods—those goods needed for subsistence—are much lower than in the United States. The level of consumption of the average person in India, who consumes mostly basic goods, is not 36 (47,300 divided by 1,300) times smaller than that of the average person in the United States. This point applies to other countries besides the United States and India: In general, the lower a country’s output per person, the lower the prices of food and basic services in that country.

So, when we focus on comparing standards of living, we get more meaningful comparisons by correcting for the two effects we just discussed—variations in exchange rates, and systematic differences in prices across countries. The details of constructing these differences are complicated, but the principle is simple: The numbers for GDP—and hence for GDP per person— are constructed using a common set of prices for all countries. Such adjusted real GDP numbers, which you can think of as measures of purchasing power across time or across countries, are called purchasing power parity (PPP) numbers. Further discussion is given in the Focus box “The Construction of PPP Numbers.”
The Construction of PPP Numbers

Consider two countries—let’s call them the United States and Russia, although we are not attempting to fit the characteristics of those two countries very closely:

In the United States, annual consumption per person equals $20,000. People in the United States each buy two goods: Every year, they buy a new car for $10,000 and spend the rest on food. The price of a yearly bundle of food in the United States is $10,000.

In Russia, annual consumption per person equals 60,000 rubles. People there keep their cars for 15 years. The price of a car is 300,000 rubles, so individuals spend on average 20,000 rubles—300,000/15—a year on cars. They buy the same yearly bundle of food as their U.S. counterparts, at a price of 40,000 rubles.

Russian and U.S. cars are of identical quality, and so are Russian and U.S. food. (You may dispute the realism of these assumptions. Whether a car in country X is the same as a car in country Y is very much the type of problem confronting economists when constructing PPP measures.)

The exchange rate is such that one dollar is equal to 30 rubles. Using this method, Russian consumption per person in Russia relative to consumption per person in the United States?

One way to answer is by taking consumption per person in Russia and converting it into dollars using the exchange rate. Using this method, Russian consumption per person in dollars is $2,000 (60,000 rubles divided by the exchange rate, 30 rubles to the dollar). According to these numbers, consumption per person in Russia is only 10% of U.S. consumption per person.

Does this answer make sense? True, Russians are poorer, but food is much cheaper in Russia. A U.S. consumer spending all of his 20,000 dollars on food would buy 2 bundles of food ($20,000/$10,000). A Russian consumer spending all of his 60,000 rubles on food would buy 1.5 bundles of food (60,000 rubles/40,000 rubles). In terms of food bundles, the difference looks much smaller between U.S. and Russian consumption per person. And given that one-half of consumption in the United States and two-thirds of consumption in Russia go to spending on food, this seems like a relevant computation.

Can we improve on our initial answer? Yes. One way is to use the same set of prices for both countries and then measure the quantities of each good consumed in each country using this common set of prices. Suppose we use U.S. prices. In terms of U.S. prices, annual consumption per person in the United States is obviously still $20,000. What is it in Russia? Every year, the average Russian buys approximately 0.07 car (one car every fifteen years) and one bundle of food. Using U.S. prices—specifically, $10,000 for a car and $10,000 for a bundle of food—gives Russian consumption per person as 

\[
\left[ (0.07 \times $10,000) + (1 \times $10,000) \right] = [$700 + $10,000] = $10,700.
\]

So, using U.S. prices to compute consumption in both countries puts annual Russian consumption per person at $10,700/$20,000 = 53.5% of annual U.S. consumption per person, a better estimate of relative standards of living than we obtained using our first method (which put the number at only 10%).

This type of computation, namely the construction of variables across countries using a common set of prices, underlies PPP estimates. Rather than using U.S. dollar prices as in our example (why use U.S. rather than Russian or, for that matter, French prices?), these estimates use average prices across countries. These average prices are called international dollar prices. Many of the estimates we use in this chapter are the result of an ambitious project known as the “Penn World Tables.” (Penn stands for the University of Pennsylvania, where the project is located.) Led by three economists—Irving Kravis, Robert Summers, and Alan Heston—over the course of more than 40 years, researchers working on the project have constructed PPP series not only for consumption (as we just did in our example), but more generally for GDP and its components, going back to 1950, for most countries in the world.

For more on the construction of PPP numbers, go to the Web site http://pwt.econ.upenn.edu/ associated with the Penn World Tables. (In the Penn tables, what is the ratio of Russian PPP GDP per person to U.S. PPP GDP per person?) The IMF and the World Bank also construct their own set of PPP numbers.

When comparing rich versus poor countries, the differences between PPP numbers and the numbers based on current exchange rates can be very large. Return to the comparison between India and the United States. We saw that, at current exchange rates, the ratio of GDP per person in the United States to GDP per person in India was 36. Using PPP numbers, the ratio is “only” 14. Although this is still a large difference, it is much smaller than the ratio we obtained using current exchange rates. Differences between PPP numbers and numbers based on current exchange rate are typically smaller when making comparisons among rich countries. If we were to compare using current exchange rates—GDP per person in the United States in 2010 was equal to 115% of the GDP per person in Germany. Based on PPP numbers, GDP per person
in the United States is in fact equal to 129% of GDP per person in Germany. More generally, PPP numbers suggest that the United States still has the highest GDP per person among the world’s major countries.

Let me end this section with three remarks before we move on and look at growth:

- What matters for people’s welfare is their consumption rather than their income. One might therefore want to use consumption per person rather than output per person as a measure of the standard of living. (This is indeed what we did in the Focus box, “The Construction of PPP Numbers.”) Because the ratio of consumption to output is rather similar across countries, the ranking of countries is roughly the same, whether we use consumption per person or output per person.

- Thinking about the production side, we may be interested in differences in productivity rather than in differences in the standard of living across countries. In this case, the right measure is output per worker—or, even better, output per hour worked if the information about total hours worked is available—rather than output per person. Output per person and output per worker (or per hour) will differ to the extent that the ratio of the number of workers (or hours) to population differs across countries. Most of the aforementioned difference between output per person in the United States and in Germany comes, for example, from differences in hours worked per person rather than from differences in productivity. Put another way, German workers are about as productive as their U.S. counterparts. However, they work fewer hours, so their standard of living, measured by output per person, is lower. In exchange, however, they enjoy more leisure time.

- The reason we ultimately care about the standard of living is presumably that we care about happiness. We may therefore ask the obvious question: Does a higher standard of living lead to greater happiness? The answer is given in the Focus box “Does Money Buy Happiness?” The answer: a qualified yes.

10-2 Growth in Rich Countries since 1950

Let’s start by looking, in this section, at growth in rich countries since 1950. In the next section, we shall look further back in time and across a wider range of countries.

Table 10-1 shows the evolution of output per person (GDP divided by population, measured at PPP prices) for France, Japan, the United Kingdom, and the United States, since 1950. We have chosen these four countries not only because they are some of the

<table>
<thead>
<tr>
<th>Country</th>
<th>Annual Growth Rate Output per Person (%)</th>
<th>Real Output per Person (2005 dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>2.5</td>
<td>7,112</td>
</tr>
<tr>
<td>Japan</td>
<td>3.9</td>
<td>3,118</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2.0</td>
<td>10,400</td>
</tr>
<tr>
<td>United States</td>
<td>1.9</td>
<td>13,183</td>
</tr>
<tr>
<td>Average</td>
<td>2.6</td>
<td>8,453</td>
</tr>
</tbody>
</table>

Notes: The data stop in 2009, the latest year (at this point) available in the Penn tables. The average in the last line is a simple unweighted average. Source: Alan Heston, Robert Summers, and Bettina Aten, Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, May 2011.
Does Money Lead to Happiness?

Does money lead to happiness? Or, put more accurately, does higher income per person lead to more happiness? The implicit assumption, when economists assess the performance of an economy by looking at its level of income per person or at its growth rate, is that this is indeed the case. Early examinations of data on the relation between income and self-reported measures of happiness suggested that this assumption may not be right. They yielded what is now known as the Easterlin paradox (so named for Richard Easterlin, who was one of the first economists to look systematically at the evidence):

■ Looking across countries, happiness in a country appeared to be higher, the higher the level of income per person. The relation, however, appeared to hold only in relatively poor countries. Looking at rich countries, say the set of OECD countries (look at Chapter 1 for the list), there appeared to be little relation between income per person and happiness.

■ Looking over time, average happiness in rich countries did not seem to increase very much, if at all, with income. (There were no reliable data for poor countries.) In other words, in rich countries, growth did not appear to increase happiness.

Looking across people within a given country, happiness appeared to be strongly correlated with income. Rich people were consistently happier than poor people. This was true in both poor and rich countries.

The first two facts suggested that, once basic needs are satisfied, higher income per person does not increase happiness. The third fact suggested that what was important was not the absolute level of income but the level of income relative to others.

If this interpretation is right, it has major implications for the way we think about the world and about economic policies. In rich countries, policies aimed at increasing income per person might be misdirected because what matters is the distribution of income rather than its average level. Globalization and the diffusion of information, to the extent that it makes people in poor countries compare themselves not to rich people in the same country but to people in richer countries, may actually decrease rather than increase happiness.

So, as you can guess, these findings have led to an intense debate and further research. As new data sets have become available better evidence has accumulated. The state of knowledge and the remaining controversies are analyzed in a recent article by Betsey Stevenson and Justin Wolfers. Their conclusions are well summarized in Figure 1 below.

![Figure 1](image-url)
The figure contains a lot of information. Let’s go through it step by step.

The horizontal axis measures PPP GDP per person for 131 countries. The scale is a logarithmic scale, so a given size interval represents a given percentage increase in GDP. The vertical axis measures average life satisfaction in each country. The source for this variable is a 2006 Gallup World Poll survey, which asked about a thousand individuals in each country the following question:

“Here is a ladder representing the “ladder of life.” Let’s suppose the top of the ladder represents the best possible life for you; and the bottom, the worst possible life for you. On which step of the ladder do you feel you personally stand at the present time?”

The ladder went from 0 to 10. The variable measured on the vertical axis is the average of the individual answers in each country.

Focus first on the dots representing each country, ignoring for the moment the lines that cross each dot. The visual impression is clear: There is a strong relation across countries between average income and average happiness. The index is around 4 in the poorest countries, around 8 in the richest. And, more importantly in view of the early Easterlin paradox, this relation appears to hold both for poor and rich countries.

Focus now on the lines through each dot. The slope of each line reflects the estimated relation between life satisfaction and income across individuals within each country. Note first that all the lines slope upward: This confirms the third leg of the Easterlin paradox: In each country, rich people are happier than poor people. Note also that the slopes of most of these lines are roughly similar to the slope of the relation across countries. This goes against the Easterlin paradox: Individual happiness increases with income, whether this is because the country is getting richer or because the individual becomes relatively richer within the country.

Stevenson and Wolfers draw a strong conclusion from their findings: While individual happiness surely depends on much more than income, it definitely increases with income. Thus, it is not a crime for economists to focus first on levels and growth rates of GDP per person. So, is the debate over? The answer is no. Even if we accept this interpretation of the evidence, clearly, many other aspects of the economy matter for welfare, income distribution surely being one of them. And not everybody is convinced by the evidence. In particular, the evidence on the relation between happiness and income per person over time within a country is not as clear as the evidence across countries or across individuals presented in Figure 1. Given the importance of the question, the debate will continue for some time.


Chapter 10  The Facts of Growth 213

world’s major economic powers, but because what has happened to them is broadly representative of what has happened in other advanced countries over the last half century or so.

Table 10-1 yields two main conclusions:

- There has been a large increase in output per person.
- There has been a convergence of output per person across countries.

Let’s look at each of these points in turn.

The Large Increase in the Standard of Living since 1950

Look at the column on the far right of the table. Output per person has increased by a factor of 3.1 since 1950 in the United States, by a factor of 4.3 in France, and by a factor of 10.2 in Japan.

These numbers show what is sometimes called the force of compounding. In a different context, you probably have heard how saving even a little while you are young will build to a large amount by the time you retire. For example, if the interest rate is 3.9% a year, an investment of one dollar, with the proceeds reinvested every year, will grow to about 10 dollars ($1 + 0.039)^{60}$) 60 years later. The same logic applies to growth rates. The average annual growth rate in Japan over the period 1950 to 2009 was equal to 3.9%. This high growth rate has led to an ten-fold increase in real output per person in Japan over the period.
Clearly, a better understanding of growth, if it leads to the design of policies that stimulate growth, can have a very large effect on the standard of living. Suppose we could find a policy measure that permanently increased the growth rate by 1% per year. This would lead, after 40 years, to a standard of living 48% higher than it would have been without the policy—a substantial difference.

**The Convergence of Output per Person**

The second and third columns of Table 10-1 show that the levels of output per person have converged (become closer) over time: The numbers for output per person are much more similar in 2009 than they were in 1950. Put another way, those countries that were behind have grown faster, reducing the gap between them and the United States.

In 1950, output per person in the United States was roughly twice the level of output per person in France and more than four times the level of output per person in Japan. From the perspective of Europe or Japan, the United States was seen as the land of plenty, where everything was bigger and better. Today these perceptions have faded, and the numbers explain why. Using PPP numbers, U.S. output per person is still the highest, but, in 2009, it was only 20% above average output per person in the other three countries, a much smaller difference than in the 1950s.

This convergence of levels of output per person across countries is not specific to the four countries we are looking at. It extends to the set of OECD countries. This is shown in Figure 10-2, which plots the average annual growth rate of output per person since 1950 against the initial level of output per person in 1950 for the set of countries that are members of the OECD today. There is a clear negative relation between the initial level of output per person and the growth rate since 1950: Countries that were behind in 1950 have typically grown faster. The relation is not perfect: Turkey, which had roughly the same low level of output per person as Japan in 1950, has had a growth rate equal to only about one-half that of Japan. But the relation is clearly there.

Some economists have pointed to a problem in graphs like Figure 10-2. By looking at the subset of countries that are members of the OECD today, what we have done in effect is to look at a club of economic winners: OECD membership is not officially based on economic success, but economic success is surely an important determinant of membership. But when you look at a club whose membership is based on economic success, you will find that those who came from behind had the fastest growth: This is

\[ 1.01^{40} - 1 = 1.48 - 1 = 48\%. \]

Unfortunately, policy measures with such magic results have proven difficult to discover!
precisely why they made it to the club! The finding of convergence could come in part from the way we selected the countries in the first place.

So a better way of looking at convergence is to define the set of countries we look at not on the basis of where they are today—as we did in Figure 10-2 by taking today’s OECD members—but on the basis of where they were in, say, 1950. For example, we can look at all countries that had an output per person of at least one-fourth of U.S. output per person in 1950, and then look for convergence within that group. It turns out that most of the countries in that group have indeed converged, and therefore convergence is not solely an OECD phenomenon. However, a few countries—Uruguay, Argentina, and Venezuela among them—have not converged. In 1950, those three countries had roughly the same output per person as France. In 2009, they had fallen far behind; their level of output per person stood only between one-fourth and one-half of the French level.

10-3  A Broader Look across Time and Space

In the previous section, we focused on growth over the last 50 years in rich countries. Let’s now put this in context by looking at the evidence both over a much longer time span and a wider set of countries.

Looking across Two Millennia

Has output per person in the currently rich economies always grown at rates similar to the growth rates in Table 10-1? The answer is no. Estimates of growth are clearly harder to construct as we look further back in time. But there is agreement among economic historians about the main evolutions over the last 2,000 years.

From the end of the Roman Empire to roughly year 1500, there was essentially no growth of output per person in Europe: Most workers were employed in agriculture in which there was little technological progress. Because agriculture’s share of output was so large, inventions with applications outside agriculture could only contribute little to overall production and output. Although there was some output growth, a roughly proportional increase in population led to roughly constant output per person. This period of stagnation of output per person is often called the Malthusian era. Thomas Robert Malthus, an English economist, at the end of the eighteenth century, argued that this proportional increase in output and population was not a coincidence. Any increase in output, he argued, would lead to a decrease in mortality, leading to an increase in population until output per person was back to its initial level. Europe was in a Malthusian trap, unable to increase its output per person.

Eventually, Europe was able to escape this trap. From about 1500 to 1700, growth of output per person turned positive, but it was still small—only around 0.1% per year. It then increased to just 0.2% per year from 1700 to 1820. Starting with the Industrial Revolution, growth rates increased, but from 1820 to 1950 the growth rate of output per person in the United States was still only 1.5% per year. On the scale of human history, therefore, sustained growth of output per person—especially the high growth rates we have seen since 1950—is definitely a recent phenomenon.

Looking across Countries

We have seen how output per person has converged among OECD countries. But what about the other countries? Are the poorest countries also growing faster? Are they converging toward the United States, even if they are still far behind?

The answer is given in Figure 10-3, which plots the average annual growth rate of output per person since 1960 against output per person for the year 1960, for 76 countries.

The numbers for 1950 are missing for too many countries to use 1950 as the initial year, as we did in Figure 10-2.
The striking feature of Figure 10-3 is that there is no clear pattern: It is not the case that, in general, countries that were behind in 1960 have grown faster. Some have, but many have clearly not.

The cloud of points in Figure 10-3 hides, however, a number of interesting patterns that appear when we put countries into different groups. Note that we have used different symbols in the figure: The diamonds represent OECD countries; the squares represent African countries; the triangles represent Asian countries. Looking at patterns by groups yields three main conclusions:

1. The picture for the OECD countries (for the rich countries) is much the same as in Figure 10-2, which looked at a slightly longer period of time (from 1950 onward, rather than from 1960). Nearly all start at high levels of output per person (say, at least one-third of the U.S. level in 1960), and there is clear evidence of convergence.

2. Convergence is also visible for many Asian countries: Most of the countries with very high growth rates over the period are in Asia. Japan was the first country to take off. Starting a decade later, in the 1960s, four countries—Singapore, Taiwan, Hong Kong, and South Korea, a group of countries sometimes called the four tigers—started catching up as well. In 1960, their average output per person was about 18% of the U.S.; by 2009, it had increased to 83% of U.S. output. More recently, the major story has been China—both because of its very high growth rates and because of its sheer size. Over the period, growth of output per person in China has been 4.4%. But, because it started very low, its output per person is still only about one-sixth of the U.S.

3. The picture is different, however, for African countries. Most African countries (represented by squares) were very poor in 1960, and most have not done well over the period. Many have suffered from either internal or external conflicts. Eight of them have had negative growth of output per person—an absolute decline in their

As we saw in Chapter 9, this fast growth came to an end in the 1990s, and Japan has remained in a slump since.

Paradoxically, the two fastest growing countries in Figure 10-3 are Botswana and Equatorial Guinea, both in Africa. In both cases, however, high growth reflects primarily favorable natural resources—diamonds in Botswana, oil in Guinea.

Figure 10-3

_Growth Rate of GDP per Person since 1960, versus GDP Per Person in 1960 (2005 dollars); 76 Countries_

There is no clear relation between the growth rate of output since 1960 and the level of output per person in 1960.

Source: See Table 10-1.
standard of living between 1960 and 2009. Growth averaged −1% in the Central African Republic and −0.7% in Niger. As a result, output per person in the Central African Republic in 2009 is 60% of its level in 1960. Some hope, however, comes from more recent numbers: Growth of output per person in Sub Saharan Africa, which averaged only 1.3% in the 1990s, was close to 5.0% in the 2000s.

Looking further back in time the following picture emerges. For much of the first millennium, and until the fifteenth century, China probably had the world’s highest level of output per person. For a couple of centuries, leadership moved to the cities of northern Italy. But, until the nineteenth century, differences across countries were typically much smaller than they are today. Starting in the nineteenth century, a number of countries, first in Western Europe, then in North and South America, started growing faster than others. Since then, a number of other countries, most notably in Asia, have started growing fast and are converging. Many others, mainly in Africa, are not.

Our main focus, in this and the next chapter, will primarily be on growth in rich and emerging countries. We shall not take on some of the wider challenges raised by the facts we have just seen, such as why growth of output per person started in earnest in the nineteenth century or why Africa has remained so poor. Doing so would take us too far into economic history and development economics. But these facts put into perspective the two basic facts we discussed earlier when looking at the OECD: Neither growth nor convergence is a historical necessity.

10-4 Thinking About Growth: A Primer

To think about growth economists use a framework developed originally by Robert Solow, from the Massachusetts Institute of Technology (MIT) in the late 1950s. The framework has proven sturdy and useful, and we will use it here. This section provides an introduction. Chapters 11 and 12 will provide a more detailed analysis, first of the role of capital accumulation and then of the role of technological progress in the process of growth.

The Aggregate Production Function

The starting point for any theory of growth must be an aggregate production function, a specification of the relation between aggregate output and the inputs in production.

The aggregate production function we introduced in Chapter 6 to study the determination of output in the short run and the medium run took a particularly simple form. Output was simply proportional to the amount of labor used by firms—more specifically, proportional to the number of workers employed by firms (equation (6.2)). So long as our focus was on fluctuations in output and employment, the assumption was acceptable. But now that our focus has shifted to growth this assumption will no longer do: It implies that output per worker is constant, ruling out growth (or at least growth of output per worker) altogether. It is time to relax it. From now on, we will assume that there are two inputs—capital and labor—and that the relation between aggregate output and the two inputs is given by:

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

As before, $Y$ is aggregate output. $K$ is capital—the sum of all the machines, plants, and office buildings in the economy. $N$ is labor—the number of workers in the economy. The function $F$, which tells us how much output is produced for given quantities of capital and labor, is the aggregate production function.
This way of thinking about aggregate production is an improvement on our treatment in Chapter 6. But it should be clear that it is still a dramatic simplification of reality. Surely, machines and office buildings play very different roles in production and should be treated as separate inputs. Surely, workers with Ph.D.’s are different from high-school dropouts; yet, by constructing the labor input as simply the number of workers in the economy, we treat all workers as identical. We will relax some of these simplifications later. For the time being, equation (10.1), which emphasizes the role of both labor and capital in production, will do.

The next step must be to think about where the aggregate production function \( F \), which relates output to the two inputs, comes from. In other words, what determines how much output can be produced for given quantities of capital and labor? The answer: the state of technology. A country with a more advanced technology will produce more output from the same quantities of capital and labor than will an economy with a primitive technology.

How should we define the state of technology? Should we think of it as the list of blueprints defining both the range of products that can be produced in the economy as well as the techniques available to produce them? Or should we think of it more broadly, including not only the list of blueprints, but also the way the economy is organized—from the internal organization of firms, to the system of laws and the quality of their enforcement, to the political system, and so on? In the next two chapters we will have in mind the narrower definition—the set of blueprints. In Chapter 13, however, we will consider the broader definition and return to what we know about the role of the other factors, from legal institutions to the quality of government.

**Returns to Scale and Returns to Factors**

Now that we have introduced the aggregate production function, the next question is: What restrictions can we reasonably impose on this function?

Consider first a thought experiment in which we double both the number of workers and the amount of capital in the economy. What do you expect will happen to output? A reasonable answer is that output will double as well: In effect, we have cloned the original economy, and the clone economy can produce output in the same way as the original economy. This property is called constant returns to scale: If the scale of operation is doubled—that is, if the quantities of capital and labor are doubled—then output will also double.

\[
2Y = F(2K, 2N)
\]

Or, more generally, for any number \( x \) (this will be useful below)

\[
xY = F(xK, xN)
\] (10.2)

We have just looked at what happens to production when both capital and labor are increased. Let’s now ask a different question. What should we expect to happen if only one of the two inputs in the economy—say capital—is increased?

Surely output will increase. That part is clear. But it is also reasonable to assume that the same increase in capital will lead to smaller and smaller increases in output as the level of capital increases. In other words, if there is little capital to start with, a little more capital will help a lot. If there is a lot of capital to start with, a little more capital may make little difference. Why? Think, for example, of a secretarial pool, composed of a given number of secretaries. Think of capital as computers. The introduction of the first computer will substantially increase the pool’s production, because some of the more time-consuming tasks can now be done automatically by the computer. As the number of computers increases and more secretaries in the pool get their own
computers, production will further increase, although perhaps by less per additional computer than was the case when the first one was introduced. Once each and every secretary has a computer, increasing the number of computers further is unlikely to increase production very much, if at all. Additional computers might simply remain unused and left in their shipping boxes and lead to no increase in output.

We shall refer to the property that increases in capital lead to smaller and smaller increases in output as **decreasing returns to capital** (a property that will be familiar to those who have taken a course in microeconomics).

A similar argument applies to the other input, labor. Increases in labor, given capital, lead to smaller and smaller increases in output. (Return to our example, and think of what happens as you increase the number of secretaries for a given number of computers.) There are **decreasing returns to labor** as well.

### Output per Worker and Capital per Worker

The production function we have written down, together with the assumption of constant returns to scale, implies that there is a simple relation between output per worker and capital per worker.

To see this, set \( x = 1/N \) in equation (10.2), so that

\[
\frac{Y}{N} = F\left(\frac{K}{N}, \frac{N}{N}\right) = F\left(\frac{K}{N}, 1\right)
\]

(10.3)

Note that \( Y/N \) is output per worker, \( K/N \) is capital per worker. So equation (10.3) tells us that the amount of output per worker depends on the amount of capital per worker. This relation between output per worker and capital per worker will play a central role in what follows, so let’s look at it more closely.

This relation is drawn in Figure 10-4. Output per worker \( (Y/N) \) is measured on the vertical axis, and capital per worker \( (K/N) \) is measured on the horizontal axis. The relation between the two is given by the upward-sloping curve. As capital per worker increases, so does output per worker. Note that the curve is drawn so that increases in capital lead to smaller and smaller increases in output. This follows from the property that there are **decreasing returns to capital**: At point \( A \), where capital per worker is low, an increase in capital per worker, represented by the horizontal distance \( AB \), leads to an increase in output per worker equal to the vertical distance \( A'B' \). At point \( C \), where capital per worker is larger, the same increase in capital per worker, represented by the

Even under constant returns to scale, there are decreasing returns to each factor, keeping the other factor constant:

**There are decreasing returns to capital**: Given labor, increases in capital lead to smaller and smaller increases in output.

**There are decreasing returns to labor**: Given capital, increases in labor lead to smaller and smaller increases in output.

Make sure you understand what is behind the algebra. Suppose capital and the number of workers both double. What happens to output per worker?

---

![Figure 10-4](image-url)

**Figure 10-4**

**Output and Capital per Worker**

Increases in capital per worker lead to smaller and smaller increases in output per worker.
Increases in capital per worker lead to smaller and smaller increases in output per worker as the level of capital per worker increases. Horizontal distance $CD$ (where the distance $CD$ is equal to the distance $AB$), leads to a much smaller increase in output per worker, only $C'D'$. This is just like our secretarial pool example, where additional computers had less and less impact on total output.

**The Sources of Growth**

We are now ready to return to our basic question: Where does growth come from? Why does output per worker—or output per person, if we assume the ratio of workers to the population as a whole remains constant over time—go up over time? Equation (10.3) gives a first answer:

- Increases in output per worker $(Y/N)$ can come from increases in capital per worker $(K/N)$. This is the relation we just looked at in Figure 10-4. As $(K/N)$ increases—that is, as we move to the right on the horizontal axis—$(Y/N)$ increases. Or they can come from improvements in the state of technology that shift the production function, $F$, and lead to more output per worker given capital per worker. This is shown in Figure 10-5. An improvement in the state of technology shifts the production function up, from $F(K/N, 1)$ to $F(K/N, 1)'$. For a given level of capital per worker, the improvement in technology leads to an increase in output per worker. For example, for the level of capital per worker corresponding to point $A$, output per worker, increases from $A'$ to $B'$. (To go back to our secretarial pool example, a reallocation of tasks within the pool may lead to a better division of labor and an increase in the output per secretary.)

Hence, we can think of growth as coming from **capital accumulation** and from **technological progress**—the improvement in the state of technology. We will see, however, that these two factors play very different roles in the growth process:

- Capital accumulation by itself cannot sustain growth. A formal argument will have to wait until Chapter 11. But you can already see the intuition behind this from Figure 10-5. Because of decreasing returns to capital, sustaining a steady increase in output per worker will require larger and larger increases in the level of capital per worker. At some stage, the economy will be unwilling or unable to save and invest enough to further increase capital. At that stage, output per worker will stop growing.

Does this mean that an economy’s **saving rate**—the proportion of income that is saved—is irrelevant? No. It is true that a higher saving rate cannot permanently increase the growth rate of output. But a higher saving rate can sustain a higher...
increase consumption, but by much less than the increase in their income. Firms that expect sales to be higher only for a year are unlikely to change their investment plans much, if at all.

Putting things together, a large decrease in the current real interest rate—from $r_A$ to $r_B$ in Figure 17-2—leads to only a small increase in output, from $Y_A$ to $Y_B$. Put another way: The IS curve, which goes through points $A$ and $B$, is steeply downward sloping.

A change in any variable in equation (17.2) other than $Y$ and $r$ shifts the IS curve:
- Changes in current taxes ($T$) or in current government spending ($G$) shift the IS curve.
  An increase in current government spending increases spending at a given interest rate, shifting the IS curve to the right; an increase in taxes shifts the IS curve to the left. These shifts are represented in Figure 17-2.
- Changes in expected future variables also shift the IS curve.
  An increase in expected future output, $Y^{e}$, shifts the IS curve to the right: Higher expected future income leads consumers to feel wealthier and spend more; higher expected future output implies higher expected profits, leading firms to invest more. Higher spending by consumers and firms leads, through the multiplier effect, to higher output. By a similar argument, an increase in expected future taxes leads consumers to decrease their current spending and shifts the IS curve to the left. And an increase in the expected future real interest rate decreases current spending, also leading to a decrease in output, shifting the IS curve to the left. These shifts are also represented in Figure 17-2.

The LM Relation Revisited

The LM relation we derived in Chapter 4 and have used until now was given by

$$\frac{M}{P} = Y L(\ i)$$

where $M/P$ is the supply of money and $Y L(\ i)$ is the demand for money. Equilibrium in financial markets requires that the supply of money be equal to the demand for money. The demand for money depends on real income and on the short-term nominal interest rate—the opportunity cost of holding money. We derived this demand for money before thinking about expectations. Now that we have introduced them, the question is whether we should modify equation (17.3). The answer—we are sure this will be good news—is: no.

Think of your own demand for money. How much money you want to hold today depends on your current level of transactions, not on the level of transactions you expect next year or the year after; there will be ample time for you to adjust your money balances to your transaction level if and when it changes in the future. And the opportunity cost of holding money today depends on the current nominal interest rate, not on the expected nominal interest rate next year or the year after. If short-term interest rates were to increase in the future, increasing the opportunity cost of holding money then, the time to reduce your money balances would be then, not now.

So, in contrast to the consumption decision, the decision about how much money to hold is myopic, depending primarily on current income and the current short-term nominal interest rate. We can still think of the demand for money as depending on the current level of output and the current nominal interest rate, and use equation (17.3) to describe the determination of the nominal interest rate in the current period.

Let’s summarize:

We have seen that expectations about the future play a major role in spending decisions. This implies that expectations enter the IS relation: Private spending depends not only on current output and the current real interest rate, but also on expected future output and the expected future real interest rate.
In contrast, the decision about how much money to hold is largely myopic: The two variables entering the \( LM \) relation are still current income and the current nominal interest rate.

17-2 Monetary Policy, Expectations, and Output

In the basic \( IS-LM \) model we developed in Chapter 5, there was only one interest rate, \( i \), which entered both the \( IS \) relation and the \( LM \) relation. When the Fed expanded the money supply, “the” interest rate went down, and spending increased. From the previous three chapters, you have learned that there are in fact many interest rates, and that we must keep at least two distinctions in mind:

1. The distinction between the nominal interest rate and the real interest rate.
2. The distinction between current and expected future interest rates.

The interest rate that enters the \( LM \) relation, which is the interest rate that the Fed affects directly, is the current nominal interest rate. In contrast, spending in the \( IS \) relation depends on both current and expected future real interest rates. Economists sometimes state this distinction even more starkly by saying that, while the Fed controls the short-term nominal interest rate, what matters for spending and output is the long-term real interest rate. Let’s look at this distinction more closely.

From the Short Nominal Rate to Current and Expected Real Rates

- Recall from Chapter 14 that the real interest rate is approximately equal to the nominal interest rate minus expected current inflation:
  \[
  r = i - \pi^e
  \]

- Similarly, the expected future real interest rate is approximately equal to the expected future nominal interest rate minus expected future inflation:
  \[
  r^e = i^e - \pi^e
  \]

When the Fed increases the money supply—decreasing the current nominal interest rate \( i \)—the effects on the current and the expected future real interest rates depend, therefore, on two factors:

- Whether the increase in the money supply leads financial markets to revise their expectations of the future nominal interest rate, \( i^e \).
- Whether the increase in the money supply leads financial markets to revise their expectations of both current and future inflation, \( \pi^e \) and \( \pi^e^{\prime} \). If, for example, the change in money leads financial markets to expect more inflation in the future—so \( \pi^e^{\prime} \) increases—the expected future real interest rate, \( r^e \), will decrease by more than the expected future nominal interest rate, \( i^e \).

To keep things simple, we shall ignore here the second factor—the role of changing expectations of inflation—and focus on the first, the role of changing expectations of the future nominal interest rate. Thus we shall assume that expected current inflation and expected future inflation are both equal to zero. In this case, we do not need to distinguish between the nominal interest rate and the real interest rate, as they are equal, and we can use the same letter to denote both. Let \( r \) denote the current real (and nominal) interest rate, and \( r^e \) denote the expected future real (and nominal) interest rate.
With this simplification, we can rewrite the $IS$ and $LM$ relations in equations (17.2) and (17.3) as:

$$IS: \quad Y = A(Y, T, r, T', r') + G \quad (17.4)$$

$$LM: \quad \frac{M}{P} = Y L(r) \quad (17.5)$$

The corresponding $IS$ and $LM$ curves are drawn in Figure 17-3. The vertical axis measures the current interest rate $r$; the horizontal axis measures current output $Y$. The $IS$ curve is steeply downward sloping. We saw the reason why earlier: For given expectations, a change in the current interest rate has a limited effect on spending, and the multiplier is small. The $LM$ curve is upward sloping. An increase in income leads to an increase in the demand for money; given the supply of money, the result is an increase in the interest rate. Equilibrium in goods and financial markets implies that the economy is at point $A$, on both the $IS$ and the $LM$ curves.

