Output, Employment, and Inflation

12

> When the demand for a commodity or service is high relative to the supply of it we expect the price to rise, the rate of rise being greater, the greater the excess demand. Conversely when the demand is low relatively to the supply we expect the price to fall, the rate of fall being greater, the greater the deficiency of demand. It seems plausible that this principle should operate as one of the factors determining the rate of change of money wage rates, which are the price of labour services.
>
> A. W. Phillips[1]

12.1 Overview

A cup of tea in London that cost 5p in 1965 goes for £1.30 or more in 2011. The baguette in Paris which cost 40 centimes (one centime = one hundredth of the old French franc) in 1965 fetches 13 times more in terms of euros 45 years later. Prices seem to grow relentlessly. Yet inflation is not just about changes in the price of tea in Britain and bread in Paris. Inflation measures the rate of increase of the price level, i.e. the number of units of pounds or euros paid for a bundle of goods. It is easy to see that the inflation rate simultaneously measures the rate at which money loses its value in terms of those goods. The phenomenon does not stop there. When the price level rises, the level of nominal wages tends to rise too. Wages rise partly due to rising productivity, but they also chase prices, which in turn chase wages. And the nominal exchange rate seems to be engaged in the same kind of race. Somehow, all nominal variables seem to leapfrog each other.

Whether very low or excruciatingly high, inflation is a key feature of modern economies. It was not always so. From time immemorial to the middle of the twentieth century, prices were pretty much trendless. Continuing inflation is a relatively new event in the history of humankind, and coincides with the emergence of modern central banks and fiat money. This should not come as a surprise once we remember the monetary neutrality principle. Ultimately, central banks and their ability to create money are the source of inflation.

So far, we have studiously avoided talking about inflation. We alluded to it in Chapter 6 when we established the principle of monetary neutrality, quickly relegating it to the faraway long run. In Chapters 10 and 11, we explicitly ruled it out by adopting the Keynesian assumption that prices are sticky. These were useful steps that allow us now to focus on this central aspect of modern economies. In this chapter, we focus on the 'medium run'. The short run corresponds to the Keynesian view, whereby prices hardly change. The long run is described by the opposite, neoclassical perspective. When enough time has elapsed, all prices are able to adjust and the economy is governed by monetary neutrality, a principle already presented in Chapter 6. In the medium run, prices move but not sufficiently to allow all markets to clear. We start by clarifying these issues.

Then we look at a very simple description of the inflation phenomenon, the **Phillips curve**. We will see that the Philipps curve has been useful for helping us think about the short-run behaviour of inflation. At the same time, it is unstable and seems to vanish at times. So we embark on a detective-like investigation of the puzzle of the 'disappearing Phillips curve'. This leads us to track down the various reasons why prices rise. We start by asking who sets prices (firms do) and why (mostly to cover their production costs). As we study production costs, we return to the wage bargaining process first encountered in Chapter 5, and find that wage negotiators

[1] The son of a Kiwi dairy farmer, A. W. Phillips (1914–1975) started out as an apprentice electrician working in an Australian mine, then left for Britain via China and Russia in the late 1930s. After a tour of duty in the Second World War and time spent as a prisoner of war, he studied at the London School of Economics and became a lecturer and later professor there. Phillips is remembered not only for his curve relating unemployment to rates of wage change, but also for the *Moniac*, a complex hydraulic representation of macroeconomy, on display in the Science Museum in London.

worry about prices of goods, for different reasons. We end up facing the apparently circular conclusion that prices drive wages and wages drive prices, in a sort of race between employers—who want high profits—and employees—who want high wages. The outcome of the analysis is an accounting of the factors that add up to a full explanation of inflation. This analysis helps us solve the Phillips curve puzzle. It also allows us to derive the **aggregate supply curve**, which will be teamed up with the aggregate demand curve developed in the next chapter to complete the system of inflation and output determination.

12.2 General Equilibrium with Flexible Prices: The Neoclassical Case

12.2.1 From the Keynesian Short Run to the Neoclassical Long Run

In Chapters 10 and 11, we made extensive use of the Keynesian assumption that prices are sticky. We justified this assumption by asserting that prices move slowly in normal times. In the short run, we said, ignoring inflation is an easy way to make things simple. When thinking about the long run, however, we need to make the diametrically opposite assumption that prices are fully flexible. This is the **neoclassical assumption** against which Keynes rebelled in the 1930s. He well knew the limits of his own assumption, but then famously wrote that 'in the long run, we are all dead'. In a parallel fashion, this section relaxes the Keynesian assumption and presents the view espoused by neoclassical economists ('neoclassicals' for short) writing long before and after Keynes. While some Keynesians and neoclassicals still fight it out, most economists now agree that the Keynesian assumption is acceptable in the short run, and the neoclassical assumption is the right way to think about the long run.[2] This section, therefore, studies the long run, but we also look at how things change over time, from the short to the long run.

In the Keynesian view, the price level is exogenous; output is endogenous and responds one-for-one to demand. In the neoclassical view, it is the price level that is endogenous and plays the equilibrating role in the goods market. It is the price of goods in terms of money. It turns out that, in this case, it is demand that adjusts to supply.

We have already started to examine the neoclassical case in Chapter 6 when we established the monetary neutrality principle. From that chapter, recall the Cambridge equation:

$$M = kPY \tag{12.1}$$

where, for the moment, we treat parameter k as constant.[3