### Monetary Policy Revisited

Now suppose the economy is in recession, and the Fed decides to increase the money supply.

Assume first that this expansionary monetary policy does not change expectations of either the future interest rate or future output. In Figure 17-4, the $LM$ shifts down, from $LM$ to $LM'$. (Because I have already used primes to denote future values of the variables, we shall use double primes, such as in $LM''$, to denote shifts in curves in this chapter.) The equilibrium moves from point $A$ to point $B$, with higher output and a lower interest rate. The steep $IS$ curve, however, implies that the increase in the money supply has only a small effect on output: Changes in the current interest rate, accompanied by changes in expectations, have only a small effect on spending, and in turn a small effect on output.

Is it reasonable, however, to assume that expectations are unaffected by an expansionary monetary policy? Isn’t it likely that, as the Fed lowers the current interest rate, financial markets now anticipate lower interest rates in the future as well, along with higher future output stimulated by this lower future interest rate? What happens if they do? At a given current interest rate, prospects of a lower future interest rate and

---

**Figure 17-3**

**The New IS–LM**

The $IS$ curve is steeply downward sloping: Other things being equal, a change in the current interest rate has a small effect on output. The $LM$ curve is upward sloping. The equilibrium is at the intersection of the $IS$ and $LM$ curves. In this diagram, I have assumed the expected rate of inflation is zero in both the current year and in the future.
Expectations

Extensions

The effects of higher future output both increase spending and output; they shift the IS curve to the right, from IS to IS*. The new equilibrium is given by point C. Thus, while the direct effect of the monetary expansion on output is limited, the full effect, once changes in expectations are taken into account, is much larger.

You have just learned an important lesson. The effects of monetary policy—the effects of any type of macroeconomic policy for that matter—depend crucially on its effect on expectations:

- If a monetary expansion leads financial investors, firms, and consumers to revise their expectations of future interest rates and output, then the effects of the monetary expansion on output may be very large.
- But if expectations remain unchanged, the effects of the monetary expansion on output will be small.

The role of expectations is even more central in a case we have left aside, namely the case where the economy is in the liquidity trap and the short-term rate is already equal to zero. The discussion of how monetary policy may still work in this case is taken up in the Focus box “The Liquidity Trap, Quantitative Easing, and the Role of Expectations.”

We can link this to our discussion to our earlier discussion in Chapter 15, about the effects of changes in monetary policy on the stock market. Many of the same issues were present there. If, when the change in monetary policy takes place, it comes as no surprise to investors, firms, and consumers, then expectations will not change. The stock market will react only a little, if at all. And demand and output will change only a little, if at all. But, if the change comes as a surprise and is expected to last, expectations of future output will go up, expectations of future interest rates will come down, the stock market will boom, and output will increase.

At this stage, you may have become very skeptical that macroeconomists can say much about the effects of policy, or the effects of other shocks: If the effects depend so much on what happens to expectations, can macroeconomists have any hope of predicting what will happen? The answer is yes.

Saying that the effect of a particular policy depends on its effect on expectations is not the same as saying that anything can happen. Expectations are not arbitrary. The manager of a mutual fund who must decide whether to invest in stocks or bonds, the firm thinking about whether or not to build a new plant, the
At the time of this writing, the scope for conventional monetary policy, namely a reduction in the nominal short-term interest rate, has simply disappeared: Since the end of 2008, the nominal short-term interest rate is close to zero, and the U.S. economy has been in a liquidity trap. This has led the Fed, as well as other central banks, to explore unconventional policies. These come by the name of “quantitative easing” or “credit easing.” We briefly talked about them in Chapter 9. We look at them more closely here.

First, the semantics: Economists typically refer to “quantitative easing” to denote operations in which, while at the zero interest rate bound, the central bank continues to increase the money supply through open market operations, either by buying more T-bills or by buying long-maturity government bonds. The purpose of quantitative easing is to increase the money supply. Economists apply the term “credit easing” to operations where the central bank buys a specific type of asset; for example, mortgage-based securities, or even stocks. The focus is then not so much on the increase in the money supply, but on the effects on the price or the interest rate on the specific asset being bought.

Why should such operations have any effect? After all, we saw in Chapter 9 how, at least under the assumptions we made there, an increase in the money supply had no effect on the short-term nominal interest rate. As both bonds and money paid the same nominal interest rate, namely zero, people willingly held more money and less bonds in response to an open market operation, leaving the nominal interest rate unchanged and equal to zero. As we saw in Chapter 15, so long as expectations of future interest rates were unchanged, arbitrage between short- and long-term bonds implied that interest rates on longer-term bonds also would not change. Nor would stock prices or other asset prices.

This simple argument is why many economists are skeptical that these unconventional policies can do much to increase spending and output. But you can see that there are a number of qualifiers in the previous paragraph. If some of the assumptions we stated there are not correct, these policies may have an effect. Economists have identified three channels through which quantitative or credit easing may affect the economy:

- **Arbitrage may not hold.** Investors may, for example, think that an asset is so risky that they do not want to hold it at all. Or, and this happened during the crisis, investors may be short of funds and may have to sell the asset, even if they wanted to keep it. This is known as a case of fire sales. In this case, by doing credit easing, (i.e., by buying the asset), the central bank can replace these investors, increase the price of the asset, and decrease the associated interest rate. Or, to take another example, if banks, for the reasons we saw in Chapter 9, suddenly lose some of their funding and are forced in turn to ration loans and turn down some borrowers, the central bank may be able to finance some of the borrowers. In short, when arbitrage fails, credit easing can work.

- **Quantitative easing may affect expectations of future nominal interest rates.** As this chapter makes clear, given inflation expectations, what matters is not so much the current nominal interest rate as future nominal interest rates. If the increase in the money supply is taken as a signal by markets that the central bank will continue to follow a very expansionary monetary policy in the future, and thus keep nominal interest rates low for a long time, this will increase spending today.

- **Quantitative easing may affect expectations of inflation.** In the strange world of the liquidity trap, higher expected inflation is good, as it leads to lower current and future expected real interest rates. Thus, if the large increase in the money supply leads people to expect more inflation in the future, this will also increase spending today.

None of these channels is a sure thing. The first is more likely to work during the acute phase of the crisis, when perceptions of risk are high and some investors have to sell assets in a hurry. The other two work through expectations and are far from mechanical. If expectations of either future nominal interest rates or of inflation do not move, easing will have no effect on spending, and in turn no effect on output.

What does the evidence suggest? To answer, we can look at what has happened in the United States during the crisis. Once the room for conventional monetary policy was exhausted, the Fed decided to use both credit and quantitative easing. In November 2008, it started a program known as Quantitative Easing I, or QEI for short, in which it purchased large amounts of mortgage-based securities. (A more appropriate name for the program would be Credit Easing I, as the purpose was clearly to decrease the interest rate on those particular assets, assets that private investors no longer wanted to hold.) In August 2010, in what is known as Quantitative Easing II, or QEII, the Fed started purchasing long-term government bonds, further increasing the money supply. Figure 1 shows the size and the composition of the assets held by the Fed since 2007. The yellow surface shows the holdings of mortgage-based securities and a few other assets; the blue surface represents the holdings of long-term government bonds (the green surface corresponds...
consumer thinking about how much she should save for retirement, all give a lot of thought to what might happen in the future. We can think of each of them as forming expectations about the future by assessing the likely course of future expected policy and then working out the implications for future activity. If they do not do it themselves (surely most of us do not spend our time solving macroeconomic models before making decisions), they do so indirectly by watching TV and reading newsletters and newspapers or finding public information on the Web, all of which in turn rely on the forecasts of public and private forecasters. Economists refer to expectations formed in this forward-looking manner as rational expectations. The introduction of the assumption of rational expectations is one of the important developments in macroeconomics in the last 35 years. It has largely shaped the way macroeconomists think about policy. It is discussed further in the Focus box “Rational Expectations.”

We could go back and think about the implications of rational expectations in the case of the monetary expansion we have just studied. It will be more fun to do this in the context of a change in fiscal policy, and this is what we now turn to.
Rational Expectations

Most macroeconomists today routinely solve their models under the assumption of rational expectations. This was not always the case. The last 40 years in macroeconomic research are often called the “rational expectations” revolution.

The importance of expectations is an old theme in macroeconomics. But until the early 1970s, macroeconomists thought of expectations in one of two ways:

- One was as animal spirits (from an expression Keynes introduced in the General Theory to refer to movements in investment that could not be explained by movements in current variables). In other words, shifts in expectations were considered important but were left unexplained.

- The other was as the result of simple, backward-looking rules. For example, people were often assumed to have static expectations; that is, to expect the future to be like the present (we used this assumption when discussing the Phillips curve in Chapter 8 and when exploring investment decisions in Chapter 16). Or people were assumed to have adaptive expectations: If, for example, their forecast of a given variable in a given period turned out to be too low, people were assumed to “adapt” by raising their expectation for the value of the variable for the following period. For example, seeing an inflation rate higher than they had expected led people to revise upward their forecast of inflation in the future.

In the early 1970s, a group of macroeconomists led by Robert Lucas (at Chicago) and Thomas Sargent (at Minnesota) argued that these assumptions did not reflect the way people form expectations. (Robert Lucas received the Nobel Prize in 1995; Thomas Sargent received the Nobel Prize in 2011.) They argued that, in thinking about the effects of alternative policies, economists should assume that people have rational expectations, that people look into the future and do the best job they can in predicting it. This is not the same as assuming that people know the future, but rather that they use the information they have in the best possible way.

Using the popular macroeconomic models of the time, Lucas and Sargent showed how replacing traditional assumptions about expectations formation by the assumption of rational expectations could fundamentally alter the results. We saw, for example, in Chapter 8 how Lucas challenged the notion that disinflation necessarily required an increase in unemployment for some time. Under rational expectations, he argued, a credible disinflation policy might be able to decrease inflation without any increase in unemployment. More generally, Lucas and Sargent’s research showed the need for a complete rethinking of macroeconomic models under the assumption of rational expectations, and this is what happened over the next two decades.

Most macroeconomists today use rational expectations as a working assumption in their models and analyses of policy. This is not because they believe that people always have rational expectations. Surely there are times when people, firms, or financial market participants lose sight of reality and become too optimistic or too pessimistic. (Recall our discussion of bubbles and fads in Chapter 15.) But these are more the exception than the rule, and it is not clear that economists can say much about those times anyway. When thinking about the likely effects of a particular economic policy, the best assumption to make seems to be that financial markets, people, and firms will do the best they can to work out the implications of that policy. Designing a policy on the assumption that people will make systematic mistakes in responding to it is unwise.

Why did it take until the 1970s for rational expectations to become a standard assumption in macroeconomics? Largely because of technical problems. Under rational expectations, what happens today depends on expectations of what will happen in the future. But what happens in the future also depends on what happens today. Solving such models is hard. The success of Lucas and Sargent in convincing most macroeconomists to use rational expectations comes not only from the strength of their case, but also from showing how it could actually be done. Much progress has been made since in developing solution methods for larger and larger models. Today, a number of large macroeconomic models are solved under the assumption of rational expectations. (The simulation of the Taylor model presented in the box on monetary policy in Chapter 7 was derived under rational expectations. You will see another example in Chapter 22.)

17-3 Deficit Reduction, Expectations, and Output

Recall the conclusions we reached in the core about the effects of a budget deficit reduction:

- In the long run, a reduction in the budget deficit is good for the economy: In the medium run, a lower budget deficit implies higher saving and higher investment. In the long run, higher investment translates into higher capital and thus higher output.
In the short run, however, a reduction in the budget deficit, unless it is offset by a monetary expansion, leads to lower spending and to a contraction in output.

It is this adverse short-run effect that—in addition to the unpopularity of increases in taxes or reductions in government programs in the first place—often deters governments from tackling their budget deficits: Why take the risk of a recession now for benefits that will accrue only in the future?

In the recent past, however, a number of economists have argued that, under some conditions, a deficit reduction might actually increase output even in the short run. Their argument: If people take into account the future beneficial effects of deficit reduction, their expectations about the future might improve enough so as to lead to an increase—rather than a decrease—in current spending, thereby increasing current output. This section explores their argument. The Focus box “Can a Budget Deficit Reduction Lead to an Output Expansion? Ireland in the 1980s” reviews some of the supporting evidence.

Assume the economy is described by equation (17.4) for the IS relation and equation (17.5) for the LM relation. Now suppose the government announces a program to reduce the deficit, through decreases both in current spending, $G$, and in future spending, $G^e$. What will happen to output this period?

### The Role of Expectations about the Future

Suppose first that expectations of future output, $Y^e$, and of the future interest rate, $r^e$, do not change. Then we get the standard answer: The decrease in government spending in the current period leads to a shift of the IS curve to the left, and so to a decrease in output.

The crucial question therefore is what happens to expectations. To answer, let us go back to what we learned in the core about the effects of a deficit reduction in the medium run and the long run:

- **In the medium run**, a deficit reduction has no effect on output. It leads, however, to a lower interest rate and to higher investment. These were two of the main lessons of Chapter 7.

  Let’s review the logic behind each:

  Recall that, when we look at the medium run, we ignore the effects of capital accumulation on output. So, in the medium run, the natural level of output depends on the level of productivity (taken as given) and on the natural level of employment. The natural level of employment depends in turn on the natural rate of unemployment. If spending by the government on goods and services does not affect the natural rate of unemployment—and there is no obvious reason why it should—then changes in spending will not affect the natural level of output. Therefore, a deficit reduction has no effect on the level of output in the medium run.

  Now recall that output must be equal to spending, and that spending is the sum of public spending and private spending. Given that output is unchanged and that public spending is lower, private spending must therefore be higher. Higher private spending requires a lower equilibrium interest rate: The lower interest rate leads to higher investment, and thus to higher private spending, which offsets the decrease in public spending and output unchanged.

- **In the long run**—that is, taking into account the effects of capital accumulation on output—higher investment leads to a higher capital stock, and, therefore, a higher level of output.

  This was the main lesson of Chapter 11. The higher the proportion of output saved (or invested; investment and saving must be equal for the goods market to
be in equilibrium in a closed economy), the higher the capital stock, and thus the higher the level of output in the long run.

We can think of our future period as including both the medium and the long run. If people, firms, and financial market participants have rational expectations, then, in response to the announcement of a deficit reduction, they will expect these developments to take place in the future. Thus, they will revise their expectation of future output \((Y^{e})\) up, and their expectation of the future interest rate \((r^{e})\) down.

**Back to the Current Period**

We can now return to the question of what happens this period in response to the announcement and start of the deficit reduction program. Figure 17-5 draws the IS and LM curves for the current period. In response to the announcement of the deficit reduction, there are now three factors shifting the IS curve:

- Current government spending \((G)\) goes down, leading the IS curve to shift to the left. At a given interest rate, the decrease in government spending leads to a decrease in total spending and so a decrease in output. This is the standard effect of a reduction in government spending, and the only one taken into account in the basic IS–LM model.

- Expected future output \((Y^{e})\) goes up, leading the IS curve to shift to the right. At a given interest rate, the increase in expected future output leads to an increase in private spending, increasing output.

- The expected future interest rate \((r^{e})\) goes down, leading the IS curve to shift to the right. At a given current interest rate, a decrease in the future interest rate stimulates spending and increases output.

What is the net effect of these three shifts in the IS curve? Can the effect of expectations on consumption and investment spending offset the decrease in government spending? Without much more information about the exact form of the IS and LM relations and about the details of the deficit reduction program, we cannot tell which shifts will dominate, and whether output will go up or down. But our analysis tells us that both cases are possible—that output may go up in response to the deficit reduction. And it gives us a few hints as to when this might happen:

- Note that the smaller the decrease in current government spending \((G)\), the smaller the adverse effect on spending today. Note also that the larger the decrease

The way this is likely to happen: Forecasts by economists will show that these lower deficits are likely to lead to higher output and lower interest rates in the future. In response to these forecasts, long-term interest rates will decrease and the stock market will increase. People and firms, reading these forecasts and looking at bond and stock prices, will revise their spending plans and increase spending.

![Figure 17-5](image-url)

**Figure 17-5**

*The Effects of a Deficit Reduction on Current Output*

When account is taken of its effect on expectations, the decrease in government spending need not lead to a decrease in output.
Can a Budget Deficit Reduction Lead to an Output Expansion? Ireland in the 1980s

Ireland went through two major deficit reduction programs in the 1980s:

1. The first program was started in 1982. In 1981, the budget deficit had reached a very high 13% of GDP. Government debt, the result of the accumulation of current and past deficits, was 77% of GDP, also a very high level. The Irish government clearly had to regain control of its finances. Over the next three years, it embarked on a program of deficit reduction, based mostly on tax increases. This was an ambitious program: Had output continued to grow at its normal growth rate, the program would have reduced the deficit by 5% of GDP.

The results, however, were dismal. As shown in line 2 of Table 1, output growth was low in 1982, and negative in 1983. Low output growth was associated with a major increase in unemployment, from 9.5% in 1981 to 15% in 1984 (line 3). Because of low output growth, tax revenues—which depend on the level of economic activity—were lower than anticipated. The actual deficit reduction from 1981 to 1984, shown in line 1, was only of 3.5% of GDP. And the result of continuing high deficits and low GDP growth was a further increase in the ratio of debt to GDP to 97% in 1984.

2. A second attempt to reduce budget deficits was made starting in February 1987. At the time, things were still very bad. The 1986 deficit was 10.7% of GDP; debt stood at 116% of GDP, a record high in Europe at the time. This new program of deficit reduction was different from the first. It was focused more on reducing the role of government and cutting government spending than on increasing taxes. The tax increases in the program were achieved through a tax reform widening the tax base—increasing the number of households paying taxes—rather than through an increase in the marginal tax rate. The program was again very ambitious: Had output continued to grow at its normal rate, the reduction in the deficit would have been 6.4% of GDP.

The results of the second program could not have been more different from the results of the first. 1987 to 1989 were years of strong growth, with average GDP growth exceeding 5%. The unemployment rate was reduced by almost 2%. Because of strong output growth, tax revenues were higher than anticipated, and the deficit was reduced by nearly 9% of GDP.

A number of economists have argued that the striking difference between the results of the two programs can be traced to the different reaction of expectations in each case. The first program, they argue, focused on tax increases and did not change what many people saw as too large a role of government in the economy. The second program, with its focus on cuts in spending and on tax reform, had a much more positive impact on expectations, and so a positive impact on spending and output.

Are these economists right? One variable, the household saving rate—defined as disposable income minus consumption, divided by disposable income—strongly suggests that expectations are an important part of the story. To interpret the behavior of the saving rate, recall the lessons from Chapter 16 about consumption behavior. When disposable income grows unusually slowly or goes down—as it does in a recession—consumption typically slows down or declines by less than disposable income because people expect things to improve in the future. Put another way, when the growth of disposable income is unusually low, the saving rate typically comes down. Now look (in line 4) at what happened from 1981 to 1984: Despite low growth throughout the period and a recession in 1983, the household saving rate actually increased slightly during the period. Put another way, people reduced their consumption by more than the reduction in their disposable income: The reason must be that they were very pessimistic about the future.

| Table 1 Fiscal and Other Macroeconomic Indicators, Ireland, 1981 to 1984, and 1986 to 1989 |
|---------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 Budget deficit (% of GDP)    | −13.0  | −13.4  | −11.4  | −9.5   | −10.7  | −8.6   | −4.5   | −1.8   |
| 2 Output growth rate (%)       | 3.3    | 2.3    | −0.2   | 4.4    | −0.4   | 4.7    | 5.2    | 5.8    |
| 3 Unemployment rate (%)        | 9.5    | 11.0   | 13.5   | 15.0   | 17.1   | 16.9   | 16.3   | 15.1   |
| 4 Household saving rate (%)    | 17.9   | 19.6   | 18.1   | 18.4   | 15.7   | 12.9   | 11.0   | 12.6   |

Source: OECD Economic Outlook, June 1998
Now turn to the period 1986 to 1989. During that period, economic growth was unusually strong. By the same argument as in the previous paragraph, we would have expected consumption to increase less strongly, and thus the saving rate to increase. Instead, the saving rate dropped sharply, from 15.7% in 1986 to 12.6% in 1989. Consumers must have become much more optimistic about the future to increase their consumption by more than the increase in their disposable income.

The next question is whether this difference in the adjustment of expectations over the two episodes can be attributed fully to the differences in the two fiscal programs. The answer is surely no. Ireland was changing in many ways at the time of the second fiscal program. Productivity was increasing much faster than real wages, reducing the cost of labor for firms. Attracted by tax breaks, low labor costs, and an educated labor force, many foreign firms were relocating to Ireland and building new plants: These factors played a major role in the expansion of the late 1980s. Irish growth was then very strong, usually more than 5% per year from 1990 to the time of the crisis in 2007. Surely, this long expansion is due to many factors. Nevertheless, the change in fiscal policy in 1987 probably played an important role in convincing people, firms—including foreign firms—and financial markets, that the government was regaining control of its finances. And the fact remains that the substantial deficit reduction of 1987–1989 was accompanied by a strong output expansion, not by the recession predicted by the basic IS–LM model.


For a more systematic look at whether and when fiscal consolidations have been expansionary (and a mostly negative answer), see “Will It Hurt? Macroeconomic Effects of Fiscal Consolidation,” Chapter 3, World Economic Outlook, International Monetary Fund, October 2010.

in expected future government spending \( (G^e) \), the larger the effect on expected future output and interest rates, thus the larger the favorable effect on spending today. This suggests that backloading the deficit reduction program toward the future, with small cuts today and larger cuts in the future is more likely to lead to an increase in output.

■ On the other hand, backloading raises other issues. Announcing the need for painful cuts in spending, and then leaving them to the future, is likely to decrease the program’s credibility—the perceived probability that the government will do what it has promised when the time comes to do it.

■ The government must play a delicate balancing act: enough cuts in the current period to show a commitment to deficit reduction; enough cuts left to the future to reduce the adverse effects on the economy in the short-run.

More generally, our analysis suggests that anything in a deficit reduction program that improves expectations of how the future will look is likely to make the short-run effects of deficit reduction less painful. We will give you two examples.

■ Measures that are perceived by firms and financial markets as reducing some of the distortions in the economy may improve expectations, and make it more likely that output increases in the short run. Take for example unemployment benefits. You saw in Chapter 6 that lower unemployment benefits lead to a decline in the natural rate of unemployment, resulting in a higher natural level of output. So, a reform of the social insurance system, which includes a reduction in the generosity of unemployment benefits, is likely to have two effects on spending and thus on output in the short run:

One is an adverse effect on the consumption of the unemployed: Lower unemployment benefits will reduce their income and their consumption. The other is a positive effect on spending through expectations: The anticipation of higher output in the future may lead to both higher consumption and higher investment.
If the second effect dominates, the outcome might be an increase in overall spending, increasing output not only in the medium run but also in the short run. (An important caveat: Even if a reduction in unemployment benefits increases output, this surely does not imply that unemployment benefits should be eliminated. Even if aggregate income goes up, we must worry about the effects on the distribution of income: The consumption of the unemployed goes down, and the pain associated with being unemployed goes up.)

- Or take an economy where the government has, in effect, lost control of its budget: Government spending is high, tax revenues are low, and the deficit is very large. In such an environment, a credible deficit reduction program is also more likely to increase output in the short run. Before the announcement of the program, people may have expected major political and economic troubles in the future. The announcement of a program of deficit reduction may well reassure them that the government has regained control, and that the future is less bleak than they anticipated. This decrease in pessimism about the future may lead to an increase in spending and output, even if taxes are increased as part of the deficit reduction program.

Let’s summarize:
A program of deficit reduction may increase output even in the short run. Whether it does or does not depends on many factors, in particular:

- The credibility of the program: Will spending be cut or taxes increased in the future as announced?
- The timing of the program: How large are spending cuts in the future relative to current spending cuts?
- The composition of the program: Does the program remove some of the distortions in the economy?
- The state of government finances in the first place: How large is the initial deficit? Is this a “last chance” program? What will happen if it fails?

This gives you a sense of both the importance of expectations in determining the outcome, and of the complexities involved in the use of fiscal policy in such a context. And it is far more than an illustrative example. At the time of this writing, it is at the center of macroeconomic policy discussions. As we have seen, the crisis has led to large increases in government debt and large budget deficits throughout advanced countries. Nearly all governments must now embark on a path of fiscal consolidation—of budget deficit reduction. At what rate should they proceed, and whether they can hope that fiscal consolidation, if done right, will be expansionary rather than contractionary, are crucial issues. The analysis we have just gone through suggests the following conclusion:

- There should be a clear, politically credible plan for how fiscal consolidation will be carried out. The more credible the plan, the more people are convinced that fiscal consolidation will be effectively carried out in the medium and long terms, the smaller the need for fiscal consolidation today, and thus the smaller the direct adverse effects of deficit reduction today.
- In most cases, governments should not expect miracles, and fiscal consolidation is likely to lead to lower growth in the short run. In a few cases, however, where financial markets are very worried about the fiscal situation, and where, as a result, the interest rate on government debt is already very high, such as in Italy and Spain, it may be that the decrease in the interest rate that would follow from a credible fiscal consolidation may offset the adverse direct effects.
Spending in the goods market depends on current and expected future output and real interest rates. Expectations affect demand and, in turn, affect output: Changes in expected future output or the expected future real interest rate lead to changes in spending and output today. By implication, the effects of fiscal and monetary policy on spending and output depend on how the policy affects expectations of future output and real interest rates. Rational expectations is the assumption that people, firms, and participants in financial markets form expectations of the future by assessing the course of future expected policy and then working out the implications for future output, future interest rates, and so on. Although it is clear that most people do not go through this exercise themselves, we can think of them as doing so indirectly by relying on the predictions of public and private forecasters.

Although there are surely cases in which people, firms, or financial investors do not have rational expectations, the assumption of rational expectations seems to be the best benchmark to evaluate the potential effects of alternative policies. Designing a policy on the assumption that people will make systematic mistakes in responding to it would be unwise. Changes in the money supply affect the short-term nominal interest rate. Spending, however, depends instead on current and expected future real interest rates. Thus, the effect of monetary policy on activity depends crucially on whether and how changes in the short-term nominal interest rate lead to changes in current and expected future real interest rates. A budget deficit reduction may lead to an increase rather than a decrease in output. This is because expectations of higher output and lower interest rates in the future may lead to an increase in spending that more than offsets the reduction in spending coming from the direct effect of the deficit reduction on total spending.

Aggregate private spending, or private spending, 359
Fire sales, 365
Quantitative Easing I and II (QE I, QE II), 365
Rational expectations, 366
Animal spirits, 367
Adaptive expectations, 367
Backloading, 371
Credibility, 371

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. Changes in the current one-year real interest rate are likely to have a much larger effect on spending than changes in expected future one-year real interest rates.
   b. The introduction of expectations in the goods market model makes the IS curve flatter, although it is still downward sloping.
   c. Current money demand depends on current and expected future nominal interest rates.
   d. The rational expectations assumption implies that consumers take into account the effects of future fiscal policy on output.
   e. Expected future fiscal policy affects expected future economic activity but not current economic activity.
   f. Depending on its effect on expectations, a fiscal contraction may actually lead to an economic expansion.
   g. Ireland’s experience with deficit reduction programs in 1982 and 1987 provides strong evidence against the hypothesis that deficit reduction can lead to an output expansion.

2. During the late 1990s, many observers claimed that the United States had transformed into a New Economy, and this justified the very high values for stock prices observed at the time.
   a. Discuss how the belief in the New Economy, combined with the increase in stock prices, affected consumption spending.
   b. Stock prices subsequently decreased. Discuss how this might have affected consumption.

3. For each of the changes in expectations in (a) through (d), determine whether there is a shift in the IS curve, the LM curve, both curves, or neither. In each case, assume that expected current and future inflation are equal to zero and that no other exogenous variable is changing.
   a. a decrease in the expected future real interest rate.
   b. an increase in the current money supply.
   c. an increase in expected future taxes.
   d. a decrease in expected future income.

4. Consider the following statement. “The rational expectations assumption is unrealistic because, essentially, it amounts to the assumption that every consumer has perfect knowledge of the economy.” Discuss.
5. A new president, who promised during the campaign that she would cut taxes, has just been elected. People trust that she will keep her promise, but expect that the tax cuts will be implemented only in the future. Determine the impact of the election on current output, the current interest rate, and current private spending. Suppose, as in part (a), that the President nominates a replacement who is expected to conduct a more contractionary monetary policy than the current Fed chair.

b. Suppose financial market participants are not surprised by the President’s choice. In other words, market participants had correctly predicted who the President would choose as nominee. Under these circumstances, is the announcement of the nominee likely to have any effect on the yield curve?

c. Suppose instead that the identity of the nominee is a surprise and that financial market participants had expected the nominee to be someone who favored an even more contractionary policy than the actual nominee. Under these circumstances, what is likely to happen to the yield curve on the day of the announcement? (Hint: Be careful. Compared to what was expected, is the actual nominee expected to follow a more contractionary or more expansionary policy?)

d. On October 24, 2005, Ben Bernanke was nominated to succeed Alan Greenspan as chairman of the Federal Reserve. Do an internet search and try to learn what happened in financial markets on the day the nomination was announced. Were financial market participants surprised by the choice? If so, was Bernanke believed to favor policies that would lead to higher or lower interest rates (as compared to the expected nominee) over the next three to five years? (You may also do a yield curve analysis of the kind described in Problem 8 for the period around Bernanke’s nomination. If you do this, use one-year and five-year interest rates.)

EXPLORE FURTHER

8. Deficits and interest rates

Go back and look again at Figure 1-4. There was a dramatic change in the U.S. budget position after 2000 (from a surplus to a large and continuing deficit). This change took place well before the crisis and the election of President Obama. The change reinvigorated the debate about the effect of fiscal policy on interest rates. This problem asks you to review theory and evidence on this topic.

a. Review what theory predicts about fiscal policy and interest rates. Suppose there is an increase in government spending and a decrease in taxes. Use an IS–LM diagram to show what will happen to the nominal interest rate in the short run and the medium run. Assuming that there is no change in monetary policy, what does the IS–LM model predict will happen to the yield curve immediately after an increase in government spending and a decrease in taxes?

During the first term of the G. W. Bush administration, the actual and projected federal budget deficits increased dramatically. Part of the increase in the deficit can be attributed to the recession of 2001. However, deficits and projected deficits continued to increase even after the recession had ended.
b. Go to the web site of the Federal Reserve Bank of St. Louis (research.stlouisfed.org/fred2/). Under “Interest Rates” and then “Treasury Constant Maturity,” obtain the data for “3-Month Constant Maturity Treasury Yield” and “5-Year Constant Maturity Treasury Yield” for each of the months in the table shown here. For each month, subtract the three-month yield from the five-year yield to obtain the interest rate spread. What happened to the interest rate spread as the budget picture worsened over the sample period? Is this result consistent with your answer to part (a)?

The analysis you carried out in this problem is an extension of work by William C. Gale and Peter R. Orszag. See “The Economic Effects of Long-Term Fiscal Discipline,” Brookings Institution, December 17, 2002. Figure 5 in this paper relates interest rate spreads to CBO five-year projected budget deficits from 1982 to 2002.

The following table provides budget projections produced by the Congressional Budget Office (CBO) over the period August 2002 to January 2004. These projections are for the total federal budget deficit, so they include Social Security, which was running a surplus over the period. In addition, each projection assumes that current policy (as of the date of the forecast) continues into the future.

<table>
<thead>
<tr>
<th>Date of Forecast</th>
<th>Projected Five-Year Deficit (as a % of GDP, negative number indicates a deficit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>August 2002</td>
<td>−0.4</td>
</tr>
<tr>
<td>January 2003</td>
<td>−0.2</td>
</tr>
<tr>
<td>August 2003</td>
<td>−2.3</td>
</tr>
<tr>
<td>January 2004</td>
<td>−2.3</td>
</tr>
</tbody>
</table>
The Open Economy

The next four chapters cover the second extension of the core. They look at the implications of openness—the fact that most economies trade both goods and assets with the rest of the world.

Chapter 18

Chapter 18 discusses the implications of openness in goods markets and financial markets. Openness in goods markets allows people to choose between domestic goods and foreign goods. An important determinant of their decisions is the real exchange rate—the relative price of domestic goods in terms of foreign goods. Openness in financial markets allows people to choose between domestic assets and foreign assets. This imposes a tight relation between the exchange rate, both current and expected, and domestic and foreign interest rates—a relation known as the interest parity condition.

Chapter 19

Chapter 19 focuses on equilibrium in the goods market in an open economy. It shows how the demand for domestic goods now depends on the real exchange rate. It shows how fiscal policy affects both output and the trade balance. It discusses the conditions under which a real depreciation improves the trade balance, and increases output.
Chapter 20

Chapter 20 characterizes goods and financial markets’ equilibrium in an open economy. In other words, it gives an open economy version of the IS–LM model we saw in the core. It shows how, under flexible exchange rates, monetary policy affects output not only through its effect on the interest rate but also through its effect on the exchange rate. It shows how fixing the exchange rate also implies giving up the ability to change the interest rate.

Chapter 21

Chapter 21 looks at the properties of different exchange rate regimes. It first shows how, in the medium run, the real exchange rate can adjust even under a fixed exchange rate regime. It then looks at exchange rate crises under fixed exchange rates, and at movements in exchange rates under flexible exchange rates. It ends by discussing the pros and cons of various exchange rate regimes, including the adoption of a common currency such as the euro.
We have assumed until now that the economy we looked at was closed—that it did not interact with the rest of the world. We had to start this way, to keep things simple and to build up intuition for the basic macroeconomic mechanisms. Figure 18-1 shows how bad, in fact, that this assumption is. The figure plots the growth rates for advanced and emerging economies since 2005. What is striking is how the growth rates have moved together: Despite the fact that the crisis originated in the United States, the outcome was a worldwide recession, with negative growth both in advanced and in emerging economies. It is therefore time to relax this assumption. Understanding the macroeconomic implications of openness will occupy us for this and the next three chapters.

“Openness” has three distinct dimensions:

1. **Openness in goods markets**—the ability of consumers and firms to choose between domestic goods and foreign goods. In no country is this choice completely free of restrictions: Even the countries most committed to free trade have **tariffs**—taxes on imported goods—and **quotas**—restrictions on the quantity of goods that can be imported—on at least some foreign goods. At the same time, in most countries, average tariffs are low and getting lower.

2. **Openness in financial markets**—the ability of financial investors to choose between domestic assets and foreign assets. Until recently even some of the richest countries in the world, such as France and Italy, had **capital controls**—restrictions on the foreign assets their domestic residents could hold and the domestic assets foreigners could hold. These restrictions have largely disappeared. As a result, world financial markets are becoming more and more closely integrated.

3. **Openness in factor markets**—the ability of firms to choose where to locate production, and of workers to choose where to work. Here also trends are clear. Multinational companies operate plants in many countries and move their operations around the world to take advantage of low costs. Much of the debate about the **North American Free Trade Agreement (NAFTA)** signed in 1993 by the United States, Canada, and Mexico centered on how it would affect the relocation of U.S. firms to Mexico. Similar fears now center around China. And immigration from low-wage countries is a hot political issue in countries from Germany to the United States.

In the short run and in the medium run—the focus of this and the next three chapters—openness in factor markets plays much less of a role than openness in either goods markets or financial markets. Thus, we shall ignore openness in factor markets and focus on the implications of the first two dimensions of openness here.
Section 18-1 looks at openness in the goods market, the determinants of the choice between domestic goods and foreign goods, and the role of the real exchange rate.

Section 18-2 looks at openness in financial markets, the determinants of the choice between domestic assets and foreign assets, and the role of interest rates and exchange rates.

Section 18-3 gives a map to the next three chapters.

18-1 Openness in Goods Markets

Let’s start by looking at how much the United States sells to and buys from the rest of the world. Then, we shall be better able to think about the choice between domestic goods and foreign goods, and the role of the relative price of domestic goods in terms of foreign goods—the real exchange rate.

Exports and Imports

Figure 18-2 plots the evolution of U.S. exports and U.S. imports, as ratios to GDP, since 1960 (“U.S. exports” means exports from the United States; “U.S. imports” means imports to the United States). The figure suggests two main conclusions.

- The U.S. economy is becoming more open over time. Exports and imports, which were equal to 5% of GDP in the early 1960s, are now equal to about 14.5% of GDP (13% for exports, 16% for imports). In other words, the United States trades more than twice as much (relative to its GDP) with the rest of the world than it did 50 years ago.
- Although imports and exports have followed the same upward trend, since the early 1980s imports have consistently exceeded exports. Put another way, for the last 30 years, the United States has consistently run a trade deficit. For four years in a row in the mid-2000s, the ratio of the trade deficit to GDP exceeded 5% of GDP. While it has decreased since the beginning of the crisis, it remains large today.
Understanding the sources and implications of this large deficit is a central issue in macroeconomics today and one to which we shall return later.

Given all the talk in the media about globalization, a volume of trade (measured by the average of the ratios of exports and imports to GDP) around 14.4% of GDP might strike you as small. However, the volume of trade is not necessarily a good measure of openness. Many firms are exposed to foreign competition but, by being competitive and keeping their prices low enough, these firms are able to retain their domestic market share and limit imports. This suggests that a better index of openness than export or import ratios is the proportion of aggregate output composed of tradable goods—goods that compete with foreign goods in either domestic markets or foreign markets. Estimates are that tradable goods represent about 60% of aggregate output in the United States today.

With exports around 13% of GDP, it is true that the United States has one of the smallest ratios of exports to GDP among the rich countries of the world. Table 18-1 gives ratios for a number of OECD countries.

The United States is at the low end of the range of export ratios. Japan’s ratio is about the same, the United Kingdom’s twice as large, and Germany’s three times as large. And the smaller European countries have very large ratios, from 54% in Switzerland to 81% in Belgium. (Belgium’s 81% ratio of exports to GDP raises an odd possibility: Can a country have exports larger than its GDP; in other words, can a country have an export ratio greater than one? The answer is: yes. The reason why is given in the Focus box “Can Exports Exceed GDP?”)

![Figure 18-2](image_url)

**U.S. Exports and Imports as Ratios of GDP since 1960**

Since 1960, exports and imports have more than doubled in relation to GDP. The United States has become a much more open economy.

Source: Series GDP, EXPGS, IMPGS. Federal Reserve Economic Data (FRED) http://research.stlouis-fed.org/fred2/

<table>
<thead>
<tr>
<th>Table 18-1</th>
<th>Ratios of Exports to GDP for Selected OECD Countries, 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
<td><strong>Export Ratio</strong></td>
</tr>
<tr>
<td>United States</td>
<td>13%</td>
</tr>
<tr>
<td>Japan</td>
<td>15%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>30%</td>
</tr>
<tr>
<td>Germany</td>
<td>46%</td>
</tr>
</tbody>
</table>

Source: OECD Economic Outlook Database

Tradable goods: cars, computers, etc. Nontradable goods: housing, most medical services, haircuts, etc.

For more on the OECD and for the list of member countries, see Chapter 1.
The Open Economy

Extensions

Do these numbers indicate that the United States has more trade barriers than, say, the United Kingdom or Belgium? No. The main factors behind these differences are geography and size. Distance from other markets explains a part of the lower Japanese ratio. Size also matters: The smaller the country, the more it must specialize in producing and exporting only a few products and rely on imports for the other products. Belgium can hardly afford to produce the range of goods produced by the United States, a country roughly 40 times its economic size.

The Choice between Domestic Goods and Foreign Goods

How does openness in goods markets force us to rethink the way we look at equilibrium in the goods market?

Until now, when we were thinking about consumers’ decisions in the goods market, we focused on their decision to save or to consume. When goods markets are open, domestic consumers face a second decision: whether to buy domestic goods or to buy foreign goods. Indeed, all buyers—including domestic and foreign firms and governments—face the same decision. This decision has a direct effect on domestic output: If buyers decide to buy more domestic goods, the demand for domestic goods increases, and so does domestic output. If they decide to buy more foreign goods, then foreign output increases instead of domestic output.

Central to this second decision (to buy domestic goods or foreign goods) is the price of domestic goods relative to foreign goods. We call this relative price the real exchange rate. The real exchange rate is not directly observable, and you will not find it in the newspapers. What you will find in newspapers are nominal exchange rates, the relative prices of currencies. So we start by looking at nominal exchange rates and then see how we can use them to construct real exchange rates.

Nominal Exchange Rates

Nominal exchange rates between two currencies can be quoted in one of two ways:

- As the price of the domestic currency in terms of the foreign currency. If, for example, we look at the United States and the United Kingdom, and think of the dollar as the domestic currency and the pound as the foreign currency, we can express the nominal exchange rate as the price of a dollar in terms of pounds. In September 2011, the exchange rate defined this way was 0.61. In other words, one dollar was worth 0.61 pounds.
As the price of the foreign currency in terms of the domestic currency. Continuing with the same example, we can express the nominal exchange rate as the price of a pound in terms of dollars. In September 2011, the exchange rate defined this way was 1.63. In other words, one pound was worth 1.63 dollars.

Either definition is fine; the important thing is to remain consistent. In this book, we shall adopt the first definition: we shall define the nominal exchange rate as the price of the domestic currency in terms of foreign currency, and denote it by $E$. When looking, for example, at the exchange rate between the United States and the United Kingdom (from the viewpoint of the United States, so the dollar is the domestic currency), $E$ will denote the price of a dollar in terms of pounds (so, for example, $E = 0.61$ in September 2011).

Exchange rates between the dollar and most foreign currencies change every day—indeed every minute of the day. These changes are called nominal appreciations or nominal depreciations—appreciations or depreciations for short.

An appreciation of the domestic currency is an increase in the price of the domestic currency in terms of a foreign currency. Given our definition of the exchange rate, an appreciation corresponds to an increase in the exchange rate.

A depreciation of the domestic currency is a decrease in the price of the domestic currency in terms of a foreign currency. So, given our definition of the exchange rate, a depreciation of the domestic currency corresponds to a decrease in the exchange rate, $E$.

You may have encountered two other words to denote movements in exchange rates: “revaluations” and “devaluations.” These two terms are used when countries operate under fixed exchange rates—a system in which two or more countries maintain a constant exchange rate between their currencies. Under such a system, increases in the exchange rate—which are infrequent by definition—are called revaluations (rather than appreciations). Decreases in the exchange rate are called devaluations (rather than depreciations).

Figure 18-3 plots the nominal exchange rate between the dollar and the pound since 1970. Note the two main characteristics of the figure:

The trend increase in the exchange rate. In 1970, a dollar was worth only 0.41 pounds. In 2011, a dollar was worth 0.61 pounds. Put another way, there was an appreciation of the dollar relative to the pound over the period.

The large fluctuations in the exchange rate. In the 1980s, a sharp appreciation, in which the dollar more than doubled in value relative to the pound, was followed by a nearly equally sharp depreciation. In the 2000s, a large depreciation was followed by a large appreciation as the crisis started, and a smaller depreciation since then.

If we are interested, however, in the choice between domestic goods and foreign goods, the nominal exchange rate gives us only part of the information we need. Figure 18-3, for example, tells us only about movements in the relative price of the two currencies, the dollar and the pound. To U.S. tourists thinking of visiting the United Kingdom, the question is not only how many pounds they will get in exchange for their dollars but how much goods will cost in the United Kingdom relative to how much they cost in the United States. This takes us to our next step—the construction of real exchange rates.

From Nominal to Real Exchange Rates

How can we construct the real exchange rate between the United States and the United Kingdom—the price of U.S. goods in terms of British goods?

Suppose the United States produced only one good, a Cadillac luxury sedan, and the United Kingdom also produced only one good, a Jaguar luxury sedan. (This is one of those “Suppose” statements that run completely against the facts, but we shall become
more realistic shortly.) Constructing the real exchange rate, the price of the U.S. goods (Cadillacs) in terms of British goods (Jaguars) would be straightforward. We would express both goods in terms of the same currency and then compute their relative price.

Suppose, for example, we expressed both goods in terms of pounds. Then

- The first step would be to take the price of a Cadillac in dollars and convert it to a price in pounds. The price of a Cadillac in the United States is $40,000. The dollar is worth, say, 0.60 pounds, so the price of a Cadillac in pounds is 40,000 dollars times 0.60 = £24,000.
- The second step would be to compute the ratio of the price of the Cadillac in pounds to the price of the Jaguar in pounds. The price of a Jaguar in the United Kingdom is £30,000. So the price of a Cadillac in terms of Jaguars—that is, the real exchange rate between the United States and the United Kingdom—would be £24,000/£30,000 = 0.80. A Cadillac would be 20% cheaper than a Jaguar.

This example is straightforward, but how do we generalize it? The United States and the United Kingdom produce more than Cadillacs and Jaguars, and we want to construct a real exchange rate that reflects the relative price of all the goods produced in the United States in terms of all the goods produced in the United Kingdom.

The computation we just went through tells us how to proceed. Rather than using the price of a Jaguar and the price of a Cadillac, we must use a price index for all goods produced in the United Kingdom and a price index for all goods produced in the United States. This is exactly what the GDP deflators we introduced in Chapter 2 do: They are, by definition, price indexes for the set of final goods and services produced in the economy.

Let \( P \) be the GDP deflator for the United States, \( P^* \) be the GDP deflator for the United Kingdom (as a rule, we shall denote foreign variables by a star), and \( E \) be the dollar–pound nominal exchange rate. Figure 18-4 goes through the steps needed to construct the real exchange rate.
The price of U.S. goods in dollars is $P$. Multiplying it by the exchange rate, $E$—the price of dollars in terms of pounds—gives us the price of U.S. goods in pounds, $EP$.

The price of British goods in pounds is $P^*$. The real exchange rate, the price of U.S. goods in terms of British goods, which we shall call $\epsilon$ (the Greek lowercase epsilon), is thus given by

$$\epsilon = \frac{EP}{P^*}$$  \hspace{1cm} (18.1)

The real exchange rate is constructed by multiplying the domestic price level by the nominal exchange rate and then dividing by the foreign price level—a straightforward extension of the computation we made in our Cadillac/Jaguar example.

Note, however, an important difference between our Cadillac/Jaguar example and this more general computation:

Unlike the price of Cadillacs in terms of Jaguars, the real exchange rate is an index number: that is, its level is arbitrary, and therefore uninformative. It is uninformative because the GDP deflators used to construct the real exchange rate are themselves index numbers; as we saw in Chapter 2, they are equal to 1 (or 100) in whatever year is chosen as the base year.

But all is not lost. Although the level of the real exchange rate is uninformative, the rate of change of the real exchange rate is informative: If, for example, the real exchange rate between the United States and the United Kingdom increases by 10%, this tells us U.S. goods are now 10% more expensive relative to British goods than they were before.

Like nominal exchange rates, real exchange rates move over time. These changes are called real appreciations or real depreciations:

- An increase in the real exchange rate—that is, an increase in the relative price of domestic goods in terms of foreign goods—is called a real appreciation.
- A decrease in the real exchange rate—that is, a decrease in the relative price of domestic goods in terms of foreign goods—is called a real depreciation.

Figure 18-5 plots the evolution of the real exchange rate between the United States and the United Kingdom since 1970, constructed using equation (18.1). For convenience, it also reproduces the evolution of the nominal exchange rate from Figure 18-3. The GDP deflators have both been set equal to 1 in the year 2000, so the nominal exchange rate and the real exchange rate are equal in that year by construction.

You should take two lessons from Figure 18-5:

- The nominal and the real exchange rate can move in opposite directions. Note how, from 1970 to 1980, while the nominal exchange rate went up, the real exchange rate actually went down.
How do we reconcile the fact that there was both a nominal appreciation (of the dollar relative to the pound) and a real depreciation (of U.S. goods relative to British goods) during the period? To see why, return to the definition of the real exchange rate, a slightly rewritten version of (18.1):

$$
\epsilon = \frac{E}{P^*}
$$

Two things happened in the 1971s:

First, $E$ increased: The dollar went up in terms of pounds—this is the nominal appreciation we saw earlier.

Second, $P/P^*$ decreased. The price level increased less in the United States than in the United Kingdom. Put another way, over the period, average inflation was lower in the United States than in the United Kingdom.

The resulting decrease in $P/P^*$ was larger than the increase in $E$, leading to a decrease in $\epsilon$, a real depreciation—a decrease in the relative price of domestic goods in terms of foreign goods.

To get a better understanding of what happened, let’s go back to our U.S tourists thinking about visiting the United Kingdom, circa 1980. They would find that they could buy more pounds per dollar than in 1971 ($E$ had increased). Did this imply their trip would be cheaper? No: When they arrived in the United Kingdom, they would discover that the prices of goods in the United Kingdom had increased much more than the prices of goods in the United States ($P^*$ has increased more than $P$, so $P/P^*$ has declined), and this more than canceled the increase in the value of the dollar in terms of pounds. They would find that their trip was actually more expensive (in terms of U.S. goods) than it would have been 10 years earlier.

There is a general lesson here. Over long periods of time, differences in inflation rates across countries can lead to very different movements in nominal exchange rates and real exchange rates. We shall return to this issue in Chapter 21.

Can there be a real appreciation with no nominal appreciation?
Can there be a nominal appreciation with no real appreciation? (The answers to both questions: yes.)
The large fluctuations in the nominal exchange rate we saw in Figure 18-3 also show up in the real exchange rate.

This not surprising: Year-to-year movements in the price ratio $P/P^*$ are typically small compared to the often sharp movements in the nominal exchange rate $E$. Thus, from year to year, or even over the course of a few years, movements in the real exchange rate ($\epsilon$) tend to be driven mostly by movements in the nominal exchange rate $E$. Note that, since the early 1990s, the nominal exchange rate and the real exchange rate have moved nearly together. This reflects the fact that, since the early 1990s, inflation rates have been very similar—and low—in both countries.

### From Bilateral to Multilateral Exchange Rates

We need to take one last step. We have so far concentrated on the exchange rate between the United States and the United Kingdom. But the United Kingdom is just one of many countries the United States trades with. Table 18-2 shows the geographic composition of U.S. trade for both exports and imports.

The main message of the table is that the United States does most of its trade with three sets of countries. The first includes its neighbors to the North and to the South, Canada and Mexico: Trade with Canada and Mexico accounts for 26% of U.S. exports and 24% of U.S. imports. The second includes the countries of Western Europe, which account for 23% of U.S. exports and 20% of U.S. imports. The third includes the Asian countries, including Japan and China, which together account for 29% of U.S. exports and 38% of U.S. imports.

How do we go from bilateral exchange rates, like the real exchange rate between the United States and the United Kingdom we focused on earlier, to multilateral exchange rates that reflect this composition of trade? The principle we want to use is simple, even if the details of construction are complicated. We want the weight of a given country to incorporate not only how much the country trades with the United States but also how much it competes with the United States in other countries. (Why not just look at trade shares between the United States and each individual country? Take two countries, the United States and country A. Suppose the United States and country A do not trade with each other—so trade shares are equal to zero—but they are both exporting to another country, call it country B. The real exchange rate between the United States and country A will matter very much for how much the United States exports to country B and thus to the U.S. export performance.) The variable constructed in this way is called the multilateral real U.S. exchange rate, or the U.S. real exchange rate for short.

<table>
<thead>
<tr>
<th>Table 18-2 The Country Composition of U.S. Exports and Imports, 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percent of Exports to</strong></td>
</tr>
<tr>
<td>Canada</td>
</tr>
<tr>
<td>Mexico</td>
</tr>
<tr>
<td>European Union</td>
</tr>
<tr>
<td>China</td>
</tr>
<tr>
<td>Japan</td>
</tr>
<tr>
<td>Rest of Asia and Pacific</td>
</tr>
<tr>
<td>Others</td>
</tr>
</tbody>
</table>

Source: Survey of Current Business, August 2011, Tables F2 and F3
Figure 18-6 shows the evolution of this multilateral real exchange rate, the price of U.S. goods in terms of foreign goods, since 1973. Like the bilateral real exchange rates we saw a few pages earlier, it is an index number and its level is arbitrary. You should note two things about Figure 18-6: first, a trend real depreciation since 1973 (in contrast to the trend nominal appreciation vis à vis the pound in Figure 18-3); second, and more strikingly, the large swings in the multilateral real exchange rate in the 1980s and, to a lesser extent, in the 2000s. These swings are so striking that they have been given various names, from the “dollar cycle” to the more graphic “dance of the dollar.” In the coming chapters, we shall examine where these swings come from and their effects on the trade deficit and economic activity.

18-2 Openness in Financial Markets

Openness in financial markets allows financial investors to hold both domestic assets and foreign assets, to diversify their portfolios, to speculate on movements in foreign interest rates versus domestic interest rates, on movements in exchange rates, and so on.

Diversify and speculate they do. Given that buying or selling foreign assets implies buying or selling foreign currency—sometimes called foreign exchange—the volume of transactions in foreign exchange markets gives us a sense of the importance of international financial transactions. In 2010, for example, the recorded daily volume of foreign exchange transactions in the world was $4 trillion, of which 85%—about $3.4 trillion—involved U.S. dollars on one side of the transaction.

To get a sense of the magnitude of these numbers, the sum of U.S. exports and imports in 2010 totaled $4.1 trillion for the year, or about $11 billion a day. Suppose the only dollar transactions in foreign-exchange markets had been, on one side, by U.S.
exporters selling their foreign currency earnings, and on the other side by U.S. importers buying the foreign currency they needed to buy foreign goods. Then the volume of transactions involving dollars in foreign exchange markets would have been $11 billion a day, or about 0.3% of the actual daily total volume of dollar transactions ($3.40 trillion) involving dollars in foreign exchange markets. This computation tells us that most of the transactions are associated not with trade, but with purchases and sales of financial assets. Moreover, the volume of transactions in foreign exchange markets is not only high but also rapidly increasing. The volume of foreign exchange transactions has more than quadrupled since 2001. Again, this increase in activity reflects mostly an increase in financial transactions rather than an increase in trade.

For a country as a whole, openness in financial markets has another important implication: It allows the country to run trade surpluses and trade deficits. Recall that a country running a trade deficit is buying more from the rest of the world than it is selling to the rest of the world. In order to pay for the difference between what it buys and what it sells, the country must borrow from the rest of the world. It borrows by making it attractive for foreign financial investors to increase their holdings of domestic assets—in effect, to lend to the country.

Let’s start by looking more closely at the relation between trade flows and financial flows. When this is done, we shall then look at the determinants of these financial flows.

The Balance of Payments

A country’s transactions with the rest of the world, including both trade flows and financial flows, are summarized by a set of accounts called the balance of payments. Table 18-3 presents the U.S. balance of payments for 2010. The table has two parts, separated by a line. Transactions are referred to either as above the line or below the line.

The Current Account

The transactions above the line record payments to and from the rest of the world. They are called current account transactions.

Table 18-3 The U.S. Balance of Payments, 2010, in Billions of U.S. Dollars

<table>
<thead>
<tr>
<th>Current Account</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exports</td>
<td>1838</td>
</tr>
<tr>
<td>Imports</td>
<td>2338</td>
</tr>
<tr>
<td>Trade balance (deficit = −) (1)</td>
<td>−500</td>
</tr>
<tr>
<td>Income received</td>
<td>663</td>
</tr>
<tr>
<td>Income paid</td>
<td>498</td>
</tr>
<tr>
<td>Net income (2)</td>
<td></td>
</tr>
<tr>
<td>Net transfers received (3)</td>
<td>−136</td>
</tr>
<tr>
<td>Current account balance (deficit = −) (1) + (2) + (3)</td>
<td>−471</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capital Account</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in foreign holdings of U.S. assets (4)</td>
<td>1260</td>
</tr>
<tr>
<td>Increase in U.S. holdings of foreign assets (5)</td>
<td>1005</td>
</tr>
<tr>
<td>Capital account balance (deficit = −) (4) − (5)</td>
<td>255</td>
</tr>
<tr>
<td>Statistical discrepancy</td>
<td>216</td>
</tr>
</tbody>
</table>

Source: Survey of Current Business, August 2011, Table F2
The difference between exports and imports is the trade balance. In 2010, imports exceeded exports, leading to a U.S. trade deficit of $500 billion—roughly 3.4% of U.S. GDP.

Exports and imports are not the only sources of payments to and from the rest of the world. U.S. residents receive income on their holdings of foreign assets, and foreign residents receive income on their holdings of U.S. assets. In 2010, income received from the rest of the world was $663 billion, and income paid to foreigners was $498 billion, for a net income balance of $165 billion.

Finally, countries give and receive foreign aid; the net value of these payments is recorded as net transfers received. These net transfers amounted in 2010 to $136 billion. This negative amount reflects the fact that, in 2010, the United States was—as it has traditionally been—a net donor of foreign aid.

The sum of net payments to and from the rest of the world is called the current account balance. If net payments from the rest of the world are positive, the country is running a current account surplus; if they are negative, the country is running a current account deficit. Adding all payments to and from the rest of the world, net payments from the United States to the rest of the world were equal in 2010 to $500 + $165 - $136 = $471 billion. Put another way, in 2010, the United States ran a current account deficit of $471 billion—roughly 3.2% of its GDP.

The Capital Account

The fact that the United States had a current account deficit of $471 billion in 2010 implies that it had to borrow $471 billion from the rest of the world—or, equivalently, that net foreign holdings of U.S. assets had to increase by $471 billion. The numbers below the line describe how this was achieved. Transactions below the line are called capital account transactions.

The increase in foreign holdings of U.S. assets was $1,260 billion: Foreign investors, be they foreign private investors, foreign governments, or foreign central banks, bought $1,260 billion worth of U.S. stocks, U.S. bonds, and other U.S. assets. At the same time, there was an increase in U.S. holdings of foreign assets of $1,005 billion: U.S. investors, private and public, bought $1,005 billion worth of foreign stocks, bonds, and other assets. The result was an increase in net U.S. foreign indebtedness (the increase in foreign holdings of U.S. assets, minus the increase in U.S. holdings of foreign assets), also called net capital flows to the United States, of $1260 - $1005 = $255 billion. Another name for net capital flows is the capital account balance: Positive net capital flows are called a capital account surplus; negative net capital flows are called a capital account deficit. So, put another way, in 2010, the United States ran a capital account surplus of $255 billion.

Shouldn’t net capital flows (equivalently, the capital account surplus) be exactly equal to the current account deficit (which we saw earlier was equal to $471 billion in 2010)?

In principle, yes. In practice, no.

The numbers for current and capital account transactions are constructed using different sources; although they should give the same answers, they typically do not. In 2010, the difference between the two—called the statistical discrepancy—was $216 billion, about 45% of the current account balance. This is yet another reminder that, even for a rich country such as the United States, economic data are far from perfect. (This problem of measurement manifests itself in another way as well. The sum of the current account deficits of all the countries in the world should be equal to zero: One country’s deficit should show up as a surplus for the other countries taken as a whole. However, this is not the case in the data: If we just add the published current account...
deficits of all the countries in the world, it would appear that the world is running a large current account deficit!)

Now that we have looked at the current account, we can return to an issue we touched on in Chapter 2, the difference between GDP, the measure of output we have used so far, and GNP, another measure of aggregate output.

GDP measures value added domestically. GNP measures the value added by domestic factors of production. When the economy is closed, the two measures are the same. When the economy is open, however, they can differ: Some of the income from domestic production goes to foreigners; and domestic residents receive some foreign income. Thus, to go from GDP to GNP, one must start from GDP, add income received from the rest of the world, and subtract income paid to the rest of the world. Put another way, GNP is equal to GDP plus net payments from the rest of the world. More formally, denoting these net income payments by NI,

$$ GNP = GDP + NI $$

In most countries, the difference between the GNP and GDP is small (relative to GDP). For example, in the United States, you can see from Table 18-3 that net income payments were equal to $165 billion: GNP exceeded GDP by $165 billion, or about 1% of GDP. For some countries, however, the difference can be large. This is explored in the Focus box “GDP versus GNP: The Example of Kuwait.”

The Choice between Domestic and Foreign Assets

Openness in financial markets implies that people (or financial institutions that act on their behalf) face a new financial decision: whether to hold domestic assets or foreign assets.

It would appear that we actually have to think about at least two new decisions, the choice of holding domestic money versus foreign money, and the choice of holding domestic interest-paying assets versus foreign interest-paying assets. But remember why people hold money: to engage in transactions. For someone who lives in the United States and whose transactions are mostly or fully in dollars, there is little point in holding foreign currency: Foreign currency cannot be used for transactions in the United States, and if the goal is to hold foreign assets, holding foreign currency is clearly less desirable than holding foreign bonds, which pay interest. This leaves us with only one new choice to think about, the choice between domestic interest-paying assets and foreign interest-paying assets.

Let’s think of these assets for now as domestic one-year bonds and foreign one-year bonds. Consider, for example, the choice between U.S. one-year bonds and U.K. one-year bonds, from the point of view of a U.S. investor.

Suppose you decide to hold U.S. bonds.

Let $i_t$ be the one-year U.S. nominal interest rate. Then, as Figure 18-7 shows, for every dollar you put in U.S. bonds, you will get $(1 + i_t)$ dollars next year. (This is represented by the arrow pointing to the right at the top of the figure.)

Suppose you decide instead to hold U.K. bonds.

To buy U.K. bonds, you must first buy pounds. Let $E_t$ be the nominal exchange rate between the dollar and the pound. For every dollar, you get $E_t$ pounds. (This is represented by the arrow pointing downward in the figure.)

Let $i_t^*$ denote the one-year nominal interest rate on U.K. bonds (in pounds). When next year comes, you will have $E_t(1 + i_t^*)$ pounds. (This is represented by the arrow pointing to the right at the bottom of the figure.)
FOCUS

When oil was discovered in Kuwait, Kuwait’s government decided that a portion of oil revenues would be saved and invested abroad rather than spent, so as to provide future Kuwaiti generations with income when oil revenues came to an end. Kuwait ran a large current account surplus, steadily accumulating large foreign assets. As a result, it has large holdings of foreign assets and receives substantial income from the rest of the world. Table 1 gives GDP, GNP, and net investment income for Kuwait, from 1989 to 1994 (you will see the reason for the choice of dates below).

Note how much larger GNP was compared to GDP throughout the period. Net income from abroad was 34% of GDP in 1989. But note also how net factor payments decreased after 1989. This is because Kuwait had to pay its allies for part of the cost of the 1990–1991 Gulf War and also had to pay for reconstruction after the war. It did so by running a current account deficit—that is, by decreasing its net holdings of foreign assets. This in turn led to a decrease in the income it earned from foreign assets and, by implication, a decrease in its net factor payments.

Since the Gulf War, Kuwait has rebuilt a sizable net foreign asset position. Net income from abroad was 37% of GDP in 2010.

Table 1 GDP, GNP, and Net Income in Kuwait, 1989–1994

<table>
<thead>
<tr>
<th>Year</th>
<th>GDP</th>
<th>GNP</th>
<th>Net Income (NI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>7143</td>
<td>9616</td>
<td>2473</td>
</tr>
<tr>
<td>1990</td>
<td>5328</td>
<td>7560</td>
<td>2232</td>
</tr>
<tr>
<td>1991</td>
<td>3131</td>
<td>4669</td>
<td>1538</td>
</tr>
<tr>
<td>1992</td>
<td>5826</td>
<td>7364</td>
<td>1538</td>
</tr>
<tr>
<td>1993</td>
<td>7231</td>
<td>8386</td>
<td>1151</td>
</tr>
<tr>
<td>1994</td>
<td>7380</td>
<td>8321</td>
<td>941</td>
</tr>
</tbody>
</table>

Source: International Financial Statistics, IMF. All numbers are in millions of Kuwaiti dinars. 1 dinar = $3.6 (2011)

You will then have to convert your pounds back into dollars. If you expect the nominal exchange rate next year to be $E_{t+1}^e$, each pound will be worth $(1/E_{t+1}^e)$ dollars. So you can expect to have $E_t^e(1 + i_t^e)(1/E_{t+1}^e)$ dollars next year for every dollar you invest now. (This is represented by the arrow pointing upward in the figure.)

We shall look at the expression we just derived in more detail soon. But note its basic implication already: In assessing the attractiveness of U.K. versus U.S. bonds, you cannot look just at the U.K. interest rate and the U.S. interest rate; you must also assess what you think will happen to the dollar/pound exchange rate between this year and next.

Let’s now make the same assumption we made in Chapter 15 when discussing the choice between short-term bonds and long-term bonds, or between bonds and stocks. Let’s assume that you and other financial investors care only about the expected rate of return and therefore want to hold only the asset with the highest expected rate of return. In this case, if both U.K. bonds and U.S. bonds are to be held, they must have...
the same expected rate of return. In other words, because of arbitrage, the following relation must hold:

$$1 + i_t = (E_t)(1 + i_t^*)(\frac{1}{E_{t+1}^e})$$

Reorganizing,

$$1 + i_t = (1 + i_t^*)(\frac{E_t}{E_{t+1}^e})$$

Equation (18.2) is called the **uncovered interest parity** relation, or simply the **interest parity condition**.

The assumption that financial investors will hold only the bonds with the highest expected rate of return is obviously too strong, for two reasons:

- It ignores transaction costs: Going in and out of U.K. bonds requires three separate transactions, each with a transaction cost.
- It ignores risk: The exchange rate a year from now is uncertain; holding U.K. bonds is therefore more risky, in terms of dollars, than holding U.S. bonds.

But as a characterization of capital movements among the major world financial markets (New York, Frankfurt, London, and Tokyo), the assumption is not far off. Small changes in interest rates and rumors of impending appreciation or depreciation can lead to movements of billions of dollars within minutes. For the rich countries of the world, the arbitrage assumption in equation (18.2) is a good approximation of reality. Other countries whose capital markets are smaller and less developed, or countries that have various forms of capital controls, have more leeway in choosing their domestic interest rate than is implied by equation (18.2). We shall return to this issue at the end of Chapter 20.

### Interest Rates and Exchange Rates

Let’s get a better sense of what the interest parity condition implies. First rewrite $E_t/E_{t+1}^e$ as $1/(1 + (E_{t+1}^e - E_t)/E_t)$. Replacing in equation (18.2) gives

$$1 + i_t = \frac{(1 + i_t^*)}{[1 + (E_{t+1}^e - E_t)/E_t]}$$

(18.3)

This gives us a relation between the domestic nominal interest rate, $i_t$, the foreign nominal interest rate, $i_t^*$, and the expected rate of appreciation of the domestic currency, $(E_{t+1}^e - E_t)/E_t$. As long as interest rates or the expected rate of depreciation are not too large—say below 20% a year—a good approximation to this equation is given by

$$i_t \approx i_t^* - \frac{E_{t+1}^e - E_t}{E_t}$$

(18.4)

This is the form of the **interest parity condition** you must remember: Arbitrage by investors implies that the domestic interest rate must be equal to the foreign interest rate minus the expected appreciation rate of the domestic currency.

Note that the expected appreciation rate of the domestic currency is also the expected depreciation rate of the foreign currency. So equation (18.4) can be equivalently stated as saying that the domestic interest rate must be equal to the foreign interest rate minus the expected depreciation rate of the foreign currency.

The word *uncovered* is to distinguish this relation from another relation called the **covered interest parity condition**. The covered interest parity condition is derived by looking at the following choice:

- Buy and hold U.S. bonds for one year. Or buy pounds today, buy one-year U.K. bonds with the proceeds, and agree to sell the pounds for dollars a year ahead at a predetermined price, called the **forward exchange rate**.

The rate of return on these two alternatives, which can both be realized at no risk today, must be the same. The covered interest parity condition is a **riskless arbitrage** condition.

Whether holding U.K. bonds or U.S. bonds is more risky actually depends on which investors we are looking at. Holding U.K. bonds is more risky from the point of view of U.S. investors. Holding U.S. bonds is more risky from the point of view of British investors. (Why?)

This follows from Proposition 3 in Appendix 2 at the end of the book.

If the dollar is expected to appreciate by 3% relative to the pound, then the pound is expected to depreciate by 3% relative to the dollar.
Buying Brazilian Bonds

Put yourself back in September 1993 (the very high interest rate in Brazil at the time helps make the point we want to get across here). Brazilian bonds are paying a monthly interest rate of 36.9%. This seems very attractive compared to the annual rate of 3% on U.S. bonds—corresponding to a monthly interest rate of about 0.2%. Shouldn’t you buy Brazilian bonds?

The discussion in this chapter tells you that, to decide, you need one more crucial element, the expected rate of depreciation of the cruzeiro (the name of the Brazilian currency at the time; the currency is now called the real) in terms of dollars.

You need this information because, as we saw in equation (18.4), the return in dollars from investing in Brazilian bonds for a month is equal to one plus the Brazilian interest rate, divided by one plus the expected rate of depreciation of the cruzeiro relative to the dollar:

$$\frac{1 + i_t}{1 + (E_{t+1} - E_t)/E_t}$$

What rate of depreciation of the cruzeiro should you expect over the coming month? A reasonable first pass is to expect the rate of depreciation during the coming month to be equal to the rate of depreciation during last month. The dollar was worth 100,000 cruzeiros at the end of July 1993 and worth 134,600 cruzeiros at the end of August 1993, so the rate of appreciation of the dollar relative to the cruzeiro—equivalently, the rate of depreciation of the cruzeiro relative to the dollar—in August was 34.6%. If depreciation is expected to continue at the same rate in September as it did in August, the expected return from investing in Brazilian bonds for one month is

$$\frac{1.369}{1.346} = 1.017$$

The expected rate of return in dollars from holding Brazilian bonds is only $(1.017 - 1) = 1.6\%$ per month, not the 36.9% per month that initially looked so attractive. Note that 1.6% per month is still much higher than the monthly interest rate on U.S. bonds (about 0.2%). But think of the risk and the transaction costs—all the elements we ignored when we wrote the arbitrage condition. When these are taken into account, you may well decide to keep your funds out of Brazil.

Let’s apply this equation to U.S. bonds versus U.K. bonds. Suppose the one-year nominal interest rate is 2.0% in the United States and 5.0% in the United Kingdom. Should you hold U.K. bonds or U.S. bonds? The answer:

- It depends whether you expect the pound to depreciate relative to the dollar over the coming year by more or less than the difference between the U.S. interest rate and the U.K. interest rate, or 3.0% in this case (5.0% − 2.0%).
- If you expect the pound to depreciate by more than 3.0%, then, despite the fact that the interest rate is higher in the United Kingdom than in the United States, investing in U.K. bonds is less attractive than investing in U.S. bonds. By holding U.K. bonds, you will get higher interest payments next year, but the pound will be worth less in terms of dollars next year, making investing in U.K. bonds less attractive than investing in U.S. bonds.
- If you expect the pound to depreciate by less than 3.0% or even to appreciate, then the reverse holds, and U.K bonds are more attractive than U.S. bonds.

Looking at it another way: If the uncovered interest parity condition holds, and the U.S. one-year interest rate is 3% lower than the U.K. interest rate, it must be that financial investors are expecting, on average, an appreciation of the dollar relative to the pound over the coming year of about 3%, and this is why they are willing to hold U.S. bonds despite their lower interest rate. (Another—and more striking—example is provided in the Focus box “Buying Brazilian Bonds”).

The arbitrage relation between interest rates and exchange rates, either in the form of equation (18.2) or equation (18.4), will play a central role in the following chapters. It suggests that, unless countries are willing to tolerate large movements in their exchange
Openness in goods and financial markets are likely to move very much together. Take the extreme case of two countries that commit to maintaining their bilateral exchange rates at a fixed value. If markets have faith in this commitment, they will expect the exchange rate to remain constant, and the expected depreciation will be equal to zero. In this case, the arbitrage condition implies that interest rates in the two countries will have to move exactly together. Most of the time, as we shall see, governments do not make such absolute commitments to maintain the exchange rate, but they often do try to avoid large movements in the exchange rate. This puts sharp limits on how much they can allow their interest rate to deviate from interest rates elsewhere in the world.

How much do nominal interest rates actually move together in major countries? Figure 18-8 plots the three-month nominal interest rate in the United States and the three-month nominal interest rate in the United Kingdom (both expressed at annual rates), since 1970. The figure shows that the movements are related but not identical. Interest rates were very high in both countries in the early 1980s, and high again—although much more so in the United Kingdom than in the United States—in the late 1980s. They have been low in both countries since the mid-1990s. At the same time, differences between the two have sometimes been quite large: In 1990, for example, the U.K. interest rate was nearly 7% higher than the U.S. interest rate. In the coming chapters, we shall return to why such differences emerge and what their implications may be.

18-3 Conclusions and a Look Ahead

We have now set the stage for the study of the open economy:

- Openness in goods markets allows people and firms to choose between domestic goods and foreign goods. This choice depends primarily on the real exchange rate—the relative price of domestic goods in terms of foreign goods.
Openness in financial markets allows investors to choose between domestic assets and foreign assets. This choice depends primarily on their relative rates of return, which depend on domestic interest rates and foreign interest rates, and on the expected rate of appreciation of the domestic currency.

In the next chapter, Chapter 19, we look at the implications of openness in goods markets. In Chapter 20, we further explore openness in financial markets. In Chapter 21, we discuss the pros and cons of different exchange rate regimes.

Summary

- Openness in goods markets allows people and firms to choose between domestic goods and foreign goods. Openness in financial markets allows financial investors to hold domestic financial assets or foreign financial assets.
- The nominal exchange rate is the price of the domestic currency in terms of foreign currency. From the viewpoint of the United States, the nominal exchange rate between the United States and the United Kingdom is the price of a dollar in terms of pounds.
- A nominal appreciation (an appreciation, for short) is an increase in the price of the domestic currency in terms of foreign currency. In other words, it corresponds to an increase in the exchange rate. A nominal depreciation (a depreciation, for short) is a decrease in the price of the domestic currency in terms of foreign currency. It corresponds to a decrease in the exchange rate.
- The real exchange rate is the relative price of domestic goods in terms of foreign goods. It is equal to the nominal exchange rate times the domestic price level divided by the foreign price level.
- A real appreciation is an increase in the relative price of domestic goods in terms of foreign goods (i.e., an increase in the real exchange rate). A real depreciation is a decrease in the relative price of domestic goods in terms of foreign goods (i.e., a decrease in the real exchange rate).
- The multilateral real exchange rate, or real exchange rate for short, is a weighted average of bilateral real exchange rates, with the weight for each foreign country equal to its share in trade.
- The balance of payments records a country’s transactions with the rest of the world. The current account balance is equal to the sum of the trade balance, net income, and net transfers the country receives from the rest of the world. The capital account balance is equal to capital flows from the rest of the world minus capital flows to the rest of the world.
- The current account and the capital account are mirror images of each other. Leaving aside statistical problems, the current account plus the capital account must sum to zero. A current account deficit is financed by net capital flows from the rest of the world, thus by a capital account surplus. Similarly, a current account surplus corresponds to a capital account deficit.
- Uncovered interest parity, or interest parity for short, is an arbitrage condition stating that the expected rates of return in terms of domestic currency on domestic bonds and foreign bonds must be equal. Interest parity implies that the domestic interest rate approximately equals the foreign interest rate minus the expected appreciation rate of the domestic currency.

Key Terms

openness in goods markets, 379
tariffs, 379
quotas, 379
openness in financial markets, 379
capital controls, 379
openness in factor markets, 379
North American Free Trade Agreement (NAFTA), 379
tradable goods, 381
real exchange rate, 382
nominal exchange rate, 383
appreciation (nominal), 383
depreciation (nominal), 383
fixed exchange rates, 383
revaluation, 383
devaluation, 383
real appreciation, 385
real depreciation, 385
bilateral exchange rate, 387
multilateral exchange rate, 387
multilateral real U.S. exchange rate, 387
trade-weighted real exchange rate, 387
effective real exchange rate, 387
Questions and Problems

QUICK CHECK
All Quick Check questions and problems are available on MyEconLab.

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. If there are no statistical discrepancies, countries with current account deficits must receive net capital inflows.
   b. While the export ratio can be larger than one—as it is in Singapore—the same cannot be true of the ratio of imports to GDP.
   c. That a rich country like Japan has such a small ratio of imports to GDP is clear evidence of an unfair playing field for U.S. exporters to Japan.
   d. Uncovered interest parity implies that interest rates must be the same across countries.
   e. If the dollar is expected to appreciate against the yen, uncovered interest parity implies that the U.S. nominal interest rate will be greater than the Japanese nominal interest rate.
   f. Given the definition of the exchange rate adopted in this chapter, if the dollar is the domestic currency and the euro the foreign currency, a nominal exchange rate of 0.75 means that 0.75 dollars is worth 0.75 euros.
   g. A real appreciation means that domestic goods become less expensive relative to foreign goods.

2. Consider two fictional economies, one called the domestic country and the other the foreign country. Given the transactions listed in (a) through (g), construct the balance of payments for each country. If necessary, include a statistical discrepancy.
   a. The domestic country purchased $100 in oil from the foreign country.
   b. Foreign tourists spent $25 on domestic ski slopes.
   c. Foreign investors were paid $15 in dividends from their holdings of domestic equities.
   d. Domestic residents gave $25 to foreign charities.
   e. Domestic businesses borrowed $65 from foreign banks.
   f. Foreign investors purchased $15 of domestic government bonds.
   g. Domestic investors sold $50 of their holdings of foreign government bonds.

3. Consider two bonds, one issued in euros (€) in Germany, and one issued in dollars ($) in the United States. Assume that both government securities are one-year bonds—paying the face value of the bond one year from now. The exchange rate, E, stands at 1 dollar = 0.75 euro. The face values and prices on the two bonds are given by

<table>
<thead>
<tr>
<th></th>
<th>Face Value</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>$10,000</td>
<td>$9,615.38</td>
</tr>
<tr>
<td>Germany</td>
<td>€10,000</td>
<td>€9,433.96</td>
</tr>
</tbody>
</table>

   a. Compute the nominal interest rate on each of the bonds.
   b. Compute the expected exchange rate next year consistent with uncovered interest parity.
   c. If you expect the dollar to depreciate relative to the euro, which bond should you buy?
   d. Assume that you are a U.S. investor. You exchange dollars for euros and purchase the German bond. One year from now, it turns out that the exchange rate, E, is actually 0.72 (1 dollar = 0.72 euro). What is your realized rate of return in dollars compared to the realized rate of return you would have made had you held the U.S. bond?
   e. Are the differences in rates of return in (d) consistent with the uncovered interest parity condition? Why or why not?

DIG DEEPER
All Dig Deeper questions and problems are available on MyEconLab.

4. Consider a world with three equal-sized economies (A, B, and C) and three goods (clothes, cars, and computers). Assume that consumers in all three economies want to spend an equal amount on all three goods. The value of production of each good in the three economies is given below.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clothes</td>
<td>10</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Cars</td>
<td>5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Computers</td>
<td>0</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

   a. What is GDP in each economy? If the total value of GDP is consumed and no country borrows from abroad, how much will consumers in each economy spend on each of the goods?
b. If no country borrows from abroad, what will be the trade balance in each country? What will be the pattern of trade in this world (i.e., which good will each country export and to whom)?

c. Given your answer to part (b), will country A have a zero trade balance with country B? with country C? Will any country have a zero trade balance with any other country?

d. The United States has a large trade deficit. It has a trade deficit with each of its major trading partners, but the deficit is much larger with some countries (e.g., China) than with others. Suppose the United States eliminates its overall trade deficit (with the world as a whole). Do you expect it to have a zero trade balance with every one of its trading partners? Does the especially large trade deficit with China necessarily indicate that China does not allow U.S. goods to compete on an equal basis with Chinese goods?

5. The exchange rate and the labor market
Suppose the domestic currency depreciates (E falls). Assume that P and P* remain constant.

a. How does the nominal depreciation affect the relative price of domestic goods (i.e., the real exchange rate)? Given your answer, what effect would a nominal depreciation likely have on (world) demand for domestic goods? on the domestic unemployment rate?

b. Given the foreign price level, P*, what is the price of foreign goods in terms of domestic currency? How does a nominal depreciation affect the price of foreign goods in terms of domestic currency? How does a nominal depreciation affect the domestic consumer price index? (Hint: Remember that domestic consumers buy foreign goods (imports) as well as domestic goods.)

c. If the nominal wage remains constant, how does a nominal depreciation affect the real wage?

d. Comment on the following statement. “A depreciating currency puts domestic labor on sale.”

EXPLORE FURTHER

6. Retrieve the nominal exchange rates between Japan and the United States from the Federal Reserve Bank of St. Louis FRED data site. It is series AEXJPUS.

a. Plot the yen versus the dollar since 1971. During which times period(s) did the yen appreciate? During which period(s) did the yen depreciate?

b. Given the current Japanese slump (although there are some encouraging signs at the time of this writing), one way of increasing demand would be to make Japanese goods more attractive. Does this require an appreciation or a depreciation of the yen?

c. What has happened to the yen during the past few years? Has it appreciated or depreciated? Is this good or bad for Japan?

7. Retrieve the most recent World Economic Outlook (WEO) from the web site of the International Monetary Fund (www.imf.org). In the Statistical Appendix, find the table titled “Balances on Current Account,” which lists current account balances around the world. Use the data for the most recent year available to answer parts (a) through (c).

a. Note the sum of current account balances around the world. As noted in the chapter, the sum of current account balances should equal zero. What does this sum actually equal? Why does this sum indicate some mismeasurement (i.e., if the sum were correct, what would it imply)?

b. Which regions of the world are borrowing and which are lending?

c. Compare the U.S. current account balance to the current account balances of the other advanced economies. Is the United States borrowing only from advanced economies?

d. The statistical tables in the WEO typically project data for two years into the future. Look at the projected data on current account balances. Do your answers to parts (b) and (c) seem likely to change in the near future?

8. Saving and investment throughout the world
Retrieve the most recent World Economic Outlook (WEO) from the web site of the International Monetary Fund (www.imf.org). In the Statistical Appendix, find the table titled “Summary of Sources and Uses of World Saving,” which lists saving and investment (as a percentage of GDP) around the world. Use the data for the most recent year available to answer parts (a) and (b).

a. Does world saving equal investment? (You may ignore small statistical discrepancies.) Offer some intuition for your answer.

b. How does U.S. saving compare to U.S. investment? How is the United States able to finance its investment? (We explain this explicitly in the next chapter, but your intuition should help you figure it out now.)

c. The same publication has a table entitled “Table B14. Advanced Economies: Current Account Transactions.” The Advanced countries are the United States, Germany, France, Italy, Spain, Japan, the United Kingdom, and Canada. Use this table to list the countries in which GNP is larger than GDP.

Further Readings

- If you want to learn more about international trade and international economics, a very good textbook is by Paul Krugman and Maurice Obstfeld, *International Economics, Theory and Policy*, 9th ed. (Pearson Addison Wesley, 2010).

- If you want to know current exchange rates between nearly any pair of currencies in the world, look at the “currency converter” on http://www.oanda.com.
In 2009, countries around the world worried about the risk of a recession in the United States. But their worries were not so much for the United States as they were for themselves. To them, a U.S. recession meant lower exports to the United States, a deterioration of their trade position, and weaker growth at home.

Were their worries justified? Figure 18-1 from the previous chapter certainly suggested they were. The U.S. recession clearly led to a world recession. To understand what happened, we must expand the treatment of the goods market in Chapter 3 of the core and account for openness in the analysis of goods markets. This is what we do in this chapter.

Section 19-1 characterizes equilibrium in the goods market for an open economy.

Sections 19-2 and 19-3 show the effects of domestic shocks and foreign shocks on the domestic economy's output and trade balance.

Sections 19-4 and 19-5 look at the effects of a real depreciation on output and the trade balance.

Section 19-6 gives an alternative description of the equilibrium that shows the close connection among saving, investment, and the trade balance.
“The domestic demand for goods” and “The demand for domestic goods” sound close but are not the same. Part of domestic demand falls on foreign goods. Part of foreign demand falls on domestic goods.

---

**19-1 The IS Relation in the Open Economy**

When we were assuming the economy was closed to trade, there was no need to distinguish between the *domestic demand for goods* and the *demand for domestic goods*: They were clearly the same thing. Now, we must distinguish between the two: Some domestic demand falls on foreign goods, and some of the demand for domestic goods comes from foreigners. Let’s look at this distinction more closely.

### The Demand for Domestic Goods

In an open economy, the *demand for domestic goods*, $Z$, is given by

$$ Z = C + I + G - IM/\epsilon + X $$

The first three terms—consumption, $C$, investment, $I$, and government spending, $G$—constitute the *domestic demand for goods*. If the economy were closed, $C + I + G$ would also be the demand for domestic goods. This is why, until now, we have only looked at $C + I + G$. But now we have to make two adjustments:

- First, we must subtract imports—that part of the domestic demand that falls on foreign goods rather than on domestic goods.

  We must be careful here: Foreign goods are different from domestic goods, so we cannot just subtract the quantity of imports, $IM$. If we were to do so, we would be subtracting apples (foreign goods) from oranges (domestic goods). We must first express the value of imports in terms of domestic goods. This is what $IM/\epsilon$ in equation (19.1) stands for: Recall from Chapter 18 that $\epsilon$, the real exchange rate, is defined as the price of domestic goods in terms of foreign goods. Equivalently, $1/\epsilon$ is the price of foreign goods in terms of domestic goods. So $IM(1/\epsilon)$—or, equivalently, $IM/\epsilon$—is the value of imports in terms of domestic goods.

- Second, we must add exports—that part of the demand for domestic goods that comes from abroad. This is captured by the term $X$ in equation (19.1).

### The Determinants of $C$, $I$, and $G$

Having listed the five components of demand, our next task is to specify their determinants. Let’s start with the first three: $C$, $I$, and $G$. Now that we are assuming the economy is open, how should we modify our earlier descriptions of consumption, investment, and government spending? The answer: not very much, if at all. How much consumers decide to spend still depends on their income and their wealth. While the real exchange rate surely affects the *composition* of consumption spending between domestic goods and foreign goods, there is no obvious reason why it should affect the overall *level* of consumption. The same is true of investment: The real exchange rate may affect whether firms buy domestic machines or foreign machines, but it should not affect total investment.

This is good news because it implies that we can use the descriptions of consumption, investment, and government spending that we developed earlier. Therefore,

**Domestic demand:**

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

We assume that consumption depends positively on disposable income, $Y - T$, and that investment depends positively on production, $Y$, and negatively on the real interest rate, $r$. We continue to take government spending, $G$, as given. We leave aside the refinements introduced in Chapters 14 to 17, where we looked at how expectations...
affect spending. We want to take things one step at a time to understand the effects of opening the economy; we shall reintroduce some of those refinements later.

**The Determinants of Imports**

Imports are the part of domestic demand that falls on foreign goods. What do they depend on? They clearly depend on domestic income: Higher domestic income leads to a higher domestic demand for all goods, both domestic and foreign. So a higher domestic income leads to higher imports. They also clearly depend on the real exchange rate—the price of domestic goods in terms of foreign goods: The more expensive domestic goods are relative to foreign goods—equivalently, the cheaper foreign goods are relative to domestic goods—the higher is the domestic demand for foreign goods. So a higher real exchange rate leads to higher imports. Thus, we write imports as

\[ IM = IM(Y, \epsilon) \]  
\[ (+, +) \]  

- An increase in domestic income, \( Y \) (equivalently, an increase in domestic output—income and output are still equal in an open economy) leads to an increase in imports. This positive effect of income on imports is captured by the positive sign under \( Y \) in equation (19.2).
- An increase in the real exchange rate, \( \epsilon \), leads to an increase in imports, \( IM \). This positive effect of the real exchange rate on imports is captured by the positive sign under \( \epsilon \) in equation (19.2). (As \( \epsilon \) goes up, note that \( IM \) goes up, but \( 1/\epsilon \) goes down, so what happens to \( IM/\epsilon \), the value of imports in terms of domestic goods, is ambiguous. We will return to this point shortly.)

**The Determinants of Exports**

Exports are the part of foreign demand that falls on domestic goods. What do they depend on? They depend on foreign income: Higher foreign income means higher foreign demand for all goods, both foreign and domestic. So higher foreign income leads to higher exports. They depend also on the real exchange rate: The higher the price of domestic goods in terms of foreign goods, the lower the foreign demand for domestic goods. In other words, the higher the real exchange rate, the lower are exports.

Let \( Y^* \) denote foreign income (equivalently, foreign output). We therefore write exports as

\[ X = X(Y^*, \epsilon) \]  
\[ (+, -) \]  

- An increase in foreign income, \( Y^* \), leads to an increase in exports.
- An increase in the real exchange rate, \( \epsilon \), leads to a decrease in exports.

**Putting the Components Together**

Figure 19-1 puts together what we have learned so far. It plots the various components of demand against output, keeping constant all other variables (the interest rate, taxes, government spending, foreign output, and the real exchange rate) that affect demand.

In Figure 19-1(a), the line \( DD \) plots domestic demand, \( C + I + G \), as a function of output, \( Y \). This relation between demand and output is familiar from Chapter 3. Under our standard assumptions, the slope of the relation between demand and output is positive but less than one: An increase in output—equivalently, an increase in income—increases demand but less than one-for-one. (In the absence of good reasons...
to the contrary, we draw the relation between demand and output, and the other relations in this chapter, as lines rather than curves. This is purely for convenience, and none of the discussions that follow depend on this assumption.

To arrive at the demand for domestic goods, we must first subtract imports. This is done in Figure 19-1(b) and it gives us the line AA. The line AA represents the domestic demand for domestic goods. The distance between DD and AA equals the value of demand, DD. The line AA represents the domestic demand for domestic goods. The distance between DD and AA equals the value of demand, DD.
imports, $\text{IM}/\epsilon$. Because the quantity of imports increases with income, the distance between the two lines increases with income. We can establish two facts about line $AA$, which will be useful later in the chapter:

- $AA$ is flatter than $DD$: As income increases, some of the additional domestic demand falls on foreign goods rather than on domestic goods. In other words, as income increases, the domestic demand for domestic goods increases less than total domestic demand.

- As long as some of the additional demand falls on domestic goods, $AA$ has a positive slope: An increase in income leads to some increase in the demand for domestic goods.

Finally, we must add exports. This is done in Figure 19-1(c) and it gives us the line $ZZ$, which is above $AA$. The line $ZZ$ represents the demand for domestic goods. The distance between $ZZ$ and $AA$ equals exports. Because exports do not depend on domestic income (they depend on foreign income), the distance between $ZZ$ and $AA$ is constant, which is why the two lines are parallel. Because $AA$ is flatter than $DD$, $ZZ$ is also flatter than $DD$.

From the information in Figure 19-1(c) we can characterize the behavior of net exports—the difference between exports and imports $(X - \text{IM}/\epsilon)$—as a function of output. At output level $Y$, for example, exports are given by the distance $AC$ and imports by the distance $AB$, so net exports are given by the distance $BC$.

This relation between net exports and output is represented as the line $NX$ (for Net eXports) in Figure 19-1(d). Net exports are a decreasing function of output: As output increases, imports increase and exports are unaffected, so net exports decrease. Call $Y_{TB}$ (TB for trade balance) the level of output at which the value of imports equals the value of exports, so that net exports are equal to zero. Levels of output above $Y_{TB}$ lead to higher imports and to a trade deficit. Levels of output below $Y_{TB}$ lead to lower imports and to a trade surplus.

### 19-2 Equilibrium Output and the Trade Balance

The goods market is in equilibrium when domestic output equals the demand—both domestic and foreign—for domestic goods:

$$ Y = Z $$

Collecting the relations we derived for the components of the demand for domestic goods, $Z$, we get

$$ Y = C(Y - T) + I(Y, r) + G - \text{IM}(Y, \epsilon)/\epsilon + X(Y^*, \epsilon) $$  \hspace{1cm} (19.4)

This equilibrium condition determines output as a function of all the variables we take as given, from taxes to the real exchange rate to foreign output. This is not a simple relation; Figure 19-2 represents it graphically, in a more user-friendly way.

In Figure 19-2(a), demand is measured on the vertical axis, output (equivalently production or income) on the horizontal axis. The line $ZZ$ plots demand as a function of output; this line just replicates the line $ZZ$ in Figure 19-1; $ZZ$ is upward sloping, but with slope less than 1.

Equilibrium output is at the point where demand equals output, at the intersection of the line $ZZ$ and the 45-degree line: point $A$ in Figure 19-2(a), with associated output level $Y$.

Figure 19-2(b) replicates Figure 19-1(d), drawing net exports as a decreasing function of output. There is in general no reason why the equilibrium level of output, $Y$, should be the same as the level of output at which trade is balanced, $Y_{TB}$. As we have...
Figure 19-2
Equilibrium Output and Net Exports
The goods market is in equilibrium when domestic output is equal to the demand for domestic goods. At the equilibrium level of output, the trade balance may show a deficit or a surplus.

As in the core, we start with the goods market; the conclusions we derive here will still largely be correct when we introduce financial markets and labor markets later on.

drawn the figure, equilibrium output is associated with a trade deficit, equal to the distance $BC$. Note that we could have drawn it differently, so equilibrium output was associated instead with a trade surplus.

We now have the tools needed to answer the questions we asked at the beginning of this chapter.

19-3 Increases in Demand, Domestic or Foreign
How do changes in demand affect output in an open economy? Let’s start with an old favorite—an increase in government spending—then turn to a new exercise, the effects of an increase in foreign demand.

Increases in Domestic Demand
Suppose the economy is in a recession and the government decides to increase government spending in order to increase domestic demand and output. What will be the effects on output and on the trade balance?

The answer is given in Figure 19-3. Before the increase in government spending, demand is given by $ZZ$ in Figure 19-3(a), and the equilibrium is at point $A$, where output equals $Y$. Let’s assume that trade is initially balanced—even though, as we have seen, there is no reason why this should be true in general. So, in Figure 19-3(b), $Y = Y_{TB}$.

What happens if the government increases spending by $\Delta G$? At any level of output, demand is higher by $\Delta G$, shifting the demand relation up by $\Delta G$ from $ZZ$ to $ZZ'$. The equilibrium point moves from $A$ to $A'$, and output increases from $Y$ to $Y'$. The
An increase in government spending leads to an increase in output and to a trade deficit.

Figure 19-3
The Effects of an Increase in Government Spending

An increase in government spending leads to a trade deficit.

Starting from trade balance, an increase in government spending leads to a trade deficit.

An increase in government spending increases output. The multiplier is smaller than in a closed economy.

The smaller multiplier and the trade deficit have the same cause: Some domestic demand falls on foreign goods.
These two implications are important. In an open economy, an increase in domestic demand has a smaller effect on output than in a closed economy, and an adverse effect on the trade balance. Indeed, the more open the economy, the smaller the effect on output and the larger the adverse effect on the trade balance. Take Belgium, for example. As we saw in Chapter 18, Belgium’s ratio of exports to GDP is very high. It is also true that Belgium’s ratio of imports to GDP is very high. When domestic demand increases in Belgium, much of the increase in demand is likely to result in an increase in the demand for foreign goods rather than an increase in the demand for domestic goods. The effect of an increase in government spending is therefore likely to be a large increase in Belgium’s trade deficit and only a small increase in its output, making domestic demand expansion a rather unattractive policy for Belgium. Even for the United States, which has a much lower import ratio, an increase in demand will be associated with a worsening of the trade balance.

**Increases in Foreign Demand**

Consider now an increase in foreign output, that is, an increase in $Y^*$. This could be due to an increase in foreign government spending, $G^*$—the policy change we just analyzed, but now taking place abroad. But we do not need to know where the increase in $Y^*$ comes from to analyze its effects on the U.S. economy.

Figure 19-4 shows the effects of an increase in foreign activity on domestic output and the trade balance. The initial demand for domestic goods is given by $ZZ$ in

![Figure 19-4](image-url)

*The Effects of an Increase in Foreign Demand*

An increase in foreign demand leads to an increase in output and to a trade surplus.
Figure 19-4(a). The equilibrium is at point \( A \), with output level \( Y \). Let’s again assume trade is balanced, so that in Figure 19-4(b) the net exports associated with \( Y \) equal zero (\( Y = Y_{TB} \)).

It will be useful below to refer to the line that shows the *domestic demand for goods* \( C + I + G \) as a function of income. This line is drawn as \( DD \). Recall from Figure 19-1 that \( DD \) is steeper than \( ZZ \). The difference between \( ZZ \) and \( DD \) equal net exports, so that if trade is balanced at point \( A \), then \( ZZ \) and \( DD \) intersect at point \( A \).

Now consider the effects of an increase in foreign output, \( Y^* \) (for the moment, ignore the line \( DD \); we only need it later). Higher foreign output means higher foreign demand, including higher foreign demand for U.S. goods. So the direct effect of the increase in foreign output is an increase in U.S. exports by some amount, which we shall denote by \( X \).

- For a given level of output, this increase in exports leads to an increase in the demand for U.S. goods by \( \Delta X \), so the line showing the demand for domestic goods as a function of output shifts up by \( \Delta X \), from \( ZZ \) to \( ZZ' \).
- For a given level of output, net exports go up by \( \Delta X \). So the line showing net exports as a function of output in Figure 19-4(b) also shifts up by \( \Delta X \), from \( NX \) to \( NX' \).

The new equilibrium is at point \( A' \) in Figure 19-4(a), with output level \( Y' \). The increase in foreign output leads to an increase in domestic output. The channel is clear: Higher foreign output leads to higher exports of domestic goods, which increases domestic output and the domestic demand for goods through the multiplier.

What happens to the trade balance? We know that exports go up. But could it be that the increase in domestic output leads to such a large increase in imports that the trade balance actually deteriorates? No: The trade balance must improve. To see why, note that, when foreign demand increases, the demand for domestic goods shifts up from \( ZZ \) to \( ZZ' \); but the line \( DD \), which gives the *domestic demand for goods* as a function of output, does not shift. At the new equilibrium level of output \( Y'' \), domestic demand is given by the distance \( DC \), and the demand for domestic goods is given by \( DA' \). Net exports are therefore given by the distance \( CA' \) —which, because \( DD \) is necessarily below \( ZZ' \), is necessarily positive. Thus, while imports increase, the increase does not offset the increase in exports, and the trade balance improves.

### Fiscal Policy Revisited

We have derived two basic results so far:

- An increase in domestic demand leads to an increase in domestic output but leads also to a deterioration of the trade balance. (We looked at an increase in government spending, but the results would have been the same for a decrease in taxes, an increase in consumer spending, and so on.)

- An increase in foreign demand (which could come from the same types of changes taking place abroad) leads to an increase in domestic output and an improvement in the trade balance.

These results, in turn, have two important implications. Both have been in evidence in the crisis.

First, and most obviously, they imply that shocks to demand in one country affect all other countries. The stronger the trade links between countries, the stronger the interactions, and the more countries will move together. This is what we saw in Figure 18-1:
While the crisis started in the United States, it quickly affected the rest of the world. Trade links were not the only reason; financial links also played a central role. But the evidence points to a strong effect of trade, starting with a decrease in exports from other countries to the United States.

Second, these interactions complicate the task of policy makers, especially in the case of fiscal policy. Let’s explore this argument more closely.

Start with the following observation: Governments do not like to run trade deficits, and for good reasons. The main reason: A country that consistently runs a trade deficit accumulates debt vis-à-vis the rest of the world, and therefore has to pay steadily higher interest payments to the rest of the world. Thus, it is no wonder that countries prefer increases in foreign demand (which improve the trade balance) to increases in domestic demand (which worsen the trade balance).

But these preferences can have disastrous implications. Consider a group of countries, all doing a large amount of trade with each other, so that an increase in demand in any one country falls largely on the goods produced in the other countries. Suppose all these countries are in recession and each has roughly balanced trade to start. In this case, each country might be very reluctant to take measures to increase domestic demand. Were it to do so, the result might be a small increase in output but also a large trade deficit. Instead, each country might just wait for the other countries to increase their demand. This way, it gets the best of both worlds, higher output and an improvement in its trade balance. But if all the countries wait, nothing will happen and the recession may last a long time.

Is there a way out? There is—at least in theory. If all countries coordinate their macroeconomic policies so as to increase domestic demand simultaneously, each can increase demand and output without increasing its trade deficit (vis-à-vis the others; their combined trade deficit with respect to the rest of the world will still increase). The reason is clear: The coordinated increase in demand leads to increases in both exports and imports in each country. It is still true that domestic demand expansion leads to larger imports; but this increase in imports is offset by the increase in exports, which comes from the foreign demand expansions.

In practice, however, policy coordination is not so easy to achieve:

Some countries might have to do more than others and may not want to do so: Suppose that only some countries are in recession. Countries that are not in a recession will be reluctant to increase their own demand; but if they do not, the countries that expand will run a trade deficit vis-à-vis countries that do not. Or suppose some countries are already running a large budget deficit. These countries will not want to cut taxes or further increase spending as this would further increase their deficits. They will ask other countries to take on more of the adjustment. Those other countries may be reluctant to do so.

Countries have a strong incentive to promise to coordinate and then not deliver on that promise: Once all countries have agreed, say, to an increase in spending, each country has an incentive not to deliver, so as to benefit from the increase in demand elsewhere and thereby improve its trade position. But if each country cheats, or does not do everything it promised, there will be insufficient demand expansion to get out of the recession.

The result is that, despite declarations by governments at international meetings, coordination often fizzles. Only when things are really bad, does coordination appear to take hold. This was the case in 2009 and is explored in the Focus box “The G20 and the 2009 Fiscal Stimulus.”
In November 2008, the leaders of the G20 met in an emergency meeting in Washington. The G20, a group of ministers of finance and central bank governors from 20 countries, including both the major advanced and the major emerging countries in the world, had been created in 1999 but had not played a major role until the crisis. With mounting evidence that the crisis was going to be both deep and widespread, the group met to coordinate their responses in terms of both macroeconomic and financial policies.

On the macroeconomic front, it had become clear that monetary policy would not be enough, and so the focus turned to fiscal policy. The decrease in output was going to lead to a decrease in revenues, and thus an increase in budget deficits. Dominique Strauss-Kahn, the then managing director of the International Monetary Fund, argued that further fiscal actions were needed and suggested taking additional discretionary measures—either decreases in taxes or increases in spending—adding up to roughly 2% of GDP on average for each country. Here is what he said:

“The fiscal stimulus is now essential to restore global growth. Each country’s fiscal stimulus can be twice as effective in raising domestic output growth if its major trading partners also have a stimulus package.”

He noted that some countries had more room for maneuver than others. “We believe that those countries—advanced and emerging economies—with the strongest fiscal policy frameworks, the best ability to finance fiscal expansion, and the most clearly sustainable debt should take the lead.”

Over the next few months, most countries indeed adopted discretionary measures, aimed at either increasing private or public spending. For the G20 as a whole, discretionary measures added up to about 2.3% of GDP in 2009. Some countries, with less fiscal room, such as Italy, did less. Some countries, such as the United States or France, did more.

Was this fiscal stimulus successful? Some have argued that it was not: After all, the world economy had large negative growth in 2009. The issue here is one of counterfactuals. What would have happened in the absence of the stimulus? Many believe that, absent the fiscal stimulus, growth would have been even more negative, perhaps catastrophically so. Counterfactuals are hard to prove or disprove, and thus the controversy is likely to go on. (On the issue of counterfactuals and the difference between economists and politicians, politicians, following is a very nice quote from U.S. congressman Barney Frank: “Not for the first time, as an elected official, I envy economists. Economists have available to them, in an analytical approach, the counterfactual. Economists can explain that a given decision was the best one that could be made, because they can show what would have happened in the counterfactual situation. They can contrast what happened to what would have happened. No one has ever gotten reelected where the bumper sticker said, ‘It would have been worse without me.’ You probably can get tenure with that. But you can’t win office.”)

Was this fiscal stimulus dangerous? Some have argued that it has led to a large increase in debt, which is now forcing governments to adjust, leading to a fiscal contraction and making recovery more difficult (we discussed this in Chapter 9 and will return to it in Chapter 23). This argument is largely misplaced. Most of the increase in debt does not come from the discretionary measures that were taken, but from the decrease in revenues that came from the decrease in output during the crisis. And a number of countries were running large deficits before the crisis. It remains true, however, that this large increase in debt is now making it more difficult to use fiscal policy to help the recovery.


### 19-4 Depreciation, the Trade Balance, and Output

Suppose the U.S. government takes policy measures that lead to a depreciation of the dollar—a decrease in the nominal exchange rate. (We shall see in Chapter 20 how it can do this by using monetary policy. For the moment we will assume the government can simply choose the exchange rate.)

Recall that the real exchange rate is given by

$$\epsilon = \frac{E P}{P^*}$$

The real exchange rate, $\epsilon$ (the price of domestic goods in terms of foreign goods) is equal to the nominal exchange rate, $E$ (the price of domestic currency in terms of foreign currency) times the domestic price level, $P$, divided by the foreign price level, $P^*$. 
In the short run, we can take the two price levels $P$ and $P^*$ as given. This implies that the nominal depreciation is reflected one for one in a real depreciation. More concretely, if the dollar depreciates vis à vis the yen by 10% (a 10% nominal depreciation), and if the price levels in Japan and the United States do not change, U.S. goods will be 10% cheaper compared to Japanese goods (a 10% real depreciation).

Let’s now ask how this real depreciation will affect the U.S. trade balance and U.S. output.

**Depreciation and the Trade Balance:**

**The Marshall-Lerner Condition**

Return to the definition of net exports:

$$NX = X - IM$$

Replace $X$ and $IM$ by their expressions from equations (19.2) and (19.3):

$$NX = X(Y^*, \epsilon) - IM(Y, \epsilon) / \epsilon$$

As the real exchange rate $\epsilon$ enters the right side of the equation in three places, this makes it clear that the real depreciation affects the trade balance through three separate channels:

- **Exports, $X$, increase.** The real depreciation makes U.S. goods relatively less expensive abroad. This leads to an increase in foreign demand for U.S. goods—an increase in U.S. exports.
- **Imports, $IM$, decrease.** The real depreciation makes foreign goods relatively more expensive in the United States. This leads to a shift in domestic demand toward domestic goods and to a decrease in the quantity of imports.
- **The relative price of foreign goods in terms of domestic goods, $1/\epsilon$, increases.** This increases the import bill, $IM/\epsilon$. The same quantity of imports now costs more to buy (in terms of domestic goods).

For the trade balance to improve following a depreciation, exports must increase enough and imports must decrease enough to compensate for the increase in the price of imports. The condition under which a real depreciation leads to an increase in net exports is known as the **Marshall-Lerner condition.** (It is derived formally in the appendix, called “Derivation of the Marshall Lerner Condition,” at the end of this chapter.) It turns out—with a complication we will state when we introduce dynamics later in this chapter—that this condition is satisfied in reality. So, for the rest of this book, we shall assume that a real depreciation—a decrease in $\epsilon$—leads to an increase in net exports—an increase in $NX$.

**The Effects of a Depreciation**

We have looked so far at the *direct* effects of a depreciation on the trade balance—that is, the effects *given U.S. and foreign output*. But the effects do not end there. The change in net exports changes domestic output, which affects net exports further.

Because the effects of a real depreciation are very much like those of an increase in foreign output, we can use Figure 19-4, the same figure that we used earlier to show the effects of an increase in foreign output.

Just like an increase in foreign output, a depreciation leads to an increase in net exports (assuming, as we do, that the Marshall-Lerner condition holds), at any level of output. Both the demand relation ($ZZ$ in Figure 19-4(a)) and the net exports relation ($NX$ in Figure 19-4(b)) shift up. The equilibrium moves from $A$ to $A'$.
and output increases from \( Y \) to \( Y' \). By the same argument we used earlier, the trade balance improves: The increase in imports induced by the increase in output is smaller than the direct improvement in the trade balance induced by the depreciation.

Let’s summarize: The depreciation leads to a shift in demand, both foreign and domestic, toward domestic goods. This shift in demand leads, in turn, to both an increase in domestic output and an improvement in the trade balance.

Although a depreciation and an increase in foreign output have the same effect on domestic output and the trade balance, there is a subtle but important difference between the two. A depreciation works by making foreign goods relatively more expensive. But this means that given their income, people—who now have to pay more to buy foreign goods because of the depreciation—are worse off. This mechanism is strongly felt in countries that go through a large depreciation. Governments trying to achieve a large depreciation often find themselves with strikes and riots in the streets, as people react to the much higher prices of imported goods. This was the case in Mexico, for example, where the large depreciation of the peso in 1994–1995—from 29 cents per peso in November 1994 to 17 cents per peso in May 1995—led to a large decline in workers’ living standards and to social unrest.

**Combining Exchange Rate and Fiscal Policies**

Suppose output is at its natural level, but the economy is running a large trade deficit. The government would like to reduce the trade deficit while leaving output unchanged. What should it do?

A depreciation alone will not do: It will reduce the trade deficit, but it will also increase output. Nor will a fiscal contraction: It will reduce the trade deficit, but it will decrease output. What should the government do? The answer: Use the right combination of depreciation and fiscal contraction. Figure 19-5 shows what this combination should be.

Suppose the initial equilibrium in Figure 19-5(a) is at \( A \), associated with output \( Y \). At this level of output, there is a trade deficit, given by the distance \( BC \) in Figure 19-5(b). If the government wants to eliminate the trade deficit without changing output, it must do two things:

- It must achieve a depreciation sufficient to eliminate the trade deficit at the initial level of output. So the depreciation must be such as to shift the net exports relation from \( NX \) to \( NX' \) in Figure 19-5(b). The problem is that this depreciation, and the associated increase in net exports, also shifts the demand relation in Figure 19-5(a) from \( ZZ \) to \( ZZ' \). In the absence of other measures, the equilibrium would move from \( A \) to \( A' \), and output would increase from \( Y \) to \( Y' \).
- In order to avoid the increase in output, the government must reduce government spending so as to shift \( ZZ' \) back to \( ZZ \). This combination of a depreciation and a fiscal contraction leads to the same level of output and an improved trade balance.

There is a general point behind this example. To the extent that governments care about both the level of output and the trade balance, they have to use both fiscal policy and exchange rate policies. We just saw one such combination. Table 19-1 gives you others, depending on the initial output and trade situation. Take, for example, the box in the top right corner of the table: Initial output is too low (put...
another way, unemployment is too high), and the economy has a trade deficit. A depreciation will help on both the trade and the output fronts: It reduces the trade deficit and increases output. But there is no reason for the depreciation to achieve both the correct increase in output and the elimination of the trade deficit. Depending on the initial situation and the relative effects of the depreciation on output and the trade balance, the government may need to complement the depreciation with either an increase or a decrease in government spending. This ambiguity is captured by the question mark in the box. Make sure that you understand the logic behind each of the other three boxes.
We have ignored dynamics so far in this chapter. It is time to reintroduce them. The
dynamics of consumption, investment, sales, and production we discussed in Chapter 3
are as relevant to the open economy as they are to the closed economy. But there are
additional dynamic effects as well, which come from the dynamics of exports and im-
ports. We focus on these effects here.

Return to the effects of the exchange rate on the trade balance. We argued earlier
that a depreciation leads to an increase in exports and to a decrease in imports. But
this does not happen overnight. Think of the dynamic effects of, say, a 10% dollar
depreciation.

In the first few months following the depreciation, the effect of the depreciation
is likely to be reflected much more in prices than in quantities. The price of imports
in the United States goes up, and the price of U.S. exports abroad goes down. But the
quantity of imports and exports is likely to adjust only slowly: It takes a while for con-
sumers to realize that relative prices have changed, it takes a while for firms to shift to
cheaper suppliers, and so on. So a depreciation may well lead to an initial deterio-
ration of the trade balance; $\epsilon$ decreases, but neither $X$ nor $IM$ adjusts very much initially,
leading to a decline in net exports $(X - IM/\epsilon)$.

As time passes, the effects of the change in the relative prices of both exports and
imports become stronger. Cheaper U.S. goods cause U.S. consumers and firms to de-
crease their demand for foreign goods: U.S. imports decrease. Cheaper U.S. goods
abroad lead foreign consumers and firms to increase their demand for U.S. goods: U.S.
exports increase. If the Marshall-Lerner condition eventually holds—and we have ar-
gued that it does—the response of exports and imports eventually becomes stronger
than the adverse price effect, and the eventual effect of the depreciation is an improve-
ment of the trade balance.

Figure 19-6 captures this adjustment by plotting the evolution of the trade balance
against time in response to a real depreciation. The pre-depreciation trade deficit is
$OA$. The depreciation initially increases the trade deficit to $OB$: $\epsilon$ decreases, but nei-
ther $IM$ nor $X$ changes right away. Over time, however, exports increase and imports
decrease, reducing the trade deficit. Eventually (if the Marshall-Lerner condition is sat-
sified), the trade balance improves beyond its initial level; this is what happens from
point $C$ on in the figure. Economists refer to this adjustment process as the $J$-curve,
because—admittedly, with a bit of imagination—the curve in the figure resembles a
“J”: first down, then up.

And even these prices may adjust slowly: Consider a dollar depreciation. If you
are an exporter to the United States, you may want to in-
crease your price less than implied by the exchange rate.
In other words, you may de-
crease your markup in order
to remain competitive with
your U.S. competitors. If you
are a U.S. exporter, you may
decrease your price abroad
by less than implied by the
exchange rate. In other
words, you may increase your
markup.

The response of the trade balance to the real exchange rate:
Initially: $X$, $IM$ unchanged, $\epsilon$ decreases $\Rightarrow (X - IM/\epsilon)$ decreases.
Eventually: $X$ increases, $IM$ decreases, $\epsilon$ decreases $\Rightarrow (X - IM/\epsilon)$ increases.
The importance of the dynamic effects of the real exchange rate on the trade balance were seen in the United States in the mid-1980s: Figure 19-7 plots the U.S. trade deficit against the U.S. real exchange rate from 1980 to 1990. As we saw in the last chapter, the period from 1980 to 1985 was one of sharp real appreciation, and the period from 1985 to 1988 one of sharp real depreciation. Turning to the trade deficit, which is expressed as a proportion of GDP, two facts are clear:

1. Movements in the real exchange rate were reflected in parallel movements in net exports. The appreciation was associated with a large increase in the trade deficit, and the later depreciation was associated with a large decrease in the trade balance.

2. There were, however, substantial lags in the response of the trade balance to changes in the real exchange rate. Note how from 1981 to 1983, the trade deficit remained small while the dollar was appreciating. And note how the steady depreciation of the dollar from 1985 onward was not immediately reflected in an improvement in the trade balance before 1987: The dynamics of the J-curve were very much at work during both episodes.

In general, the econometric evidence on the dynamic relation among exports, imports, and the real exchange rate suggests that in all OECD countries, a real depreciation eventually leads to a trade balance improvement. But it also suggests that this process takes some time, typically between six months and a year. These lags have implications not only for the effects of a depreciation on the trade balance, but also for the effects of a depreciation on output. If a depreciation initially decreases net exports, it also initially exerts a contractionary effect on output. Thus, if a government relies on a depreciation both to improve the trade balance and to expand domestic output, the effects will go the “wrong” way for a while.
Saving, Investment, and the Current Account Balance

You saw in Chapter 3 how we could rewrite the condition for equilibrium in the goods market as the condition that investment was equal to saving—the sum of private saving and public saving. We can now derive the corresponding condition for the open economy, and you will see how useful this alternative way of looking at the equilibrium can be.

Start from our equilibrium condition

\[ Y = C + I + G - IM/\epsilon + X \]

Move consumption, \( C \), from the right side to the left side of the equation, subtract taxes, \( T \), from both sides, denote net exports \(- IM/\epsilon + X\) by \( NX\) to get

\[ Y - T - C = I + (G - T) + NX \]

Recall that, in an open economy, the income of domestic residents is equal to output, \( Y \), plus net income from abroad, \( NI \), plus net transfers received. Denote these transfers by \( NT \), and add \( NI \) and \( NT \) to both sides of the equation:

\[ (Y + NI + NT - T) - C = I + (G - T) + (NX + NI + NT) \]

Note that the term in parentheses on left side is equal to disposable income, so the left side is equal to disposable income minus consumption (i.e., saving \( S \)). Note that the sum of net exports, net income from abroad, and net transfers on the right side is equal to the current account. Denote the current account by \( CA \) and rewrite the previous equation as:

\[ S = I + (G - T) + CA \]

Reorganize the equation to read:

\[ CA = S + (T - G) - I \quad (19.5) \]

The current account balance is equal to saving—the sum of private saving and public saving—minus investment. A current account surplus implies that the country is saving more than it invests. A current account deficit implies that the country is saving less than it invests.

One way of getting more intuition for this relation is to go back to the discussion of the current account and the capital account in Chapter 18. There we saw that a current account surplus implies net lending from the country to the rest of the world, and a current account deficit implies net borrowing by the country from the rest of the world. So, consider a country that invests more than it saves, so that \( S + (T - G) - I \) is negative. That country must be borrowing the difference from the rest of the world; it must therefore be running a current account deficit. Symmetrically, a country that lends to the rest of the world is a country that saves more than it invests.

Note some of the things that equation (19.5) says:

- An increase in investment must be reflected in either an increase in private saving or public saving, or in a deterioration of the current account balance—a smaller current account surplus, or a larger current account deficit, depending on whether the current account is initially in surplus or in deficit.
- A deterioration in the government budget balance—either a smaller budget surplus or a larger budget deficit—must be reflected in an increase in either private saving, or in a decrease in investment, or else in a deterioration of the current account balance.
- A country with a high saving rate (private plus government) must have either a high investment rate or a large current account surplus.
The U.S. Current Account Deficit: Origins and Implications

Equation (19.5) tells us that we can look at the current account deficit as the difference between investment and saving. With this in mind, Figure 1 plots investment and saving as ratios to GDP for the United States since 1980; the difference between the two gives the ratio of the current account deficit to GDP. As you can see, the United States has consistently run a current account deficit since 1980. The deficit reached 5% of GDP in the mid-2000s. It is now a little smaller, at around 3%.

A current account deficit can reflect high investment; this can be seen as good news, indicating the country is borrowing abroad to increase its capital and, by implication, its output in the future. Or it can be due to low saving; this can be seen as bad news, indicating that the country is borrowing from abroad to finance consumption and may have a tough time repaying the debt in the future. Figure 1 suggests that, in the case of the United States, the cause of the deficit is low saving, not high investment. The trend evolutions are clear. The ratio of U.S. saving to GDP has declined from 20% in 1980 to 15% just before the crisis, and to 12% in 2010. The ratio of investment to GDP has also declined, although by less, from 22% in 1980 to 20% just before the crisis, and to 15% in 2010.

Looking at the Components of Saving

The next step is to ask what lies behind this trend decrease in U.S. saving. Think of saving as the sum of saving by households, saving by corporations—retained earnings—and saving by the government. All three are plotted in Figure 2, again as ratios to GDP. Leaving aside for the moment the current crisis, the years since 2007, Figure 2 suggests the need to distinguish among three periods:

- A first period, from 1980 to 1992, characterized by a slow but steady decrease in household saving and negative but stable government saving. (Government saving is not exactly the same as the budget surplus: The budget surplus is equal to government saving minus government investment. Put another way, government saving is equal to the budget surplus plus government investment. In 2010, government investment was roughly equal to 2.5% of U.S. GDP.)
- A second period, from 1992 to 2000, characterized by a further slow and steady decrease in household saving, but now more than offset by a steady increase in government saving, from ~2% of GDP in 1992 to nearly 5% of GDP in 2000. This reflects the improvement in the budget under the Clinton administration, which we briefly discussed in Chapter 5. The result was, as we saw in Figure 1, a steady reduction in the current account deficit, the gap between investment and savings.
- A third period, from 2000 to the crisis, characterized by low but stable household saving, but a steady decrease in government saving, reflecting what has become known as the “Bush tax cuts.” We discussed them in a Focus box, “The U.S. Recession of 2001,” in Chapter 5.

Figure 1  Ratios of Saving and Investment to GDP in the United States since 1980

Since the 1980s, the United States has run a current account deficit. This deficit reflects low saving rather than high investment.

Source: NIPA Table 5-1. Saving and Investment by Sector
These tax cuts, initially aimed at fighting the 2001 recession, remained on the books after the recession was over and led to increasing budget deficits, even before the crisis.

Investment, Saving, and the Crisis
Let’s now turn to the crisis, the years since 2007. Figure 1 shows how much saving and investment have declined, with the decline in investment being larger than the decline in saving, leading to a reduction in the current account deficit. Figure 2 shows that behind the decline in saving are very different evolutions of household and corporate saving, on the one hand, and government saving, on the other. Both household and corporate saving have increased substantially (in the case of household saving reversing the long downward trend). But, more than offsetting this increase in private saving, government saving has dramatically decreased, down to $-7\%$ of GDP in 2010.

The easiest way to think about why this has happened is to think back in terms of equation (19.1) (recall that equations (19.1) and (19.5) are equivalent but often give two useful ways of thinking about the evolution of saving, investment, the trade deficit, and the current account deficit.). As we have seen, the crisis has led consumers and firms to be much less optimistic about the future. Consumers have decided to save more, to consume less. Firms have decided to invest less. In order, however, to limit the decrease in demand and thus in output, the U.S. government has increased spending and decreased taxes, leading to a large decrease in government saving and the dramatic increase in the budget deficit we have discussed throughout the book. The results of these changes in behavior during the crisis period have thus been higher private saving, lower government saving, lower investment, and a small decrease in the current account deficit.

What Should and Will Happen Next?
Should the United States reduce its current account deficit?

Even before the crisis, most economists argued that it should. They argued that the current account deficit reflected insufficient saving, both on the part of households and on the part of the government. The low household saving rate was widely seen as reflecting excessively optimistic expectations about the future. Low saving by the government and the increasing budget deficits in the 2000s were seen as dangerous, reflecting a failure of the political system to balance the budget (more on this in Chapter 22). The argument was therefore: Take measures to increase household saving, and decrease budget deficits. To the extent that these lead to a decrease in domestic demand, the argument was to offset this decrease by an increase in net exports, or, put another way, through a decrease in the trade deficit. And to do so, allow for the dollar to depreciate so as to increase exports and decrease imports.

If anything, the crisis has made the need to decrease the current account deficit more urgent. As a result of the crisis, consumption and investment have decreased. In order to sustain demand, the government has relied on monetary and fiscal policy. As we have seen, conventional
monetary policy can no longer be used. The fiscal expansion has led to large deficits, and the government must now reduce them. Thus, if the recovery is to continue, another source of demand must come. From equation (19.1), it is clear that the only source of demand left is net exports. In other words, in order to sustain the recovery, the United States must reduce its trade deficit and, by implication, its current account deficit.

How can this be done? From Section 19-4, the answer would seem to be simple. What is needed is a dollar depreciation, which would increase exports and decrease imports. So the question is why this is not happening. A full discussion will have to wait until the next chapter, when we look at the link between financial decisions and the exchange rate. But, in short, the answer is that there is a high demand for U.S. assets on the part of foreign investors. Thus, they have been willing to finance the U.S. current account deficit, and there has been little downward pressure on the dollar. Whether this remains the case in the future is, at this stage, an open question.

References. In 2011, the G20 asked the IMF for an assessment of the causes of current account deficits in a number of countries, including the United States, Germany, and China. You can find the reports at http://www.imf.org/external/np/g20/map2011.htm. This box is largely based on the United States report.

Note also, however, what equation (19.5) does not say. It does not say, for example, whether a government budget deficit will lead to a current account deficit, or, instead, to an increase in private saving, or to a decrease in investment. To find out what happens in response to a budget deficit, we must explicitly solve for what happens to output and its components using the assumptions that we have made about consumption, investment, exports, and imports. That is, we need to do the complete analysis laid out in this chapter. Using only equation (19.5) can, if you are not careful, be very misleading. To see how misleading, consider, for example, the following argument (which is so common that you may have read something similar in newspapers):

“It is clear the United States cannot reduce its large current account deficit through a depreciation.” Look at equation (19.5): It shows that the current account deficit is equal to investment minus saving. Why should a depreciation affect either saving or investment? So, how can a depreciation affect the current account deficit?

The argument might sound convincing, but we know it is wrong. We showed earlier that a depreciation leads to an improvement in a country’s trade position and, by implication—given net income and transfers—an improvement in the current account. So what is wrong with the argument? A depreciation actually does affect saving and investment: It does so by affecting the demand for domestic goods, thereby increasing output. Higher output leads to an increase in saving over investment, or, equivalently, to a decrease in the current account deficit.

A good way of making sure that you understand the material in this section is to go back and look at the various cases we have considered, from changes in government spending, to changes in foreign output, to combinations of depreciation and fiscal contraction, and so on. Trace what happens in each case to each of the four components of equation (19.5): private saving, public saving (equivalently, the budget surplus), investment, and the current account balance. Make sure, as always, that you can tell the story in words.

A good way of making sure you understand the material in the whole chapter is to read the Focus box on the U.S. Current Account Deficit. It will show you how the concepts we have developed in this chapter can be used to understand the origins and implications of one of the main issues facing U.S. policy makers.
In an open economy, the demand for domestic goods is equal to the domestic demand for goods (consumption, plus investment, plus government spending) minus the value of imports (in terms of domestic goods), plus exports. In an open economy, an increase in domestic demand leads to a smaller increase in output than it would in a closed economy because some of the additional demand falls on imports. For the same reason, an increase in domestic demand also leads to a deterioration of the trade balance. An increase in foreign demand leads, as a result of increased exports, to both an increase in domestic output and an improvement of the trade balance. Because increases in foreign demand improve the trade balance and increases in domestic demand worsen the trade balance, countries might be tempted to wait for increases in foreign demand to move them out of a recession. When a group of countries is in recession, coordination can, in principle, help their recovery.

If the Marshall-Lerner condition is satisfied—and the empirical evidence indicates that it is—a real depreciation leads to an improvement in net exports. A real depreciation leads first to a deterioration of the trade balance, and then to an improvement. This adjustment process is known as the J-curve.

The condition for equilibrium in the goods market can be rewritten as the condition that saving (public and private) minus investment must be equal to the current account balance. A current account surplus corresponds to an excess of saving over investment. A current account deficit usually corresponds to an excess of investment over saving.

In an open economy, the demand for domestic goods is equal to the domestic demand for goods (consumption, plus investment, plus government spending) minus the value of imports (in terms of domestic goods), plus exports.

Demand for domestic goods, 400
Domestic demand for goods, 400
Policy coordination, 408
G20, 409
Marshall-Lerner condition, 410
J-curve, 413

Summary

- In an open economy, the demand for domestic goods is equal to the domestic demand for goods (consumption, plus investment, plus government spending) minus the value of imports (in terms of domestic goods), plus exports.
- In an open economy, an increase in domestic demand leads to a smaller increase in output than it would in a closed economy because some of the additional demand falls on imports. For the same reason, an increase in domestic demand also leads to a deterioration of the trade balance.
- An increase in foreign demand leads, as a result of increased exports, to both an increase in domestic output and an improvement of the trade balance.
- Because increases in foreign demand improve the trade balance and increases in domestic demand worsen the trade balance, countries might be tempted to wait for increases in foreign demand to move them out of a recession. When a group of countries is in recession, coordination can, in principle, help their recovery.
- If the Marshall-Lerner condition is satisfied—and the empirical evidence indicates that it is—a real depreciation leads to an improvement in net exports.
- A real depreciation leads first to a deterioration of the trade balance, and then to an improvement. This adjustment process is known as the J-curve.
- The condition for equilibrium in the goods market can be rewritten as the condition that saving (public and private) minus investment must be equal to the current account balance. A current account surplus corresponds to an excess of saving over investment. A current account deficit usually corresponds to an excess of investment over saving.

Key Terms

demand for domestic goods, 400
domestic demand for goods, 400
policy coordination, 408

Questions and Problems

Quick Check

1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. The current U.S. trade deficit is the result of unusually high investment, not the result of a decline in national saving.
   b. The national income identity implies that budget deficits cause trade deficits.
   c. Opening the economy to trade tends to increase the multiplier because an increase in expenditure leads to more imports.
   d. If the trade deficit is equal to zero, then the domestic demand for goods and the demand for domestic goods are equal.
   e. A real depreciation leads to an immediate improvement in the trade balance.
   f. A small open economy can reduce its trade deficit through fiscal contraction at a smaller cost in output than can a large open economy.
   g. The current high U.S. trade deficit is solely the result of a real appreciation of U.S. goods between 1995 and 2002.
   h. In the United States, GDP is larger than GNP.

2. Real and nominal exchange rates and inflation

   Using the definition of the real exchange rate (and Propositions 7 and 8 in Appendix 2 at the end of the book), you can show that

   \[ \frac{\epsilon_t - \epsilon_{t-1}}{\epsilon_{t-1}} = \frac{E_t - E_{t-1}}{E_{t-1}} + \pi_t - \pi_t^* \]

   In words, the percentage real appreciation equals the percentage nominal appreciation plus the difference between domestic and foreign inflation.
   a. If domestic inflation is higher than foreign inflation, but the domestic country has a fixed exchange rate, what happens to the real exchange rate over time? Assume that the Marshall-Lerner condition holds. What happens to the trade balance over time? Explain in words.
   b. Suppose the real exchange rate is constant—say, at the level required for net exports (or the current account) to equal zero. In this case, if domestic inflation is higher than foreign inflation, what must happen over time to maintain a trade balance of zero?

3. A European recession and the U.S. economy

   a. In 2010, European Union spending on U.S. goods accounted for 23% of U.S. exports (see Table 18-2), and U.S.
exports amounted to 13% of U.S. GDP (see Table 18-1). What was the share of European Union spending on U.S. goods relative to U.S. GDP?

b. Assume that the multiplier in the United States is 2 and that a major slump in Europe would reduce output and imports from the U.S. by 5% (relative to its normal level). Given your answer to part (a), what is the impact on U.S. GDP of the European slump?

c. If the European slump also leads to a slowdown of the other economies that import goods from the United States, the effect could be larger. To put a bound to the size of the effect, assume that U.S. exports decrease by 5% (as a result of changes in foreign output) in one year. What is the impact of a 5% drop in exports on U.S. GDP?

d. Comment on this statement. “Unless Europe can avoid a major slump following the problems with sovereign debt and the Euro, U.S. growth will grind to a halt.”

4. Reproduce the results in Table 19-1.

DIG DEEPER
All Dig Deeper questions and problems are available on MyEconLab.

5. Net exports and foreign demand

a. Suppose there is an increase in foreign output. Show the effect on the domestic economy (i.e., replicate Figure 19-4). What is the effect on domestic output? On domestic net exports?

b. If the interest rate remains constant, what will happen to domestic investment? If taxes are fixed, what will happen to the domestic budget deficit?

c. Using equation (19.5), what must happen to private saving? Explain.

d. Foreign output does not appear in equation (19.5), yet it evidently affects net exports. Explain how this is possible.

6. Eliminating a trade deficit

a. Consider an economy with a trade deficit (\( NX < 0 \)) and with output equal to its natural level. Suppose that, even though output may deviate from its natural level in the short run, it returns to its natural level in the medium run. Assume that the natural level is unaffected by the real exchange rate. What must happen to the real exchange rate over the medium run to eliminate the trade deficit (i.e., to increase \( NX \) to 0)?

b. Now write down the national income identity. Assume again that output returns to its natural level in the medium run. If \( NX \) increases to 0, what must happen to domestic demand (\( C + I + G \)) in the medium run? What government policies are available to reduce domestic demand in the medium run? Identify which components of domestic demand each of these policies affect.

7. Multipliers, openness, and fiscal policy

Consider an open economy characterized by the equations below.

\[
\begin{align*}
C &= c_0 + c_1(Y - T) \\
I &= d_0 + d_1Y \\
IM &= m_1Y \\
X &= x_1Y^* 
\end{align*}
\]

The parameters \( m_1 \) and \( x_1 \) are the propensities to import and export. Assume that the real exchange rate is fixed at a value of 1 and treat foreign income, \( Y^* \), as fixed. Also assume that taxes are fixed and that government purchases are exogenous (i.e., decided by the government). We explore the effectiveness of changes in \( G \) under alternative assumptions about the propensity to import.

a. Write the equilibrium condition in the market for domestic goods and solve for \( Y \).

b. Suppose government purchases increase by one unit. What is the effect on output? (Assume that \( 0 < m_1 < c_1 + d_1 < 1 \). Explain why.)

c. How do net exports change when government purchases increase by one unit?

Now consider two economies, one with \( m_1 = 0.5 \) and the other with \( m_1 = 0.1 \). Each economy is characterized by \( (c_1 + d_1) = 0.6 \).

d. Suppose one of the economies is much larger than the other. Which economy do you expect to have the larger effect of changes in \( G \) under alternative assumptions about the propensity to import?

e. Calculate your answers to parts (b) and (c) for each economy by substituting the appropriate parameter values.

f. In which economy will fiscal policy have a larger effect on output? In which economy will fiscal policy have a larger effect on net exports?

8. Policy coordination and the world economy

Consider an open economy in which the real exchange rate is fixed and equal to one. Consumption, investment, government spending, and taxes are given by

\[
\begin{align*}
C &= 10 + 0.8(Y - T), \\
I &= 10, \\
G &= 10, \text{ and } T = 10
\end{align*}
\]

Imports and exports are given by

\[
\begin{align*}
IM &= 0.3Y \quad \text{and} \quad X = 0.3 Y^*
\end{align*}
\]

where \( Y^* \) denotes foreign output.

a. Solve for equilibrium output in the domestic economy, given \( Y^* \). What is the multiplier in this economy? If we were to close the economy—so exports and imports were identically equal to zero—what would the multiplier be? Why would the multiplier be different in a closed economy?

b. Assume that the foreign economy is characterized by the same equations as the domestic economy (with asterisks reversed). Use the two sets of equations to solve for the equilibrium output of each country. [Hint: Use the equations for the foreign economy to solve for \( Y^* \) as a function of \( Y \) and substitute this solution for \( Y^* \) in part (a).] What is the multiplier for each country now? Why is it different from the open economy multiplier in part (a)?

c. Assume that the domestic government, \( G \), has a target level of output of 125. Assuming that the foreign government does not change \( G^* \), what is the increase in \( G \) necessary to achieve the target output in the domestic economy? Solve for net exports and the budget deficit in each country.

d. Suppose each government has a target level of output of 125 and that each government increases government
spending by the same amount. What is the common increase in $G$ and $G^*$ necessary to achieve the target output in both countries? Solve for net exports and the budget deficit in each country.

e. Why is fiscal coordination, such as the common increase in $G$ and $G^*$ in part (d), difficult to achieve in practice?

**EXPLORE FURTHER**

9. *The U.S. trade deficit, current account deficit, and investment*

a. Define national saving as private saving plus the government surplus—i.e., as $S + T − G$. Now, using equation (19.5), describe the relation among the current account deficit, net investment income, and the difference between national saving and domestic investment.

b. Go to the statistical tables of the most recent *Economic Report of the President* ([www.gpoaccess.gov/eop/](http://www.gpoaccess.gov/eop/)). In Table B-1, “Gross Domestic Product,” retrieve annual data for GDP, gross domestic investment, and net exports from 1980 to the most recent year available. Divide gross domestic investment and net exports by GDP for each year to express their values as a percentage of GDP.

c. The trade surplus in 1980 was roughly zero. Subtract the value of net exports (as a percentage of GDP) in 1981 from the value of net exports (as a percentage of GDP) in the most recent year available. Do the same for gross domestic investment. Has the decline in net exports been matched by an equivalent increase in investment? What do your calculations imply about the change in national saving between 1981 and the present?

d. When the United States began experiencing trade deficits during the 1980s, some officials in the Reagan administration argued that the trade deficits reflected attractive investment opportunities in the United States. Consider three time periods: 1981 to 1990, 1990 to 2000, and 2000 to the present. Apply the analysis of part (c) to each of these time periods (i.e., calculate the change in net exports and gross domestic investment as a percentage of GDP). How does the change in net exports from the 1980 values compare to the change in investment during each period? How did national saving change during each period?

e. Is a trade deficit more worrisome when not accompanied by a corresponding increase in investment? Explain your answer.

f. The question above focuses on the trade deficit rather than the current account deficit. How does net investment income (NI) relate to the difference between the trade deficit and the current account deficit in the United States? Find Table B-103 “U.S. International Transactions” from the Economic Report of the President. Use your work in part (b) to calculate NI as a percent of GDP. Is this value rising or falling over time? What is the implication of such changes?

**Further Readings**

- A good discussion of the relation among trade deficits, current account deficits, budget deficits, private saving, and investment is given in Barry Bosworth’s *Saving and Investment in a Global Economy* (Brookings Institution, 1993).


**APPENDIX: Derivation of the Marshall-Lerner Condition**

Start from the definition of net exports

$$NX = X - IM/ε$$

Assume trade to be initially balanced, so that $NX = 0$ and $X = IM/ε$, or, equivalently, $εX = IM$. The Marshall-Lerner condition is the condition under which a real depreciation, a decrease in $ε$, leads to an increase in net exports.

To derive this condition, first multiply both sides of the equation above by $ε$ to get

$$εNX = εX - IM$$

Now consider a change in the real exchange rate of $Δε$. The effect of the change in the real exchange rate on the left side of the equation is given by $(Δε)NX + ε(ΔNX)$. Note that, if trade is initially balanced, $NX = 0$, so the first term in this expression is equal to zero, and the effect of the change on the left side is simply given by $ε(ΔNX)$. The effect of the change in the real exchange rate on the right side of the equation is given by $(Δε)X + ε(ΔX) - (ΔIM)$. Putting the two sides together gives

$$ε(ΔNX) = (Δε)X + ε(ΔX) - (ΔIM)$$

Divide both sides by $εX$ to get:

$$\frac{ε(ΔNX)}{εX} = \frac{(Δε)X}{εX} + \frac{ε(ΔX)}{εX} - \frac{Δ(IM)}{εX}$$
Simplify, and use the fact that, if trade is initially balanced, $eX = IM$ to replace $eX$ by $IM$ in the last term on the right. This gives

$$\frac{\Delta NX}{X} = \frac{\Delta \epsilon}{\epsilon} + \frac{\Delta X}{X} - \frac{\Delta IM}{IM}$$

The change in the trade balance (as a ratio to exports) in response to a real depreciation is equal to the sum of three terms:

- The first term is equal to the proportional change in the real exchange rate. It is negative if there is a real depreciation.
- The second term is equal to the proportional change in exports. It is positive if there is a real depreciation.
- The third term is equal to minus the proportional change in imports. It is positive if there is a real depreciation.

The Marshall-Lerner condition is the condition that the sum of these three terms be positive. If it is satisfied, a real depreciation leads to an improvement in the trade balance.

A numerical example will help here. Suppose that a 1% depreciation leads to a proportional increase in exports of 0.9%, and to a proportional decrease in imports of 0.8%. (Econometric evidence on the relation of exports and imports to the real exchange rate suggest that these are indeed reasonable numbers.) In this case, the right-hand side of the equation is equal to $-1\% + 0.9\% - (-0.8\%) = 0.7\%$. Thus, the trade balance improves: The Marshall-Lerner condition is satisfied.
In Chapter 19, we treated the exchange rate as one of the policy instruments available to the government. But the exchange rate is not a policy instrument. Rather, it is determined in the foreign exchange market—a market where, as you saw in Chapter 18, there is an enormous amount of trading. This fact raises two obvious questions: What determines the exchange rate? How can policy makers affect it?

These questions motivate this chapter. More generally, we examine the implications of equilibrium in both the goods market and financial markets, including the foreign exchange market. This allows us to characterize the joint movements of output, the interest rate, and the exchange rate in an open economy. The model we develop is an extension to the open economy of the IS–LM model you first saw in Chapter 5 and is known as the Mundell-Fleming model—after the two economists, Robert Mundell and Marcus Fleming, who first put it together in the 1960s. (The model presented here retains the spirit of the original Mundell-Fleming model but differs in its details.)

Section 20-1 looks at equilibrium in the goods market.

Section 20-2 looks at equilibrium in financial markets, including the foreign exchange market.

Section 20-3 puts the two equilibrium conditions together and looks at the determination of output, the interest rate, and the exchange rate.

Section 20-4 looks at the role of policy under flexible exchange rates.

Section 20-5 looks at the role of policy under fixed exchange rates.
Equilibrium in the goods market was the focus of Chapter 19, where we derived the equilibrium condition (equation (19.4)):

\[ Y = C(Y - T) + I(Y, r) + G - IM(Y, \varepsilon)/\varepsilon + X(Y^*, \varepsilon) \]

For the goods market to be in equilibrium, output (the left side of the equation) must be equal to the demand for domestic goods (the right side of the equation). The demand for domestic goods is equal to consumption, \( C \), plus investment, \( I \), plus government spending, \( G \), minus the value of imports, \( IM/\varepsilon \), plus exports, \( X \).

- Consumption, \( C \), depends positively on disposable income \( Y - T \).
- Investment, \( I \), depends positively on output, \( Y \), and negatively on the real interest rate, \( r \).
- Government spending, \( G \), is taken as given.
- The quantity of imports, \( IM \), depends positively on both output, \( Y \), and the real exchange rate, \( \varepsilon \). The value of imports in terms of domestic goods is equal to the quantity of imports divided by the real exchange rate.
- Exports, \( X \), depend positively on foreign output, \( Y^* \), and negatively on the real exchange rate, \( \varepsilon \).

It will be convenient in what follows to regroup the last two terms under “net exports,” defined as exports minus the value of imports:

\[ NX(Y, Y^*, \varepsilon) = X(Y^*, \varepsilon) - IM(Y, \varepsilon)/\varepsilon \]

It follows from our assumptions about imports and exports that net exports, \( NX \), depend on domestic output, \( Y \), foreign output, \( Y^* \), and the real exchange rate \( \varepsilon \): An increase in domestic output increases imports, thus decreasing net exports. An increase in foreign output increases exports, thus increasing net exports. An increase in the real exchange rate leads to a decrease in net exports.

Using this definition of net exports, we can rewrite the equilibrium condition as

\[ Y = C(Y - T) + I(Y, r) + G + NX(Y, Y^*, \varepsilon) \]  

(20.1)

For our purposes, the main implication of equation (20.1) is that both the real interest rate and the real exchange rate affect demand, and in turn equilibrium output:

- An increase in the real interest rate leads to a decrease in investment spending, and, as a result, to a decrease in the demand for domestic goods. This leads, through the multiplier, to a decrease in output.
- An increase in the real exchange rate leads to a shift in demand toward foreign goods, and, as a result, to a decrease in net exports. The decrease in net exports decreases the demand for domestic goods. This leads, through the multiplier, to a decrease in output.

For the remainder of the chapter, we shall simplify equation (20.1) in two ways:

- Given our focus on the short run, we assumed in our previous treatment of the IS–LM model that the (domestic) price level was given. We shall make the same assumption here and extend this assumption to the foreign price level, so the real exchange rate (\( \varepsilon = EP/P^* \)) and the nominal exchange rate, \( E \), move together. A decrease in the nominal exchange rate—a nominal depreciation—leads, one-for-one,
to a decrease in the real exchange rate—a real depreciation. Conversely, an increase in the nominal exchange rate—a nominal appreciation—leads, one-for-one, to an increase in the real exchange rate—a real appreciation. If, for notational convenience, we choose \( P \) and \( P^* \) so that \( P/P^* = 1 \) (and we can do so because both are index numbers), then \( \epsilon = E \) and we can replace \( \epsilon \) by \( E \) in equation (20.1).

Because we take the domestic price level as given, there is no inflation, neither actual nor expected. Therefore, the nominal interest rate and the real interest rate are the same, and we can replace the real interest rate, \( r \), in equation (20.1) by the nominal interest rate, \( i \).

With these two simplifications, equation (20.1) becomes

\[
Y = C(Y - T) + I(Y, i) + G + NX(Y, Y^*, E)
\]

(20.2)

In words: Goods market equilibrium implies that output depends negatively on both the nominal interest rate and the nominal exchange rate.

### 20-2 Equilibrium in Financial Markets

When we looked at financial markets in the IS–LM model, we assumed that people chose only between two financial assets, money and bonds. Now that we look at a financially open economy, we must also take into account the fact that people have a choice between domestic bonds and foreign bonds. Let's consider each choice in turn.

#### Money versus Bonds

When we looked at the determination of the interest rate in the IS–LM model in Chapter 5, we wrote the condition that the supply of money be equal to the demand for money as

\[
\frac{M}{P} = Y L(i)
\]

(20.3)

We took the real supply of money (the left side of equation (20.3)) as given. We assumed that the real demand for money (the right side of equation (20.3)) depended on the level of transactions in the economy, measured by real output, \( Y \), and on the opportunity cost of holding money rather than bonds; that is, the nominal interest rate on bonds, \( i \).

How should we change this characterization now that the economy is open? You will like the answer: not very much, if at all.

In an open economy, the demand for domestic money is still mostly a demand by domestic residents. There is not much reason for, say, the residents of Japan to hold euro currency or euro demand deposits. Transactions in Japan require payment in yen, not in euros. If residents of Japan want to hold euro-denominated assets, they are better off holding euro bonds, which at least pay a positive interest rate. And the demand for money by domestic residents in any country still depends on the same factors as in Chapter 4: their level of transactions, which we measure by domestic real output, and the opportunity cost of holding money, the nominal interest rate on bonds.

Therefore, we can still use equation (20.3) to think about the determination of the nominal interest rate in an open economy. The interest rate must be such that the supply of money and the demand for money are equal. An increase in the money supply leads to a decrease in the interest rate. An increase in money demand, say as a result of an increase in output, leads to an increase in the interest rate.
Domestic Bonds versus Foreign Bonds

As we look at the choice between domestic bonds and foreign bonds, we shall rely on the assumption we introduced in Chapter 19: Financial investors, domestic or foreign, go for the highest expected rate of return. This implies that, in equilibrium, both domestic bonds and foreign bonds must have the same expected rate of return; otherwise, investors would be willing to hold only one or the other, but not both, and this could not be an equilibrium. (Like most economic relations, this relation is only an approximation to reality and does not always hold. More on this in the Focus box “Sudden Stops, Safe Havens, and the Limits of the Interest Parity Condition.”)

As we saw in Chapter 18 (equation (18.2)), this assumption implies that the following arbitrage relation—the interest parity condition—must hold:

\[ (1 + i_t) = (1 + i_t^* \frac{E_t}{E_{t+1}}) \]

where \( i_t \) is the domestic interest rate, \( i_t^* \) is the foreign interest rate, \( E_t \) is the current exchange rate, and \( E_{t+1}^e \) is the future expected exchange rate. The left side of the equation gives the return, in terms of domestic currency, from holding domestic bonds. The right side of the equation gives the expected return, also in terms of domestic currency, from holding foreign bonds. In equilibrium, the two expected returns must be equal.

Multiply both sides by \( E_{t+1}^e \) and reorganize to get

\[ E_t = \frac{1 + i_t}{1 + i_t^*} E_{t+1}^e \]  

(20.4)

For now, we shall take the expected future exchange rate as given and denote it as \( E_e \) (we shall relax this assumption in Chapter 21). Under this assumption, and dropping time indexes, the interest parity condition becomes

\[ E = \frac{1 + i}{1 + i^*} E_e \]

(20.5)

This relation tells us that the current exchange rate depends on the domestic interest rate, on the foreign interest rate, and on the expected future exchange rate:

- An increase in the domestic interest rate leads to an increase in the exchange rate.
- An increase in the foreign interest rate leads to a decrease in the exchange rate.
- An increase in the expected future exchange rate leads to an increase in the current exchange rate.

This relation plays a central role in the real world and will play a central role in this chapter. To understand the relation further, consider the following example.

Consider financial investors—investors, for short—choosing between U.S. bonds and Japanese bonds. Suppose that the one-year interest rate on U.S. bonds is 2%, and the one-year interest rate on Japanese bonds is also 2%. Suppose that the current exchange rate is 100 (one dollar is worth 100 yens), and the expected exchange rate a year from now is also 100. Under these assumptions, both U.S. and Japanese bonds have the same expected return in dollars, and the interest parity condition holds.

Suppose now that investors now expect the exchange rate to be 10% higher a year from now, so \( E_e \) is now equal to 110. At an unchanged current exchange rate, U.S. bonds are now much more attractive than Japanese bonds: U.S. bonds offer an interest rate of 2% in dollars. Japanese bonds still offer an interest rate of 2% in yens, but the yen a year from today are now expected to be worth 10% less in terms of dollars. In terms of
Sudden Stops, Safe Havens, and the Limits to the Interest Parity Condition

The interest parity condition assumes that financial investors care only about expected returns. As we discussed in Chapter 15, investors care not only about expected returns, but also about risk and about liquidity—how easy it is to buy or sell the asset. Much of the time, we can ignore these other factors. Sometimes, however, these factors

Bond Fund Flows

Equity Flows

Figure 1  The Volatility of Capital Flows to Emerging Countries since January 2010

Capital flows have been very volatile during the crisis, reflecting mostly changes in perceived uncertainty.

Source: International Monetary Fund
play a big role in investors’ decisions and in determining exchange rate movements.

This is an issue that many emerging countries know well. Perceptions of risk play an important role in the decision of large foreign investors, such as pension funds, to invest or not invest in their country. Sometimes, the perception that risk has decreased leads many foreign investors to simultaneously buy assets in the country. Sometimes, the perception that risk has increased leads the same investors to want to sell all the assets they have in the country, no matter what the interest rate. These selling episodes, which have affected many Latin American and Asian emerging economies, are known as sudden stops. During these episodes, the interest parity condition fails, and the exchange rate may decrease a lot, without much change in domestic or foreign interest rates.

Indeed, the start of the crisis in 2008 and 2009 was associated with large capital movements which had little to do with expected returns. Worried about uncertainty, many investors from advanced countries decided to take their funds home, where they felt safer. The result was large capital outflows from a number of emerging countries, leading to strong downward pressure on their exchange rates and serious financial problems: For example, some domestic banks that had so far relied on foreign investors for funds found themselves short of funds, which forced them in turn to cut lending to domestic firms and households. This was an important channel of transmission of the crisis from the United States to the rest of the world. And, as the crisis continues, continuing fluctuations in uncertainty are leading to large fluctuations in capital flows, despite relatively stable interest rates. This is best shown in Figure 1, which plots the net flows from funds that invest in emerging market bonds (figure on top) and in emerging market stocks (figure on bottom) from January 2010 to August 2011. What you should take from Figure 1 is the volatility of these net flows: This volatility is not primarily due to movements in interest rates, either in advanced or in emerging countries, but to fluctuations in perceived uncertainty.

A symmetrical phenomenon is at play in some advanced countries. Because of their characteristics, some countries are seen as particularly attractive by investors. This is the case for the United States.

Even in normal times, there is a large foreign demand for U.S. T-bills. The reason is the size and the liquidity of U.S. T-bill market: One can sell or buy large quantities of T-bills quickly and without moving the price very much. Going back to the discussion of the U.S. trade deficit in Chapter 19, one reason why the United States has been able to run a trade deficit, and thus to borrow from the rest of the world for such a long time, is the very high demand for T-bills.

In crisis times, the preference for U.S. T-bill becomes even stronger. The United States is widely seen by investors as being a safe haven, a country in which it is safe to move funds. The result is that times of higher uncertainty are often associated with a stronger demand for U.S. assets and thus upward pressure on the dollar. You can see this in Figure 18-6, where the beginning of the crisis was associated with a strong dollar appreciation. There is some irony here, given that the crisis originated in the United States.

Indeed, some economists wonder how long, given the large budget deficits that it is running, the United States will continue to be perceived as a safe haven. If this were to change, the dollar would depreciate.

Further reading: Among the countries affected by large capital outflows in 2008 and 2009 were also a number of small advanced countries, notably Ireland and Iceland. A number of these countries had built up the same financial vulnerabilities as the United States (those we studied in Chapter 9), and a number of them suffered badly. A very good and easy read is Michael Lewis’s chapters on Ireland and Iceland in Boomerang: Travels in a New Third World, (Norton 2011).

of dollars, the return on Japanese bonds is therefore 2% (the interest rate) − 10% (the expected depreciation of the yen relative to the dollar), or $-8\%$.

So what will happen? At the initial exchange rate of 100, investors want to shift out of Japanese bonds into U.S. bonds. To do so, they must first sell Japanese bonds for yens, then sell yens for dollars, and then use the dollars to buy U.S. bonds. As investors sell yens and buy dollars, the dollar appreciates. By how much? Equation (20.5) gives us the answer: $E = (1.02/1.02) 110 = 110$. The current exchange rate must increase in the same proportion as the expected future exchange rate. Put another way, the dollar must appreciate today by 10\%. When it has appreciated by 10\% so $E = E^e = 110$, the expected returns on U.S. and Japanese bonds are again equal, and there is equilibrium in the foreign exchange market.

Suppose instead that, as a result of a U.S. monetary contraction, the U.S. interest rate increases from 2\% to 5\%. Assume that the Japanese interest rate remains unchanged at 2\%, and that the expected future exchange rate remains unchanged at 100.
At an unchanged current exchange rate, U.S. bonds are now again much more attractive than Japanese bonds. U.S. bonds yield a return of 5% in dollars. Japanese bonds give a return of 2% in yens, and—because the exchange rate is expected to be the same next year as it is today—an expected return of 5% in dollars as well.

So what will happen? Again, at the initial exchange rate of 100, investors want to shift out of Japanese bonds into U.S. bonds. As they do so, they sell yens for dollars, and the dollar appreciates. By how much? Equation (20.5) gives the answer: \( E = (1.05/1.02)100 \approx 103 \). The current exchange rate increases by approximately 3%.

Why 3%? Think of what happens when the dollar appreciates. If, as we have assumed, investors do not change their expectation of the future exchange rate, then the more the dollar appreciates today, the more investors expect it to depreciate in the future (as it is expected to return to the same value in the future). When the dollar has appreciated by 3% today, investors expect it to depreciate by 3% during the coming year. Equivalently, they expect the yen to appreciate relative to the dollar by 3% over the coming year. The expected rate of return in dollars from holding Japanese bonds is therefore 2% (the interest rate in yens) + 3% (the expected yen appreciation), or 5%. This expected rate of return is the same as the rate of return on holding U.S. bonds, so there is equilibrium in the foreign exchange market.

Note that our argument relies heavily on the assumption that, when the interest rate changes, the expected exchange rate remains unchanged. This implies that an appreciation today leads to an expected depreciation in the future—because the exchange rate is expected to return to the same, unchanged, value. We shall relax the assumption that the future exchange rate is fixed in Chapter 21. But the basic conclusion will remain: An increase in the domestic interest rate relative to the foreign interest rate leads to an appreciation.

Figure 20-1 plots the relation between the domestic interest rate, \( i \), and the exchange rate, \( E \), implied by equation (20.5)—the interest parity relation. The relation is drawn for a given expected future exchange rate, \( E^e \), and a given foreign interest rate, \( i^* \), and is represented by an upward-sloping line: The higher the domestic interest rate, the higher the exchange rate. Equation (20.5) also implies that when the domestic interest rate is equal to the foreign interest rate (\( i = i^* \)), the exchange rate will be \( E^e \).
The Open Economy

Extensions

The exchange rate is equal to the expected future exchange rate \( E = E^e \). This implies that the line corresponding to the interest parity condition goes through point \( A \) in the figure.

20-3 Putting Goods and Financial Markets Together

We now have the elements we need to understand the movements of output, the interest rate, and the exchange rate.

Goods-market equilibrium implies that output depends, among other factors, on the interest rate and the exchange rate:

\[
Y = C(Y - T) + I(Y, i) + G + NX(Y, Y^*, E)
\]

The interest rate in turn is determined by the equality of money supply and money demand:

\[
\frac{M}{P} = Y L(i)
\]

And the interest parity condition implies a negative relation between the domestic interest rate and the exchange rate:

\[
E = \frac{1 + i}{1 + i^*} E^e
\]

Together, these three relations determine output, the interest rate, and the exchange rate. Working with three relations is not very easy. But we can easily reduce them to two by using the interest parity condition to eliminate the exchange rate in the goods-market-equilibrium relation. Doing this gives us the following two equations, the open economy versions of our familiar IS and LM relations:

\[
IS: \quad Y = C(Y - T) + I(Y, i) + G + NX(Y, Y^*, \frac{1 + i}{1 + i^*} E^e)
\]

\[
LM: \quad \frac{M}{P} = Y L(i)
\]

Take the IS relation first and consider the effects of an increase in the interest rate on output. An increase in the interest rate now has two effects:

- The first effect, which was already present in a closed economy, is the direct effect on investment: A higher interest rate leads to a decrease in investment, a decrease in the demand for domestic goods, and a decrease in output.

- The second effect, which is only present in the open economy, is the effect through the exchange rate: An increase in the domestic interest rate leads to an increase in the exchange rate—an appreciation. The appreciation, which makes domestic goods more expensive relative to foreign goods, leads to a decrease in net exports, and therefore to a decrease in the demand for domestic goods and a decrease in output.

Both effects work in the same direction: An increase in the interest rate decreases demand directly, and indirectly—through the adverse effect of the appreciation on demand.
The IS relation between the interest rate and output is drawn in Figure 20-2(a), for given values of all the other variables in the relation—namely \( T, G, Y^*, i^*, \) and \( E_e \). The IS curve is downward sloping: An increase in the interest rate leads to lower output. The curve looks very much the same as in the closed economy, but it hides a more complex relation than before: The interest rate affects output not only directly, but also indirectly through the exchange rate.

The LM relation is exactly the same as in the closed economy. The LM curve is upward sloping. For a given value of the real money stock, \( M/P \), an increase in output leads to an increase in the demand for money, and to an increase in the equilibrium interest rate.

Equilibrium in the goods and financial markets is attained at point \( A \) in Figure 20-2(a), with output level \( Y \) and interest rate \( i \). The equilibrium value of the exchange rate cannot be read directly from the graph. But it is easily obtained from Figure 20-2(b), which replicates Figure 20-1 and gives the exchange rate associated with a given interest rate found at point \( B \). The exchange rate associated with the equilibrium interest rate \( i \) is equal to \( E \).

Let’s summarize: We have derived the IS and the LM relations for an open economy:

The IS curve is downward sloping: An increase in the interest rate leads directly, and indirectly (through the exchange rate), to a decrease in demand and a decrease in output.

The LM curve is upward sloping: An increase in income increases the demand for money, leading to an increase in the equilibrium interest rate.

Equilibrium output and the equilibrium interest rate are given by the intersection of the IS and the LM curves. Given the foreign interest rate and the expected future exchange rate, the equilibrium interest rate determines the equilibrium exchange rate.

### 20-4 The Effects of Policy in an Open Economy

Having derived the IS–LM model for the open economy, we now put it to use and look at the effects of policy.

#### The Effects of Fiscal Policy in an Open Economy

Let’s look, again, at a change in government spending. Suppose that, starting from a balanced budget, the government decides to increase defense spending without
An increase in government spending shifts the IS curve to the right. It shifts neither the LM curve nor the interest parity line.

An increase in government spending leads to an increase in output, an increase in the interest rate, and an appreciation. What happens to the level of output? To the composition of output? To the interest rate? To the exchange rate?

The answers are given in Figure 20-3. The economy is initially at point $A$. The increase in government spending by, say, $\Delta G > 0$, increases output at a given interest rate, shifting the IS curve to the right, from IS to IS’ in Figure 20-3(a). Because government spending does not enter the LM relation, the LM curve does not shift. The new equilibrium is at point $A’$, with a higher level of output, $Y’$, and a higher interest rate, $i’$.

In panel (b), the higher interest rate leads to an increase in the exchange rate—an appreciation. So an increase in government spending leads to an increase in output, an increase in the interest rate, and an appreciation.

In words: An increase in government spending leads to an increase in demand, leading to an increase in output. As output increases, so does the demand for money, leading to upward pressure on the interest rate. The increase in the interest rate, which makes domestic bonds more attractive, leads to an appreciation. The higher interest rate and the appreciation both decrease the domestic demand for goods, offsetting some of the effect of government spending on demand and output.

Can we tell what happens to the various components of demand?

- Clearly, consumption and government spending both increase—consumption goes up because of the increase in income; government spending goes up by assumption.
- What happens to investment is ambiguous. Recall that investment depends on both output and the interest rate: $I = I(Y, i)$. On the one hand, output goes up, leading to an increase in investment. But on the other, the interest rate also goes up, leading to a decrease in investment. Depending on which of these two effects dominates, investment can go up or down. In short: The effect of government spending on investment was ambiguous in the closed economy; it remains ambiguous in the open economy.
- Recall that net exports depend on domestic output, foreign output, and the exchange rate: $NX = NX(Y, Y^*, E)$. Thus, both the increase in output and the appreciation combine to decrease net exports: The increase in output increases imports, and the appreciation decreases exports and increases imports. As a result, the budget deficit leads to a deterioration of the trade balance. If trade is balanced to start, then the budget deficit leads to a trade deficit. Note that, while an increase
in the budget deficit increases the trade deficit, the effect is far from mechanical. It works through the effect of the budget deficit on output and on the exchange rate, and, in turn, on the trade deficit.

The Effects of Monetary Policy in an Open Economy

Now that we have looked at fiscal policy, we look at our other favorite policy experiment, a monetary contraction. Look at Figure 20-4(a). At a given level of output, a decrease in the money stock by, say, $\Delta M < 0$ leads to an increase in the interest rate: The $LM$ curve shifts up, from $LM$ to $LM'$. Because money does not directly enter the $IS$ relation, the $IS$ curve does not shift. The equilibrium moves from point $A$ to point $A'$. In Figure 20-4(b), the increase in the interest rate leads to an appreciation.

So a monetary contraction leads to a decrease in output, an increase in the interest rate, and an appreciation. The story is easy to tell. A monetary contraction leads to an increase in the interest rate, making domestic bonds more attractive and triggering an appreciation. The higher interest rate and the appreciation both decrease demand and output. As output falls, money demand falls, leading to a lower interest rate and offsetting some of the initial increase in the interest rate and some of the initial appreciation.

This version of the $IS$–$LM$ model for the open economy was first put together in the 1960s by the two economists we mentioned at the outset of the chapter, Robert Mundell, at Columbia University, and Marcus Fleming, at the International Monetary Fund. How well does the Mundell-Fleming model fit the facts? The answer is: typically quite well, and this is why the model is still very much in use today. Like all simple models, it often needs to be extended; for example to take into account the liquidity trap or the role of risk in affecting portfolio decisions, two important aspects of the crisis. But it is always a good starting point to organize thoughts. (To test the predictions of the model, one could hardly design a better experiment than the sharp monetary and fiscal policy changes the U.S. economy went through in the late 1970s and early 1980s. This is the topic of the Focus box “Monetary Contraction and Fiscal Expansion: The United States in the Early 1980s.” The Mundell-Fleming model and its predictions pass with flying colors.)
Monetary Contraction and Fiscal Expansion: The United States in the Early 1980s

The early 1980s in the United States were dominated by sharp changes both in monetary policy and in fiscal policy. We have already discussed the origins of the change in monetary policy in Chapter 8. By the late 1970s, the Chairman of the Fed, Paul Volcker, concluded that U.S. inflation was too high and had to be reduced. Starting in late 1979, Volcker embarked on a path of sharp monetary contraction, realizing this might lead to a recession in the short run but lower inflation in the medium run.

The change in fiscal policy was triggered by the election of Ronald Reagan in 1980. Reagan was elected on the promise of more conservative policies, namely a scaling down of taxation and the government’s role in economic activity. This commitment was the inspiration for the Economic Recovery Act of August 1981. Personal income taxes were cut by a total of 23%, in three installments from 1981 to 1983. Corporate taxes were also reduced. These tax cuts were not, however, accompanied by corresponding decreases in government spending, and the result was a steady increase in budget deficits, which reached a peak in 1983 at 5.6% of GDP. Table 1 gives spending and revenue numbers for 1980–1984.

What were the Reagan administration’s motivations for cutting taxes without implementing corresponding cuts in spending? These are still being debated today, but there is agreement that there were two main motivations:

One motivation came from the beliefs of a fringe, but influential, group of economists called the supply siders, who argued that a cut in tax rates would cause people and firms to work much harder and more productively, and that the resulting increase in activity would actually lead to an increase, not a decrease, in tax revenues. Whatever the merits of the argument appeared to be then, it proved wrong: Even if some people did work harder and more productively after the tax cuts, tax revenues decreased and the fiscal deficit increased.

The other motivation was more cynical. It was a bet that the cut in taxes, and the resulting increase in deficits, would scare Congress into cutting spending or, at the very least, into not increasing spending further. This motivation turned out to be partly right; Congress found itself under enormous pressure not to increase spending, and the

| Table 1 | The Emergence of Large U.S. Budget Deficits, 1980–1984, (Percent of GDP) |
|---|---|---|---|---|---|
| Spending | 22.0 | 22.8 | 24.0 | 25.0 | 23.7 |
| Revenues | 20.2 | 20.8 | 20.5 | 19.4 | 19.2 |
| Personal taxes | 9.4 | 9.6 | 9.9 | 8.8 | 8.2 |
| Corporate taxes | 2.6 | 2.3 | 1.6 | 1.6 | 2.0 |
| Budget surplus | −1.8 | −2.0 | −3.5 | −5.6 | −4.5 |

Numbers are for fiscal years, which start in October of the previous calendar year. All numbers are expressed as a percentage of GDP. A budget deficit is a negative budget surplus.

Source: Historical Tables, Office of Management and Budget

| Table 2 | Major U.S. Macroeconomic Variables, 1980–1984 |
|---|---|---|---|---|---|
| GDP growth (%) | −0.5 | 1.8 | −2.2 | 3.9 | 6.2 |
| Unemployment rate (%) | 7.1 | 7.6 | 9.7 | 9.6 | 7.5 |
| Inflation (CPI) (%) | 12.5 | 8.9 | 3.8 | 3.8 | 3.9 |
| Interest rate (real) (%) | 11.5 | 14.0 | 10.6 | 8.6 | 9.6 |
| Real exchange rate | 85 | 101 | 111 | 117 | 129 |
| Trade surplus (% of GDP) | −0.5 | −0.4 | −0.6 | −1.5 | −2.7 |

Inflation: rate of change of the CPI. The nominal interest rate is the three-month T-bill rate. The real interest rate is equal to the nominal rate minus the forecast of inflation by DRI, a private forecasting firm. The real exchange rate is the trade-weighted real exchange rate, normalized so that 1973 = 100. A negative trade surplus is a trade deficit.
growth of spending in the 1980s was surely lower than it would have been otherwise. Nonetheless, the adjustment of spending was not enough to offset the shortfall in tax revenues and avoid the rapid increase in deficits.

Whatever the reason for the deficits, the effects of the monetary contraction and fiscal expansion were in line with what the Mundell-Fleming model predicts. Table 2 gives the evolution of the main macroeconomic variables from 1980 to 1984.

From 1980 to 1982, the evolution of the economy was dominated by the effects of the monetary contraction. Interest rates, both nominal and real, increased sharply, leading both to a large dollar appreciation and to a recession. The goal of lowering inflation was achieved; by 1982, inflation was down to about 4%, down from 12.5% in 1980. Lower output and the dollar appreciation had opposing effects on the trade balance (lower output leading to lower imports and an improvement in the trade balance; the appreciation of the dollar leading to a deterioration in the trade balance), resulting in little change in the trade deficit before 1982.

From 1982 on, the evolution of the economy was dominated by the effects of the fiscal expansion. As our model predicts, these effects were strong output growth, high interest rates, and further dollar appreciation. The effects of high output growth and the dollar appreciation were an increase in the trade deficit to 2.7% of GDP by 1984. By the mid-1980s, the main macroeconomic policy issue had become that of the twin deficits: the budget deficit and the trade deficit. The twin deficits were to remain one of the central macroeconomic issues throughout the 1980s and early 1990s.

### 20-5 Fixed Exchange Rates

We have assumed so far that the central bank chose the money supply and let the exchange rate freely adjust in whatever manner was implied by equilibrium in the foreign exchange market. In many countries, this assumption does not reflect reality: Central banks act under implicit or explicit exchange rate targets and use monetary policy to achieve those targets. The targets are sometimes implicit, sometimes explicit; they are sometimes specific values, sometimes bands or ranges. These exchange rate arrangements (or regimes, as they are called) come under many names. Let’s first see what the names mean.

**Pegs, Crawling Pegs, Bands, the EMS, and the Euro**

At one end of the spectrum are countries with flexible exchange rates such as the United States, the United Kingdom, Japan, and Canada. These countries have no explicit exchange rate targets. Although their central banks probably do not ignore movements in the exchange rate, they have shown themselves quite willing to let their exchange rates fluctuate considerably.

At the other end are countries that operate under fixed exchange rates. These countries maintain a fixed exchange rate in terms of some foreign currency. Some peg their currency to the dollar. For example, from 1991 to 2001, Argentina pegged its currency, the peso, at the highly symbolic exchange rate of one dollar for one peso (more on this in Chapter 21). Other countries used to peg their currency to the French franc (most of these are former French colonies in Africa); as the French franc has been replaced by the euro, they are now pegged to the euro. Still other countries peg their currency to a basket of foreign currencies, with the weights reflecting the composition of their trade.

The label “fixed” is a bit misleading: It is not the case that the exchange rate in countries with fixed exchange rates never actually changes. But changes are rare. An extreme case is that of the African countries pegged to the French franc. When their exchange rates were readjusted in January 1994, this was the first adjustment in 45 years! Because these changes are rare, economists use specific words to distinguish them from the daily changes that occur under flexible exchange rates. A decrease in the exchange rate under a regime of fixed exchange rates is called a **devaluation** rather than a depreciation, and an increase in the exchange rate under a regime of fixed exchange rates is called a **revaluation** rather than an appreciation.

These terms were first introduced in Chapter 18.
Between these extremes are countries with various degrees of commitment to an exchange rate target. For example, some countries operate under a crawling peg. The name describes it well: These countries typically have inflation rates that exceed the U.S. inflation rate. If they were to peg their nominal exchange rate against the dollar, the more rapid increase in their domestic price level above the U.S. price level would lead to a steady real appreciation and rapidly make their goods uncompetitive. To avoid this effect, these countries choose a predetermined rate of depreciation against the dollar. They choose to “crawl” (move slowly) vis-à-vis the dollar.

Yet another arrangement is for a group of countries to maintain their bilateral exchange rates (the exchange rate between each pair of countries) within some bands. Perhaps the most prominent example was the European Monetary System (EMS), which determined the movements of exchange rates within the European Union from 1978 to 1998. Under the EMS rules, member countries agreed to maintain their exchange rate relative to the other currencies in the system within narrow limits or bands around a central parity—a given value for the exchange rate. Changes in the central parity and devaluations or revaluations of specific currencies could occur, but only by common agreement among member countries. After a major crisis in 1992, which led a number of countries to drop out of the EMS altogether, exchange rate adjustments became more and more infrequent, leading a number of countries to move one step further and adopt a common currency, the euro. The conversion from domestic currencies to the euro began on January 1, 1999, and was completed in early 2002. We shall return to the implications of the move to the euro in Chapter 21.

We shall discuss the pros and cons of different exchange regimes in the next chapter. But first, we must understand how pegging (also called fixing) the exchange rate affects monetary policy and fiscal policy. This is what we do in the rest of this section.

Pegging the Exchange Rate, and Monetary Control

Suppose a country decides to peg its exchange rate at some chosen value, call it $E$. How does it actually achieve this? The government cannot just announce the value of the exchange rate and remain idle. Rather, it must take measures so that its chosen exchange rate will prevail in the foreign exchange market. Let’s look at the implications and mechanics of pegging.

Pegging or no pegging, the exchange rate and the nominal interest rate must satisfy the interest parity condition

\[
(1 + i_t) = (1 + i_t^*) \left( \frac{E_t}{E_{t+1}^*} \right)
\]

Now suppose the country pegs the exchange rate at $E$, so the current exchange rate $E_t = E$. If financial and foreign exchange markets believe that the exchange rate will remain pegged at this value, then their expectation of the future exchange rate, $E_{t+1}^*$, is also equal to $E$, and the interest parity relation becomes

\[
(1 + i_t) = (1 + i_t^*), \Rightarrow i_t = i_t^*
\]

In words: If financial investors expect the exchange rate to remain unchanged, they will require the same nominal interest rate in both countries. Under a fixed exchange rate and perfect capital mobility, the domestic interest rate must be equal to the foreign interest rate.
This condition has one further important implication. Return to the equilibrium condition that the supply of money and demand for money be equal. Now that \( i = i^* \), this condition becomes:

\[
\frac{M}{P} = YL(i^*)
\]  

(20.6)

Suppose an increase in domestic output increases the demand for money. In a closed economy, the central bank could leave the money stock unchanged, leading to an increase in the equilibrium interest rate. In an open economy, and under flexible exchange rates, the central bank can still do the same: The result will be both an increase in the interest rate and an appreciation. But under fixed exchange rates, the central bank cannot keep the money stock unchanged. If it did, the domestic interest rate would increase above the foreign interest rate, leading to an appreciation. To maintain the exchange rate, the central bank must increase the supply of money in line with the increase in the demand for money so the equilibrium interest rate does not change. Given the price level, \( P \), nominal money, \( M \), must adjust so that equation (20.6) holds.

Let’s summarize: Under fixed exchange rates, the central bank gives up monetary policy as a policy instrument. With a fixed exchange rate, the domestic interest rate must be equal to the foreign interest rate. And the money supply must adjust so as to maintain the interest rate.

**Fiscal Policy under Fixed Exchange Rates**

If monetary policy can no longer be used under fixed exchange rates, what about fiscal policy? To answer this question, we use Figure 20-5.

Figure 20-5 starts by replicating Figure 20-3(a), which we used earlier to analyze the effects of fiscal policy under flexible exchange rates. In that case, we saw that a fiscal expansion (\( \Delta G > 0 \)) shifted the IS curve to the right. Under flexible exchange rates, the money stock remained unchanged, leading to a movement in the equilibrium from point \( A \) to point \( B \), with an increase in output from \( Y_A \) to \( Y_B \), an increase in the interest rate, and an appreciation.

These results depend very much on the interest rate parity condition, which in turn depends on the assumption of perfect capital mobility—that financial investors go for the highest expected rate of return. The case of fixed exchange rates with imperfect capital mobility, which is more relevant for middle-income countries, such as in Latin America or Asia, is treated in the appendix to this chapter.

---

**Figure 20-5**

*The Effects of a Fiscal Expansion under Fixed Exchange Rates*

Under flexible exchange rates, a fiscal expansion increases output from \( Y_A \) to \( Y_B \). Under fixed exchange rates, output increases from \( Y_A \) to \( Y_C \).
German Reunification, Interest Rates, and the EMS

Under a fixed exchange rate regime such as the European Monetary System (EMS) (let’s ignore here the degree of flexibility that was afforded by the bands), no individual country can change its interest rate if the other countries do not change theirs as well. So, how do interest rates actually change? Two arrangements are possible. One is for all the member countries to coordinate changes in their interest rates. Another is for one of the countries to take the lead and for the other countries to follow—this is in effect what happened in the EMS, with Germany as the leader.

During the 1980s, most European central banks shared similar goals and were happy to let the Bundesbank (the German central bank) take the lead. But in 1990, German unification led to a sharp divergence in goals between the Bundesbank and the central banks of the other EMS countries. Large budget deficits, triggered by transfers to people and firms in Eastern Germany, together with an investment boom, led to a large increase in demand in Germany. The Bundesbank’s fear that this shift would generate too strong an increase in activity led it to adopt a restrictive monetary policy. The result was strong growth in Germany together with a large increase in interest rates. This may have been the right policy mix for Germany. But for the other European countries, this policy mix was much less appealing. They were not experiencing the same increase in demand, but to stay in the EMS, they had to match German interest rates. The net result was a sharp decrease in demand and output in the other countries. These results are presented in Table 1, which gives nominal interest rates, real interest rates, inflation rates, and GDP growth from 1990 to 1992 for Germany and for two of its EMS partners, France and Belgium.

Note first how the high German nominal interest rates were matched by both France and Belgium. In fact, nominal interest rates were actually higher in France than in Germany in all three years! This is because France needed higher interest rates than Germany to maintain the Deutsche Mark/franc parity. The reason is that financial markets were not sure that France would actually keep the parity of the franc relative to the DM. Worried about a possible devaluation of the franc, financial investors asked for a higher interest rate on French bonds than on German bonds.

Although France and Belgium had to match—or, as we have just seen, more than match—German nominal rates, both countries had less inflation than Germany. The result was very high real interest rates, much higher than the rate in Germany: In both France and Belgium, average real interest rates from 1990 to 1992 were close to 7%. And in both countries, the period 1990–1992 was characterized by slow growth and rising unemployment. Unemployment in France in 1992 was 10.4%, up from 8.9% in 1990. The corresponding numbers for Belgium were 12.1% and 8.7%.

A similar story was unfolding in the other EMS countries. By 1992, average unemployment in the European Union, which had been 8.7% in 1990, had increased to 10.3%. The effects of high real interest rates on spending were not the only source of this slowdown, but they were the main one.

By 1992, an increasing number of countries were wondering whether to keep defending their EMS parity or to give it up and lower their interest rates. Worried about the risk of devaluations, financial markets started to ask for higher interest rates in those countries where they thought devaluations were more likely. The result was two major exchange rate crises, one in the fall of 1992, and the other in the summer of 1993. By the end of these two crises, two countries, Italy and the United Kingdom, had left the EMS. We shall look at these crises, their origins, and their implications, in Chapter 21.

### Table 1 German Reunification, Interest Rates, and Output Growth: Germany, France, and Belgium, 1990–1992

<table>
<thead>
<tr>
<th></th>
<th>Nominal Interest Rates (%)</th>
<th>Inflation (%)</th>
<th>Real Interest Rates (%)</th>
<th>GDP Growth (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>8.5</td>
<td>9.2</td>
<td>9.5</td>
<td>2.7</td>
</tr>
<tr>
<td>France</td>
<td>10.3</td>
<td>9.6</td>
<td>10.3</td>
<td>2.9</td>
</tr>
<tr>
<td>Belgium</td>
<td>9.6</td>
<td>9.4</td>
<td>9.4</td>
<td>2.9</td>
</tr>
</tbody>
</table>

The nominal interest rate is the short-term nominal interest rate. The real interest rate is the realized real interest rate over the year—that is, the nominal interest rate minus actual inflation over the year. All rates are annual.

Source: OECD Economic Outlook
However, under fixed exchange rates the central bank cannot let the currency appreciate. Because the increase in output leads to an increase in the demand for money, the central bank must accommodate this increased demand for money by increasing the money supply. In terms of Figure 20-5, the central bank must shift the LM curve down as the IS curve shifts to the right, so as to maintain the same interest rate and thus the same exchange rate. The equilibrium therefore moves from A to C, with higher output $Y_C$ and unchanged interest and exchange rates. So, under fixed exchange rates, fiscal policy is more powerful than it is under flexible exchange rates. This is because fiscal policy triggers monetary accommodation.

As this chapter comes to an end, a question should have started to form in your mind: Why would a country choose to fix its exchange rate? You have seen a number of reasons why this appears to be a bad idea:

- By fixing the exchange rate, a country gives up a powerful tool for correcting trade imbalances or changing the level of economic activity.
- By committing to a particular exchange rate, a country also gives up control of its interest rate. Not only that, but the country must match movements in the foreign interest rate, at the risk of unwanted effects on its own activity. This is what happened in the early 1990s in Europe. Because of the increase in demand due to the reunification of West and East Germany, Germany felt it had to increase its interest rate. To maintain their parity with the Deutsche Mark, other countries in the European Monetary System were forced to also increase their interest rates, something that they would rather have avoided. (This is the topic of the Focus box “German Reunification, Interest Rates, and the EMS.”)
- Although the country retains control of fiscal policy, one policy instrument may not be enough. As you saw in Chapter 19, for example, a fiscal expansion can help the economy get out of a recession, but only at the cost of a larger trade deficit. And a country that wants, for example, to decrease its budget deficit cannot, under fixed exchange rates, use monetary policy to offset the contractionary effect of its fiscal policy on output.

So why do some countries fix their exchange rate? Why have 17 European countries—with more to come—adopted a common currency? To answer these questions, we must do some more work. We must look at what happens not only in the short run—which is what we did in this chapter—but also in the medium run, when the price level can adjust. We must look at the nature of exchange rate crises. Once we have done this, we shall then be able to assess the pros and cons of different exchange rate regimes. These are the topics we take up in Chapter 21.

Summary

- In an open economy, the demand for domestic goods depends both on the interest rate and on the exchange rate. An increase in the interest rate decreases the demand for domestic goods. An increase in the exchange rate—an appreciation—also decreases the demand for domestic goods.
- The interest rate is determined by the equality of money demand and money supply. The exchange rate is determined by the interest parity condition, which states that domestic and foreign bonds must have the same expected rate of return in terms of domestic currency.
- Given the expected future exchange rate and the foreign interest rate, increases in the domestic interest rate lead to an increase in the exchange rate—an appreciation. Decreases in the domestic interest rate lead to a decrease in the exchange rate—a depreciation.
- Under flexible exchange rates, an expansionary fiscal policy leads to an increase in output, an increase in the interest rate, and an appreciation.
Under fixed exchange rates and the interest parity condition, a country must maintain an interest rate equal to the foreign interest rate. The central bank loses the use of monetary policy as a policy instrument. Fiscal policy becomes more powerful than under flexible exchange rates, however, because fiscal policy triggers monetary accommodation, and so does not lead to offsetting changes in the domestic interest rate and exchange rate.

Under flexible exchange rates, a contractionary monetary policy leads to a decrease in output, an increase in the interest rate, and an appreciation.

There are many types of exchange rate arrangements. They range from fully flexible exchange rates to crawling pegs, to fixed exchange rates (or pegs), to the adoption of a common currency. Under fixed exchange rates, a country maintains a fixed exchange rate in terms of a foreign currency or a basket of currencies.

Key Terms

Mundell-Fleming model, 423
sudden stops, 428
safe haven, 428
supply siders, 434
twin deficits, 435
peg, 435

Questions and Problems

QUICK CHECK
All Quick Check questions and problems are available on MyEconLab.
1. Using the information in this chapter, label each of the following statements true, false, or uncertain. Explain briefly.
   a. A fiscal expansion tends to increase net exports.
   b. Fiscal policy has a greater effect on output in an economy with fixed exchange rates than in an economy with flexible exchange rates.
   c. Other things being equal, the interest parity condition implies that the domestic currency will appreciate in response to an increase in the expected exchange rate.
   d. If financial investors expect the dollar to depreciate against the yen over the coming year, one-year interest rates will be higher in the United States than in Japan.
   e. If the Japanese interest rate is equal to zero, foreigners will not want to hold Japanese bonds.
   f. Under fixed exchange rates, the money stock must be constant.

2. Consider an open economy with flexible exchange rates. Suppose output is at the natural level, but there is a trade deficit. What is the appropriate fiscal and monetary policy mix?

3. In this chapter, we showed that a monetary expansion in an economy operating under flexible exchange rates leads to an increase in output and a depreciation of the domestic currency.
   a. How does a monetary expansion (in an economy with flexible exchange rates) affect consumption and investment?
   b. How does a monetary expansion (in an economy with flexible exchange rates) affect net exports?

4. Flexible exchange rates and foreign macroeconomic policy.
   Consider an open economy with flexible exchange rates. Let UIP stand for the uncovered interest parity condition.
   a. In an IS–LM–UIP diagram, show the effect of an increase in foreign output, \( Y^* \), on domestic output, \( Y \). Explain in words.
   b. In an IS–LM–UIP diagram, show the effect of an increase in the foreign interest rate, \( i^* \), on domestic output, \( Y \). Explain in words.
   c. Given the discussion of the effects of fiscal policy in this chapter, what effect is a foreign fiscal expansion likely to have on foreign output, \( Y^* \), and on the foreign interest rate, \( i^* \)? Given the discussion of the effects of monetary policy in this chapter, what effect is a foreign monetary expansion likely to have on \( Y^* \) and \( i^* \)?
   d. Given your answers to parts (a), (b), and (c), how does a foreign fiscal expansion affect domestic output? How does a foreign monetary expansion affect domestic output? (Hint: One of these policies has an ambiguous effect on output.)

DIG DEEPER
All Dig Deeper questions and problems are available on MyEconLab.
5. Fixed exchange rates and foreign macroeconomic policy.
   Consider a fixed exchange rate system, in which a group of countries (called follower countries) peg their currencies to the currency of one country (called the leader country). Since the currency of the leader country is not fixed against the currencies of countries outside the fixed exchange rate system, the leader country can conduct monetary policy as it wishes. For this problem, consider the domestic country to be a follower country and the foreign country to be the leader country.
   a. Redo the analysis of Problem 4(a).
   b. Redo the analysis of Problem 4(b).
c. Using your answers to parts (a) and (b) and Problem 4(c), how does a foreign monetary expansion (by the leader country) affect domestic output? How does a foreign fiscal expansion (by the leader country) affect domestic output? (You may assume that the effect of Y* on domestic output is small.) How do your answers differ from those in 4(d)?

6. The exchange rate as an automatic stabilizer
Consider an economy that suffers a fall in business confidence (which tends to reduce investment). Let UIP stand for the uncovered interest parity condition.

a. Suppose the economy has a flexible exchange rate. In an IS–LM–UIP diagram, show the short-run effect of the fall in business confidence on output, the interest rate, and the exchange rate. How does the change in the exchange rate, by itself, tend to affect output? Does the change in the exchange rate dampen (make smaller) or amplify (make larger) the effect of the fall in business confidence on output?

b. Suppose instead the economy has a fixed exchange rate. In an IS–LM–UIP diagram, show how the economy responds to the fall in business confidence. What must happen to the money supply in order to maintain the fixed exchange rate? How does the effect on output in this economy, with fixed exchange rates, compare to the effect you found for the economy in part (a), with flexible exchange rates?

c. Explain how the exchange rate acts as an automatic stabilizer in an economy with flexible exchange rates.

EXPLORE FURTHER
7. Demand for U.S. assets, the dollar, and the trade deficit
This question explores how an increase in demand for U.S. assets may have slowed the depreciation of the dollar that many economists believe is warranted by the large U.S. trade deficit and the need to stimulate the demand for domestic goods after the crisis. Here, we modify the IS–LM–UIP framework (where UIP stands for uncovered interest parity) to analyze the effects of an increase in demand for U.S. assets. Write the uncovered interest parity condition as

\[ (1 + i_i) = (1 + i_r^*) \left( \frac{E^i}{E^r_{t+1}} \right) - x \]

where the parameter x represents factors affecting the relative demand for domestic assets. An increase in x means that investors are willing to hold domestic assets at a lower interest rate (given the foreign interest rate, and the current and expected exchange rates).

a. Solve the UIP condition for the current exchange rate, \( E^i \).

b. Substitute the result from part (a) in the IS curve and construct the UIP diagram. As in the text, you may assume that P and P* are constant and equal to one.

c. Suppose that as a result of a large trade deficit in the domestic economy, financial market participants believe that the domestic currency must depreciate in the future. Therefore, the expected exchange rate, \( E^r_{t+1} \), decreases. Show the effect of the decrease in the expected exchange rate in the IS–LM–UIP diagram. What are the effects on the exchange rate and the trade balance? (Hint: In analyzing the effect on the trade balance, remember why the IS curve shifted in the first place.)

d. Now suppose that the relative demand for domestic assets, x, increases. As a benchmark, suppose that the increase in x is exactly enough to return the IS curve to its original position, before the decrease in the expected exchange rate. Show the combined effects of the decrease in \( E^r_{t+1} \) and the increase in x in your IS–LM–UIP diagram. What are the ultimate effects on the exchange rate and the trade balance?

e. Based on your analysis, is it possible that an increase in demand for U.S. assets could prevent the dollar from depreciating? Is it possible that an increase in demand for U.S. assets could worsen the U.S. trade balance? Explain your answers.

By the time you read this book, it is possible that relative demand for U.S. assets could be weaker than it was at the time of this writing and that the dollar could be depreciating. Think about how you would use the framework of this problem to assess the current situation.

8. Expected depreciation of the dollar
Martin Feldstein, as mentioned at the end of Chapter 19, is one prominent economist who argues that the dollar may need to depreciate by as much as 20% to 40% in real terms to achieve a reasonable improvement in the trade balance.

a. Go the web site of The Economist (www.economist.com) and find data on 10-year interest rates. Look in the section “Markets & Data” and then the subsection “Economic and Financial Indicators.” Look at the interest rates for the United States, Japan, China, Britain, Canada, and the Euro area. For each country (treating the Euro area as a country), calculate the spreads as that country’s interest rate minus the U.S. interest rate.

b. From the uncovered interest parity condition, the spreads from part (a) are the annualized expected appreciation rates of the dollar against other currencies. To calculate the 10-year expected appreciation, you must compound. (So, if x is the spread, the 10-year expected appreciation is \[ \left( 1 + x \right)^{10} - 1 \]. Be careful about decimal points.) Is the dollar expected to depreciate by much in nominal terms against any currency other than the yen?

c. Given your answer to part (b), if we accept that significant real depreciation of the dollar is likely in the next decade, how must it be accomplished? Does your answer seem plausible?

d. What do your answers to parts (b) and (c) suggest about the relative strength of demand for dollar assets, independent of the exchange rate? You may want to review Problem 7 before answering this question.
APPENDIX: Fixed Exchange Rates, Interest Rates, and Capital Mobility

The assumption of perfect capital mobility is a good approximation of what happens in countries with highly developed financial markets and few capital controls, such as the United States, the United Kingdom, Japan, and the Euro area. But this assumption is more questionable in countries that have less developed financial markets or have capital controls in place. In these countries, domestic financial investors may have neither the savvy nor the legal right to buy foreign bonds when domestic interest rates are low. The central bank may thus be able to decrease the interest rate while maintaining a given exchange rate.

To look at these issues, we need to have another look at the balance sheet of the central bank. In Chapter 4, we assumed the only asset held by the central bank was domestic bonds. In an open economy, the central bank actually holds two types of assets: (1) domestic bonds and (2) foreign exchange reserves, which we shall think of as foreign currency—although they also take the form of foreign bonds or foreign interest-paying assets. Think of the balance sheet of the central bank as represented in Figure 1:

On the asset side are bonds and foreign exchange reserves, and on the liability side is the monetary base. There are now two ways in which the central bank can change the monetary base: either by purchases or sales of bonds in the bond market, or by purchases or sales of foreign currency in the foreign exchange market. (If you did not read Section 4-4 in Chapter 4, replace “monetary base” with “money supply” and you will still get the basic argument.)

Perfect Capital Mobility and Fixed Exchange Rates
Consider first the effects of an open market operation under the joint assumptions of perfect capital mobility and fixed exchange rates (the assumptions we made in the last section of this chapter).

- Assume the domestic interest rate and the foreign interest rate are initially equal, so \( i = i^* \). Now suppose the central bank embarks on an expansionary open market operation, buying bonds in the bond market in amount \( \Delta B \) and creating money—increasing the monetary base—in exchange. This purchase of bonds leads to a decrease in the domestic interest rate, \( i \). This is, however, only the beginning of the story.
- Now that the domestic interest rate is lower than the foreign interest rate, financial investors prefer to hold foreign bonds. To buy foreign bonds, they must first buy foreign currency. They then go to the foreign exchange market and sell domestic currency for foreign currency.

If the central bank did nothing, the price of domestic currency would fall, and the result would be a depreciation. Under its commitment to a fixed exchange rate, the central bank cannot allow the currency to depreciate. So it must intervene in the foreign exchange market and sell foreign currency for domestic currency. As it sells foreign currency and buys domestic money, the monetary base decreases.

How much foreign currency must the central bank sell? It must keep selling until the monetary base is back to its pre-open market operation level, so the domestic interest rate is again equal to the foreign interest rate. Only then are financial investors willing to hold domestic bonds.

How long do all these steps take? Under perfect capital mobility, all this may happen within minutes or so of the original open market operation. After these steps, the balance sheet of the central bank looks as represented in Figure 2. Bond holdings are up by \( \Delta B \), reserves of foreign currency are down by \( \Delta B \), and the monetary base is unchanged, having gone up by \( \Delta B \) in the open market operation and down by \( \Delta B \) as a result of the sale of foreign currency in the foreign exchange market.

Let’s summarize: Under fixed exchange rates and perfect capital mobility, the only effect of the open market operation is to change the composition of the central bank’s balance sheet but not the monetary base (nor the interest rate.)

Imperfect Capital Mobility and Fixed Exchange Rates
Let’s now move away from the assumption of perfect capital mobility. Suppose it takes some time for financial investors to shift between domestic bonds and foreign bonds.

Now an expansionary open market operation can initially bring the domestic interest rate below the foreign interest rate. But over time, investors shift to foreign bonds, leading to an increase in the demand for foreign currency in the foreign exchange market. To avoid a depreciation of the domestic currency, the central bank must again stand ready to sell foreign currency and buy domestic currency. Eventually, the central bank buys enough domestic currency to offset the effects of the initial open market operation. The monetary base is back to its pre-open market operation level, and so is the interest rate.

### Figure 1

<table>
<thead>
<tr>
<th>Balance Sheet of the Central Bank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assets</strong></td>
</tr>
<tr>
<td>Bonds</td>
</tr>
<tr>
<td>Foreign exchange reserves</td>
</tr>
</tbody>
</table>

### Figure 2

<table>
<thead>
<tr>
<th>Balance Sheet of the Central Bank after an Open Market Operation, and the Induced Intervention in the Foreign Exchange Market</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assets</strong></td>
</tr>
<tr>
<td>Bonds: ( \Delta B )</td>
</tr>
<tr>
<td>Reserves: ( -\Delta B )</td>
</tr>
</tbody>
</table>
rate. The central bank holds more domestic bonds and smaller reserves of foreign currency.

The difference between this case and the case of perfect capital mobility is that, by accepting a loss in foreign exchange reserves, the central bank is now able to decrease interest rates for some time. If it takes just a few days for financial investors to adjust, the trade-off can be very unattractive—as many countries that have suffered large losses in reserves without much effect on the interest rate have discovered at their expense. But, if the central bank can affect the domestic interest rate for a few weeks or months, it may, in some circumstances, be willing to do so.

Now let’s deviate further from perfect capital mobility. Suppose, in response to a decrease in the domestic interest rate, financial investors are either unwilling or unable to move much of their portfolio into foreign bonds. For example, there are administrative and legal controls on financial transactions, making it illegal or very expensive for domestic residents to invest outside the country. This is the relevant case for a number of emerging economies, from Latin America to China.

After an expansionary open market operation, the domestic interest rate decreases, making domestic bonds less attractive. Some domestic investors move into foreign bonds, selling domestic currency for foreign currency. To maintain the exchange rate, the central bank must buy domestic currency and supply foreign currency. However, the foreign exchange intervention by the central bank may now be small compared to the initial open market operation. And, if capital controls truly prevent investors from moving into foreign bonds at all, there may be no need for such a foreign exchange intervention.

Even leaving this extreme case aside, the net effects of the initial open market operation and the following foreign exchange interventions are likely to be an increase in the monetary base; a decrease in the domestic interest rate; an increase in the central bank’s bond holdings; and some—but limited—loss in reserves of foreign currency. With imperfect capital mobility, a country has some freedom to move the domestic interest rate while maintaining its exchange rate. This freedom depends primarily on three factors:

- The degree of development of its financial markets, and the willingness of domestic and foreign investors to shift between domestic assets and foreign assets.
- The degree of capital controls it is able to impose on both domestic and foreign investors.
- The amount of foreign exchange reserves it holds: The higher the reserves it has, the more it can afford the loss in reserves it is likely to sustain if it decreases the interest rate at a given exchange rate.

With the large movements in capital flows we documented earlier in the chapter, all of these issues are hot topics. Many countries are considering a more active use of capital controls than in the past. Many countries are also accumulating large reserves as a precaution against large capital outflows.

**Key Term**

foreign-exchange reserves, 442
In July 1944, representatives of 44 countries met in Bretton Woods, New Hampshire, to design a new international monetary and exchange rate system. The system they adopted was based on fixed exchange rates, with all member countries other than the United States fixing the price of their currency in terms of dollars. In 1973, a series of exchange rate crises brought an abrupt
end to the system—and an end to what is now called “the Bretton Woods period.” Since then, the world has been characterized by many exchange rate arrangements. Some countries operate under flexible exchange rates; some operate under fixed exchange rates; some go back and forth between regimes. Which exchange rate regime to choose is one of the most debated issues in macroeconomics and, as the cartoon suggests, a decision facing every country in the world. This chapter discusses this issue.

Section 21-1 looks at the medium run. It shows that, in sharp contrast to the results we derived for the short run in Chapter 20, an economy ends up with the same real exchange rate and output level in the medium run, regardless of whether it operates under fixed exchange rates or flexible exchange rates. This obviously does not make the exchange rate regime irrelevant—the short run matters very much—but it is an important extension and qualification to our previous analysis.

Section 21-2 takes another look at fixed exchange rates and focuses on exchange rate crises. During a typical exchange rate crisis, a country operating under a fixed exchange rate is forced, often under dramatic conditions, to abandon its parity and to devalue. Such crises were behind the breakdown of the Bretton Woods system. They rocked the European Monetary System in the early 1990s, and were a major element of the Asian Crisis of the late 1990s. It is important to understand why they happen, and what they imply.

Section 21-3 takes another look at flexible exchange rates and focuses on the behavior of exchange rates under a flexible exchange rate regime. It shows that the behavior of exchange rates and the relation of the exchange rate to monetary policy are, in fact more complex than we assumed in Chapter 20. Large fluctuations in the exchange rate, and the difficulty of using monetary policy to affect the exchange rate, make a flexible exchange rate regime less attractive than it appeared to be in Chapter 20.

Section 21-4 puts all these conclusions together and reviews the case for flexible or fixed rates. It discusses two recent and important developments: the use of a common currency in much of Europe and the move toward strong forms of fixed exchange rate regimes, from currency boards to dollarization.

21-1 The Medium Run

When we focused on the short run in Chapter 20, we drew a sharp contrast between the behavior of an economy with flexible exchange rates and an economy with fixed exchange rates.

- Under flexible exchange rates, a country that needed to achieve a real depreciation (for example, to reduce its trade deficit or to get out of a recession) could do so by relying on an expansionary monetary policy to achieve both a lower interest rate and a decrease in the exchange rate—a depreciation.

- Under fixed exchange rates, a country lost both of these instruments: By definition, its nominal exchange rate was fixed and thus could not be adjusted. Moreover, the fixed exchange rate and the interest parity condition implied that the country could not adjust its interest rate; the domestic interest rate had to remain equal to the foreign interest rate.

This appeared to make a flexible exchange rate regime much more attractive than a fixed exchange rate regime: Why should a country give up two macroeconomic instruments—the exchange rate and the interest rate? As we now shift focus from the short run to the medium run, you shall see that this earlier conclusion needs to be qualified. Although our conclusions about the short run were valid, we shall see that, in the medium run, the difference between the two regimes fades away. More specifically, in
the medium run, the economy reaches the same real exchange rate and the same level of output whether it operates under fixed or under flexible exchange rates.

The intuition for this result is actually easy to give. Recall the definition of the real exchange rate:

$$\epsilon = \frac{E P}{P^*}$$

The real exchange rate, $\epsilon$, is equal to the nominal exchange rate, $E$ (the price of domestic currency in terms of foreign currency) times the domestic price level, $P$, divided by the domestic price level, $P^*$. There are, therefore, two ways in which the real exchange rate can adjust:

- Through a change in the nominal exchange rate $E$: This can only be done under flexible exchange rates. And if we assume the domestic price level, $P$, and the foreign price level, $P^*$, do not change in the short run, it is the only way to adjust the real exchange rate in the short run.

- Through a change in the domestic price level, $P$, relative to the foreign price level, $P^*$. In the medium run, this option is open even to a country operating under a fixed (nominal) exchange rate. And this is indeed what happens under fixed exchange rates: The adjustment takes place through the price level rather than through the nominal exchange rate.

Let us go through this argument step by step. To begin, let’s derive the aggregate demand and aggregate supply relations for an open economy under a fixed exchange rate.

### Aggregate Demand under Fixed Exchange Rates

In an open economy with fixed exchange rates, we can write the aggregate demand relation as

$$Y = Y\left(\frac{E P}{P^*}, G, T\right)$$

(21.1)

Output, $Y$, depends on the real exchange rate, $E P / P^*$ ($E$ denotes the fixed nominal exchange rate; $P$ and $P^*$ denote the domestic and foreign price levels, respectively), government spending, $G$, and taxes, $T$. An increase in the real exchange rate—a real appreciation—leads to a decrease in output. An increase in government spending leads to an increase in output; an increase in taxes to a decrease in output.

The derivation of equation (21.1) is better left to Appendix 1 at the end of this chapter, which is titled “Deriving Aggregate Demand under Fixed Exchange Rates.” The intuition behind the equation is straightforward, however:

Recall that, in the closed economy, the aggregate demand relation took the same form as equation (21.1), except for the presence of the real money stock, $M / P$, instead of the real exchange rate, $E P / P^*$.

The reason for the presence of $M / P$ in the closed economy was the following: By controlling the money supply, the central bank could change the interest rate and affect output. In an open economy, and under fixed exchange rates and perfect capital mobility, the central bank can no longer change the interest rate—which is pinned down by the foreign interest rate. Put another way, under fixed exchange rates, the central bank gives up monetary policy as a policy instrument. This is why the money stock no longer appears in the aggregate demand relation.
At the same time, the fact that the economy is open implies that we must include a variable which we did not include when we looked at the closed economy earlier, namely the real exchange rate, $\frac{E}{P}/P^*$. As we saw in Chapter 19, an increase in the real exchange rate leads to a decrease in the demand for domestic goods and thus a decrease in output. Conversely, a decrease in the real exchange rate leads to an increase in output.

Note that, just as in the closed economy, the aggregate demand relation (21.1) implies a negative relation between the price level and output. But, while the sign of the effect of the price level on output remains the same, the channel is very different:

- In the closed economy, the price level affects output through its effect on the real money stock and, in turn, its effect on the interest rate.
- In the open economy under fixed exchange rates, the price level affects output through its effect on the real exchange rate. Given the fixed nominal exchange rate, $E$, and the foreign price level, $P^*$, an increase in the domestic price level, $P$, leads to an increase in the real exchange rate $\frac{EP}{P^*}$—a real appreciation. This real appreciation leads to a decrease in the demand for domestic goods, and, in turn, to a decrease in output. Put simply: An increase in the price level makes domestic goods more expensive, thus decreasing the demand for domestic goods, in turn decreasing output.

**Equilibrium in the Short Run and in the Medium Run**

The aggregate demand curve associated with equation (21.1) is drawn as the $AD$ curve in Figure 21-1. It is downward sloping: An increase in the price level decreases output. As always, the relation is drawn for given values of the other variables; in this case for given values of $E$, $P^*$, $G$, and $T$.

For the aggregate supply curve, we rely on the relation we derived in the core.

Going back to the aggregate supply relation we derived in Chapter 7, equation (7.2):

$$P = P^e (1 + m) F\left(1 - \frac{Y}{L}, z\right)$$

(21.2)
The price level, $P$, depends on the expected price level, $P^e$, and on the level of output, $Y$. Recall the two mechanisms at work:

- The expected price level matters because it affects nominal wages, which in turn affect the price level.
- Higher output matters because it leads to higher employment, which leads to lower unemployment, which leads to higher wages, which lead to a higher price level.

The aggregate supply curve is drawn as the $AS$ curve in Figure 21-1 for a given value of the expected price level. It is upward sloping: Higher output leads to a higher price level.

The short-run equilibrium is given by the intersection of the aggregate demand curve and the aggregate supply curve, point $A$ in Figure 21-1. As was the case in the closed economy, there is no reason why the short-run equilibrium level of output, $Y$, should be equal to the natural level of output, $Y_n$. As the figure is drawn, $Y$ is less than $Y_n$, so output is below the natural level of output.

What happens over time? The basic answer is familiar from our earlier look at adjustment in a closed economy, and is shown in Figure 21-2. As long as output remains below the natural level of output, the aggregate supply shifts down to $AS'$. The reason: When output is below the natural level of output, the price level turns out to be lower than was expected. This leads wage setters to revise their expectation of the price level downward, leading to a lower price level at a given level of output—hence, the downward shift of the aggregate supply curve. So, starting from point $A$, the economy moves over time along the aggregate demand curve, until it reaches point $B$. At point $B$, output is equal to the natural level of output. The price level is lower than it was at point $A$; by implication the real exchange rate is lower than it was at point $A$.

In words: As long as output is below the natural level of output, the price level decreases. The decrease in the price level over time leads to a steady real depreciation. This real depreciation then leads to an increase in output until output has returned to its natural level.

**Figure 21-2**

**Adjustment under Fixed Exchange Rates**

The aggregate supply curve shifts down over time, leading to a decrease in the price level, to a real depreciation, and to an increase in output. The process ends when output has returned to its natural level.
In the medium run, despite the fact that the nominal exchange rate is fixed, the economy still achieves the real depreciation needed to return output to its natural level. This is an important qualification to the conclusions we reached in the previous chapter—where we were focusing only on the short run:

- In the short run, a fixed nominal exchange rate implies a fixed real exchange rate.
- In the medium run, the real exchange rate can adjust even if the nominal exchange rate is fixed. This adjustment is achieved through movements in the price level.

The Case For and Against a Devaluation

The result that, even under fixed exchange rates, the economy returns to the natural level of output in the medium run is important. But it does not eliminate the fact that the process of adjustment may be long and painful. Output may remain too low and unemployment may remain too high for a long time.

Are there faster and better ways to return output to normal? The answer, within the model we have just developed, is a clear yes.

Suppose that the government decides, while keeping the fixed exchange rate regime, to allow for a one-time devaluation. For a given price level, a devaluation (a decrease in the nominal exchange rate) leads to a real depreciation (a decrease in the real exchange rate), and therefore to an increase in output. In other words, a devaluation shifts the aggregate demand curve to the right: Output is higher at a given price level.

This has a straightforward implication: A devaluation of the right size can take the economy directly from \( Y \) to \( Y_n \). This is shown in Figure 21-3. Suppose the economy is initially at point \( A \), the same point \( A \) as in Figure 21-2. The right size depreciation shifts the aggregate demand curve up from \( AD \) to \( AD' \), taking the equilibrium from points \( A \) to \( C \). At point \( C \), output is equal to the natural level of output, \( Y_n \), and the real exchange rate is the same as at point \( B \). (We know this because output is the same at points \( B \) and \( C \). From equation (21.1), and without changes in \( G \) or \( T \), this implies that the real exchange rate must also be the same.)

**Figure 21-3**

*Adjustment with a Devaluation*

The right size devaluation can shift aggregate demand to the right, leading the economy to go to point \( C \). At point \( C \), output is back to the natural level of output.
That the devaluation of the “right size” can return output to the natural level of output right away sounds too good to be true—and, in reality, it is. Achieving the “right size” devaluation—the devaluation that takes output to $Y_n$ right away—is easier to achieve in a graph than in reality:

- In contrast to our simple aggregate demand relation (21.1), the effects of the depreciation on output do not happen right away: As you saw in Chapter 19, the initial effects of a depreciation on output can be contractionary, as people pay more for imports, and the quantities of imports and exports have not yet adjusted.

- Also, in contrast to our simple aggregate supply relation (21.2), there is likely to be a direct effect of the devaluation on the price level. As the price of imported goods increases, the price of a consumption basket increases. This increase is likely to lead workers to ask for higher nominal wages, forcing firms to increase their prices as well.

But these complications do not affect the basic conclusion: A devaluation can hasten the return of output to its natural level. And so, whenever a country under fixed exchange rates faces either a large trade deficit or a severe recession, there is a lot of political pressure either to give up the fixed exchange rate regime altogether, or, at least, to have a one-time devaluation. Perhaps the most forceful presentation of this view was made 85 years ago by Keynes, who argued against Winston Churchill’s decision to return the British pound in 1925 to its pre–World War I parity with gold. His arguments are presented in the Focus box “The Return of Britain to the Gold Standard: Keynes versus Churchill.” Most economic historians believe that history proved Keynes right, and that overvaluation of the pound was one of the main reasons for Britain’s poor economic performance after World War I.

Those who oppose a shift to flexible exchange rates or who oppose a devaluation argue that there are good reasons to choose fixed exchange rates, and that too much willingness to devalue defeats the purpose of adopting a fixed exchange rate regime in the first place. They argue that too much willingness on the part of governments to consider devaluations actually leads to an increased likelihood of exchange rate crises. To understand their arguments, we now turn to these crises: what triggers them, and what their implications might be.

### 21-2 Exchange Rate Crises under Fixed Exchange Rates

Suppose a country has chosen to operate under a fixed exchange rate. Suppose also that financial investors start believing there may soon be an exchange rate adjustment—either a devaluation or a shift to a flexible exchange rate regime accompanied by a depreciation.

We just saw why this might be the case:

- The real exchange rate may be too high. Or, put another way, the domestic currency may be overvalued. In this case, a real depreciation is called for. Although this could be achieved in the medium run without a devaluation, financial investors might conclude that the government will take the quickest way out—and devalue.

Such an overvaluation often happens in countries that peg their nominal exchange rate to the currency of a country with lower inflation. Higher relative inflation implies a steadily increasing price of domestic goods relative to foreign goods, a steady real appreciation, and so a steady worsening of the trade position. As time passes, the need for an adjustment of the real exchange rate increases, and
In 1925, Britain decided to return to the gold standard. The gold standard was a system in which each country fixed the price of its currency in terms of gold and stood ready to exchange gold for currency at the stated parity. This system implied fixed exchange rates between countries.

The gold standard had been in place from 1870 until World War I. Because of the need to finance the war, and to do so in part by money creation, Britain suspended the gold standard in 1914. In 1925, Winston Churchill, then Britain’s Chancellor of the Exchequer (the British equivalent of Secretary of the Treasury in the United States), decided to return to the gold standard, and to return to it at the pre-war parity—that is, at the pre-war value of the pound in terms of gold. But, because prices had increased faster in Britain than in many of its trading partners, returning to the pre-war parity implied a large real appreciation: At the same nominal exchange rate as before the war, British goods were now relatively more expensive relative to foreign goods. (Go back to the definition of the real exchange rate, \( \varepsilon = \frac{EP}{P^*} \): The price level in Britain, \( P \), had increased more than the foreign price level, \( P^* \). At a given nominal exchange rate, \( E \), this implied that \( \varepsilon \) was higher, that Britain suffered from a real appreciation.)

Keynes severely criticized the decision to return to the pre-war parity. In The Economic Consequences of Mr. Churchill, a book he published in 1925, Keynes argued as follows: If Britain were going to return to the gold standard, it should have done so at a lower price of currency in terms of gold; that is, at a lower nominal exchange rate than the pre-war nominal exchange rate. In a newspaper article, he articulated his views as follows:

“...There remains, however, the objection to which I have never ceased to attach importance, against the return to gold in actual present conditions, in view of the possible consequences on the state of trade and employment. I believe that our price level is too high, if it is converted to gold at the par of exchange, in relation to gold prices elsewhere; and if we consider the prices of those articles only which are not the subject of international trade, and of services, i.e., wages, we shall find that these are materially too high—not less than 5 per cent, and probably 10 per cent. Thus, unless the situation is saved by a rise of prices elsewhere, the Chancellor is committing us to a policy of forcing down money wages by perhaps 2 shillings in the Pound.

I do not believe that this can be achieved without the gravest danger to industrial profits and industrial peace. I would much rather leave the gold value of our currency where it was some months ago than embark on a struggle with every trade union in the country to reduce money wages. It seems wiser and simpler and saner to leave the currency to find its own level for some time longer rather than force a situation where employers are faced with the alternative of closing down or of lowering wages, cost what the struggle may.

For this reason, I remain of the opinion that the Chancellor of the Exchequer has done an ill-judged thing—ill judged because we are running the risk for no adequate reward if all goes well.”

Keynes’s prediction turned out to be right. While other countries were growing, Britain remained in recession for the rest of the decade. Most economic historians attribute a good part of the blame to the initial overvaluation.

Source: “The Nation and Athenaeum,” May 2, 1925

The expression to let a currency “float” is to allow a move from a fixed to a flexible exchange rate regime. A floating exchange rate regime is the same as a flexible exchange rate regime.

Because it is more convenient, we use the approximation, equation (18.4), rather than the original interest parity condition, equation (18.2).

Financial investors become more and more nervous. They start thinking that a devaluation might be coming.

Internal conditions may call for a decrease in the domestic interest rate. As we have seen, a decrease in the domestic interest rate cannot be achieved under fixed exchange rates. But it can be achieved if the country is willing to shift to a flexible exchange rate regime. If a country lets the exchange rate float and then decreases its domestic interest rate, we know from Chapter 20 that this will trigger a decrease in the nominal exchange rate—a nominal depreciation.

As soon as financial markets believe a devaluation may be coming, then maintain the exchange rate requires an increase—often a large one—in the domestic interest rate.

To see this, return to the interest parity condition we derived in Chapter 18:

\[
i_t = i_t^* - \frac{(E_{t+1}^* - E_t)}{E_t}
\]

(21.3)
In Chapter 18, we interpreted this equation as a relation among the one-year domestic and foreign nominal interest rates, the current exchange rate, and the expected exchange rate a year hence. But the choice of one year as the period was arbitrary. The relation holds over a day, a week, a month. If financial markets expect the exchange rate to be 2% lower a month from now, they will hold domestic bonds only if the one-month domestic interest rate exceeds the one-month foreign interest rate by 2% (or, if we express interest rates at an annual rate, if the annual domestic interest rate exceeds the annual foreign interest rate by \(2\% \times 12 = 24\%\)).

Under fixed exchange rates, the current exchange rate, \(E_t\), is set at some level, say \(E_t = E\). If markets expect the parity will be maintained over the period, then \(E_{t+1} = E\), and the interest parity condition simply states that the domestic and the foreign interest rates must be equal.

Suppose, however, participants in financial markets start anticipating a devaluation—a decrease in the exchange rate. Suppose they believe that, over the coming month, there is a 75% chance the parity will be maintained and a 25% chance there will be a 20% devaluation. The term \((E_{t+1} - E_t)/E_t\) in the interest parity equation (21.3), which we assumed equal to zero earlier, now equals

\[
0.75 \times 0\% + 0.25 \times (-20\%) = -5\% \quad \text{(a 75% chance of no change plus a 25% chance of a devaluation of 20%).}
\]

This implies that, if the central bank wants to maintain the existing parity, it must now set a monthly interest rate 5% higher than before—60% higher at an annual rate (12 months \(\times 5\%\) per month); 60% is the interest differential needed to convince investors to hold domestic bonds rather than foreign bonds! Any smaller interest differential, and investors will not want to hold domestic bonds.

What, then, are the choices confronting the government and the central bank?

- First, the government and the central bank can try to convince markets they have no intention of devaluing. This is always the first line of defense: Communiqués are issued, and prime ministers go on TV to reiterate their absolute commitment to the existing parity. But words are cheap, and they rarely convince financial investors.
- Second, the central bank can increase the interest rate, but by less than would be needed to satisfy equation (21.3)—in our example, by less than 60%. Although domestic interest rates are high, they are not high enough to fully compensate for the perceived risk of devaluation. This action typically leads to a large capital outflow, because financial investors still prefer to get out of domestic bonds and into foreign bonds. They sell domestic bonds, getting the proceeds in domestic currency. They then go to the foreign exchange market to sell domestic currency for foreign currency, in order to buy foreign bonds. If the central bank did not intervene in the foreign exchange market, the large sales of domestic currency for foreign currency would lead to a depreciation. If it wants to maintain the exchange rate, the central bank must therefore stand ready to buy domestic currency and sell foreign currency at the current exchange rate. In doing so, it often loses most of its reserves of foreign currency. (The mechanics of central bank intervention were described in the appendix to Chapter 20.)
- Eventually—after a few hours or a few weeks—the choice for the central bank becomes either to increase the interest rate enough to satisfy equation (21.3) or to validate the market’s expectations and devalue. Setting a very high short-term domestic interest rate can have a devastating effect on demand and on output—no firm wants to invest; no consumer wants to borrow when interest rates are very high. This course of action makes sense only if (1) the perceived probability of a devaluation is small, so the interest rate does not have to be too high; and (2) the government believes markets will soon become convinced that no devaluation is coming, allowing domestic interest rates to decrease. Otherwise, the only option is to devalue. (All in most countries, the government is formally in charge of choosing the parity, the central bank is formally in charge of maintaining it. In practice, choosing and maintaining the parity are joint responsibilities of the government and the central bank.

In the summer of 1998, Boris Yeltsin announced that the Russian government had no intention of devaluing the ruble. Two weeks later, the ruble collapsed.
The 1992 EMS Crisis

An example of the problems we discussed in this section is the exchange rate crisis that shook the European Monetary System in the early 1990s.

At the start of the 1990s, the European Monetary System (EMS) appeared to work well. The EMS had started in 1979. It was an exchange rate system based on fixed parities with bands: Each member country (among them, France, Germany, Italy, and, beginning in 1990, the United Kingdom) had to maintain its exchange rate vis-à-vis all other member countries within narrow bands. The first few years had been rocky, with many realignments—adjustment of parities—among member countries. From 1987 to 1992, however, there were only two realignments, and there was increasing talk about narrowing the bands further and even moving to the next stage—to the adoption of a common currency.

In 1992, however, financial markets became increasingly convinced that more realignments were soon to come. The reason was one we have already seen in Chapter 20, namely the macroeconomic implications of Germany’s reunification. Because of the pressure on demand coming from reunification, the Bundesbank (the German central bank) was maintaining high interest rates to avoid too large an increase in output and an increase in inflation in Germany. While Germany’s EMS partners needed lower interest rates to reduce a growing unemployment problem, they had to match the German interest rates to maintain their EMS parities. To financial markets, the position of Germany’s EMS partners looked increasingly untenable. Lower interest rates outside Germany, and thus devaluations of many currencies relative to the Deutsche Mark (DM), appeared increasingly likely.

Throughout 1992, the perceived probability of a devaluation forced a number of EMS countries to maintain higher nominal interest rates than even those in Germany. Still, the first major crisis did not come until September 1992.

In early September 1992, the belief that a number of countries were soon going to devalue led to speculative attacks on a number of currencies, with financial investors selling in anticipation of an oncoming devaluation. All the lines of defense described earlier were used by the monetary authorities and the governments of the countries under attack. First, solemn communiqués were issued, but with no discernible effect. Then, interest rates were increased. For example, Sweden’s overnight interest rate (the rate for lending and borrowing overnight) increased to 500% (expressed at an annual rate)! But interest rates were not increased by enough to prevent capital outflows and large losses of foreign exchange reserves by the central banks under pressure.

At that point, different countries took different courses of action: Spain devalued its exchange rate. Italy and the

![Figure 1: Exchange Rates of Selected European Countries Relative to the Deutsche Mark, January 1992 to December 1993](https://example.com/image.png)

Source: IMF database
these steps were very much in evidence during the exchange rate crisis which affected much of Western Europe in 1992. See the Focus box “The 1992 EMS Crisis.”

To summarize: Expectations that a devaluation may be coming can trigger an exchange rate crisis. Faced with such expectations, the government has two options:

- Give in and devalue, or
- Fight and maintain the parity, at the cost of very high interest rates and a potential recession. Fighting may not work anyway: The recession may force the government to change policy later on, or force the government out of office.

An interesting twist here is that a devaluation can occur even if the belief that a devaluation was coming was initially groundless. In other words, even if the government initially has no intention of devaluing, it might be forced to do so if financial markets believe that it will devalue: The cost of maintaining the parity would be a long period of high interest rates and a recession; the government might prefer to devalue instead.

21-3 Exchange Rate Movements under Flexible Exchange Rates

In the model we developed in Chapter 20, there was a simple relation between the interest rate and the exchange rate: The lower the interest rate, the lower the exchange rate. This implied that a country that wanted to maintain a stable exchange rate just had to maintain its interest rate close to the foreign interest rate. A country that wanted to achieve a given depreciation just had to decrease its interest rate by the right amount.

In reality, the relation between the interest rate and the exchange rate is not so simple. Exchange rates often move even in the absence of movements in interest rates. Furthermore, the size of the effect of a given change in the interest rate on the exchange rate is hard to predict. This makes it much harder for monetary policy to achieve its desired outcome.
To see why things are more complicated, we must return once again to the interest parity condition we derived in Chapter 18 (equation (18.2)):

\[(1 + i_t) = (1 + \bar{r}_t^e) \left( \frac{E_t}{E_{t+1}^e} \right)\]

As we did in Chapter 20 (equation (20.5)), multiply both sides by \(E_{t+1}^e\), and reorganize to get

\[E_t = \frac{1 + i_t}{1 + \bar{r}_t^e} E_{t+1}^e \quad (21.4)\]

Think of the time period (from \(t\) to \(t + 1\)) as one year. The exchange rate this year depends on the one-year domestic interest rate, the one-year foreign interest rate, and the exchange rate expected for next year.

We assumed in Chapter 20 that the expected exchange rate next year, \(E_{t+1}^e\), was constant. But this was a simplification. The exchange rate expected one year hence is not constant. Using equation (21.4), but now for next year, it is clear that the exchange rate next year will depend on next year’s one-year domestic interest rate, the one-year foreign interest rate, the exchange rate expected for the year after, and so on. So, any change in expectations of current and future domestic and foreign interest rates, as well as changes in the expected exchange rate in the far future, will affect the exchange rate today.

Let’s explore this more closely. Write equation (21.4) for year \(t + 1\) rather than for year \(t\):

\[E_{t+1} = \frac{1 + i_{t+1}}{1 + \bar{r}_{t+1}^e} E_{t+2}^e\]

The exchange rate in year \(t + 1\) depends on the domestic interest rate and the foreign interest rate for year \(t + 1\), as well as on the expected future exchange rate in year \(t + 2\). So, the expectation of the exchange rate in year \(t + 1\), held as of year \(t\), is given by:

\[E_{t+1}^e = \frac{1 + \bar{r}_{t+1}^e}{1 + \bar{r}_{t+1}^e} E_{t+2}^e\]

Replacing \(E_{t+1}^e\) in equation (21.4) with the expression above gives:

\[E_t = \frac{(1 + i_t)(1 + \bar{r}_{t+1}^e)}{(1 + \bar{r}_t^e)(1 + \bar{r}_{t+1}^e)} E_{t+2}^e\]

The current exchange rate depends on this year’s domestic and foreign interest rates, on next year’s expected domestic and foreign interest rates, and on the expected exchange rate two years from now. Continuing to solve forward in time in the same way (by replacing \(E_{t+2}^e, E_{t+3}^e\), and so on until, say, year \(t + n\)), we get:

\[E_t = \frac{(1 + i_t)(1 + \bar{r}_{t+1}^e) \cdots (1 + \bar{r}_{t+n}^e)}{(1 + \bar{r}_t^e)(1 + \bar{r}_{t+1}^e) \cdots (1 + \bar{r}_{t+n}^e)} E_{t+n+1}^e \quad (21.5)\]

Suppose we take \(n\) to be large, say 10 years (equation (21.5) holds for any value of \(n\)). This relation tells us that the current exchange rate depends on two sets of factors:

- Current and expected domestic and foreign interest rates for each year over the next 10 years.
- The expected exchange rate 10 years from now.

For some purposes, it is useful to go further and derive a relation among current and expected future domestic and foreign real interest rates, the current real exchange rate, and the expected future real exchange rate. This is done in Appendix 2 at the end of this chapter. (The derivation is not much fun, but it is a useful way of brushing up on
the relation between real interest rates and nominal interest rates, and real exchange rates and nominal exchange rates.) Equation (21.5) is sufficient to make three important points, each outlined in more detail below:

- The level of today's exchange rate will move one for one with the future expected exchange rate.
- Today's exchange rate will move when future expected interest rates move in either country.
- Because today's exchange rate moves with any change in expectations, the exchange rate will be volatile, that is, move frequently and perhaps by large amounts.

### Exchange Rates and the Current Account

Any factor that moves the expected future exchange rate, $E_{t+n}^e$, moves the current exchange rate, $E_t$. Indeed, if the domestic interest rate and the foreign interest rate are expected to be the same in both countries from $t$ to $t + n$, the fraction on the right in equation (21.5) is equal to 1, so the relation reduces to $E_t = E_{t+n}^e$. In words: The effect of any change in the expected future exchange rate on the current exchange rate is one-for-one.

If we think of $n$ as large (say 20 years or more), we can think of $E_{t+n}^e$ as the exchange rate required to achieve current account balance in the medium or long run: Countries cannot borrow—run a current account deficit—forever, and will not want to lend—run a current account surplus—forever either. Thus, any news that affects forecasts of the current account balance in the future is likely to have an effect on the expected future exchange rate, and in turn on the exchange rate today. For example, the announcement of a larger-than-expected current account deficit may lead investors to conclude that a depreciation will eventually be needed to repay the increased debt. Thus, $E_{t+n}^e$ will decrease, leading in turn to a decrease in $E_t$ today.

### Exchange Rates and Current and Future Interest Rates

Any factor that moves current or expected future domestic or foreign interest rates between years $t$ and $t + n$ moves the current exchange rate, too. For example, given foreign interest rates, an increase in current or expected future domestic interest rates leads to an increase in $E_t$—an appreciation.

This implies that any variable that causes investors to change their expectations of future interest rates will lead to a change in the exchange rate today. For example, the “dance of the dollar” in the 1980s that we discussed in earlier chapters—the sharp appreciation of the dollar in the first half of the decade, followed by an equally sharp depreciation later—can be largely explained by the movement in current and expected future U.S. interest rates relative to interest rates in the rest of the world during that period. During the first half of the 1980s, tight monetary policy and expansionary fiscal policy combined to increase both U.S. short-term interest rates and long-term interest rates; with the increase in long-term rates reflecting anticipations of high short-term interest rates in the future. This increase in both current and expected future interest rates was, in turn, the main cause of the dollar appreciation. Both fiscal and monetary policy were reversed in the second half of the decade, leading to lower U.S. interest rates and a depreciation of the dollar.

### Exchange Rate Volatility

The third implication follows from the first two. In reality, and in contrast to our analysis in Chapter 20, the relation between the interest rate, $i_t$, and the exchange rate, $E_t$, is anything but mechanical. When the central bank cuts the interest rate, financial markets have to assess whether this action signals a major shift in monetary policy and
the cut in the interest rate is just the first of many such cuts, or whether this cut is just a temporary movement in interest rates. Announcements by the central bank may not be very useful: The central bank itself may not even know what it will do in the future. Typically, it will be reacting to early signals, which may be reversed later. Investors also have to assess how foreign central banks will react: whether they will stay put or follow suit and cut their own interest rates. All this makes it much harder to predict what the effect of the change in the interest rate will be on the exchange rate.

Let’s be more concrete. Go back to equation (21.5). Assume that \( E_{t+n} = 1 \). Assume that current and expected future domestic interest rates, and current and expected future foreign interest rates, are all equal to 5%. The current exchange rate is then given by:

\[
E_t = \frac{(1.05)^n}{(1.05)^n} 1 = 1
\]

Now consider a monetary expansion, which decreases the current domestic interest rate, \( i_t \), from 5% to 3%. Will this lead to a decrease in \( E_t \)—to a depreciation—and if so by how much? The answer: It all depends:

Suppose the interest rate is expected to be lower for just one year, so the \( n - 1 \) expected future interest rates remain unchanged. The current exchange rate then decreases to:

\[
E_t = \frac{(1.03)(1.05)^{n-1}}{(1.05)^n} = \frac{1.03}{1.05} = 0.98
\]

The expansionary monetary policy leads to a decrease in the exchange rate—a depreciation—of only 2%.

Suppose instead that, when the current interest rate declines from 5% to 3%, investors expect the decline to last for five years (so \( i_{t+4} = \cdots = i_{t+1} = i_t = 3\% \)). The exchange rate then decreases to:

\[
E_t = \frac{(1.03)^5(1.05)^{n-5}}{(1.05)^n} = \frac{(1.03)^5}{(1.05)^5} = 0.90
\]

The expansionary monetary policy now leads to a decrease in the exchange rate—a depreciation—of 10%, a much larger effect.

You can surely think of still more outcomes. Suppose investors had anticipated that the central bank was going to decrease interest rates, and the actual decrease turns out to be smaller than they anticipated. In this case, the investors will revise their expectations of future nominal interest rates upward, leading to an appreciation rather than a depreciation of the currency.

When, at the end of the Bretton Woods period, countries moved from fixed exchange rates to flexible exchange rates, most economists had expected that exchange rates would be stable. The large fluctuations in exchange rates that followed—and have continued to this day—came as a surprise. For some time, these fluctuations were thought to be the result of irrational speculation in foreign exchange markets. It was not until the mid-1970s that economists realized that these large movements could be explained, as we have explained here, by the rational reaction of financial markets to news about future interest rates and the future exchange rate. This has an important implication:

A country that decides to operate under flexible exchange rates must accept the fact that it will be exposed to substantial exchange rate fluctuations over time.
21-4 Choosing between Exchange Rate Regimes

Let us now return to the question that motivates this chapter: Should countries choose flexible exchange rates or fixed exchange rates? Are there circumstances under which flexible rates dominate, and others under which fixed rates dominate?

Much of what we have seen in this and the previous chapter would seem to favor flexible exchange rates:

- Section 21-1 argued that the exchange rate regime may not matter in the medium run. But it is still the case that it matters in the short run. In the short run, countries that operate under fixed exchange rates and perfect capital mobility give up two macroeconomic instruments: the interest rate and the exchange rate. This not only reduces their ability to respond to shocks but can also lead to exchange rate crises.

- Section 21-2 argued that, in a country with fixed exchange rates, the anticipation of a devaluation leads investors to ask for very high interest rates. This in turn makes the economic situation worse and puts more pressure on the country to devalue. This is another argument against fixed exchange rates.

- Section 21-3 introduced one argument against flexible exchange rates, namely that, under flexible exchange rates, the exchange rate is likely to fluctuate a lot and be difficult to control through monetary policy.

On balance, it therefore appears that, from a macroeconomic viewpoint, flexible exchange rates dominate fixed exchange rates. This indeed appears to be the consensus that has emerged among economists and policy makers. The consensus goes like this:

In general, flexible exchange rates are preferable. There are, however, two exceptions: First, when a group of countries is already tightly integrated, a common currency may be the right solution. Second, when the central bank cannot be trusted to follow a responsible monetary policy under flexible exchange rates, a strong form of fixed exchange rates, such as a currency board or dollarization, may be the right solution.

Let us discuss in turn each of these two exceptions.

Common Currency Areas

Countries that operate under a fixed exchange rate regime are constrained to have the same interest rate. But how costly is that constraint? If the countries face roughly the same macroeconomic problems and the same shocks, they would have chosen similar policies in the first place. Forcing them to have the same monetary policy may not be much of a constraint.

This argument was first explored by Robert Mundell, who looked at the conditions under which a set of countries might want to operate under fixed exchange rates, or even adopt a common currency. For countries to constitute an optimal currency area, Mundell argued, they need to satisfy one of two conditions:

- The countries have to experience similar shocks. We just saw the rationale for this: If they experience similar shocks, then they would have chosen roughly the same monetary policy anyway.

- Or, if the countries experience different shocks, they must have high factor mobility. For example, if workers are willing to move from countries that are doing poorly to countries that are doing well, factor mobility rather than macroeconomic policy can allow countries to adjust to shocks. When the unemployment rate is high in a...
country, workers leave that country to take jobs elsewhere, and the unemployment rate in that country decreases back to normal. If the unemployment rate is low, workers come to the country, and the unemployment rate in the country increases back to normal. The exchange rate is not needed.

Following Mundell’s analysis, most economists believe, for example, that the common currency area composed of the 50 states of the United States is close to an optimal currency area. True, the first condition is not satisfied: Individual states suffer from different shocks. California is more affected by shifts in demand from Asia than the rest of the United States. Texas is more affected by what happens to the price of oil, and so on. But the second condition is largely satisfied. There is considerable labor mobility across states in the United States. When a state does poorly, workers leave that state. When it does well, workers come to that state. State unemployment rates quickly return to normal, not because of state-level macroeconomic policy, but because of labor mobility.

Clearly, there are also many advantages of using a common currency. For firms and consumers within the United States, the benefits of having a common currency are obvious; imagine how complicated life would be if you had to change currency every time you crossed a state line. The benefits go beyond these lower transaction costs. When prices are quoted in the same currency, it becomes much easier for buyers to compare prices, and competition between firms increases, benefiting consumers. Given these benefits and the limited macroeconomic costs, it makes good sense for the United States to have a single currency.

In adopting the euro, Europe made the same choice as the United States. When the process of conversion from national currencies to the euro ended in early 2002, the euro became the common currency for 11 European countries. (See the Focus box “The Euro: A Short History.”) The count of countries using the euro at time of writing is now 17. Is the economic argument for this new common currency area as compelling as it is for the United States?

There is little question that a common currency yields for Europe many of the same benefits that it has for the United States. A report by the European Commission estimates that the elimination of foreign exchange transactions within the Euro area leads to a reduction in costs of 0.5% of the combined GDP of these countries. There are also clear signs that the use of a common currency is already increasing competition. When shopping for cars, for example, European consumers now search for the lowest euro price anywhere in the area using the euro. This has already led to a decline in the price of cars in a number of countries.

There is, however, less agreement on whether Europe constitutes an optimal common currency area. This is because neither of the two Mundell conditions appears to be satisfied. Although the future may be different, European countries have experienced very different shocks in the past. Recall our discussion of Germany’s reunification and how differently it affected Germany and the other European countries. Furthermore, labor mobility is very low in Europe and likely to remain low. Workers move much less within European countries than they do within the United States. Because of language and cultural differences among European countries, mobility between countries is even lower.

The risk is, therefore, that one or more Euro area members suffers from a large decline in demand and output but is unable to use either the interest rate or the exchange rate to increase its level of economic activity. As we saw in Section 21-1, the adjustment can still take place in the medium run. But, as we also saw there, this adjustment is likely to be long and painful. This is no longer a hypothetical worry:

Some euro countries, in particular Greece and Portugal, are suffering from low output and a large trade deficit. Without the option of a devaluation, achieving a real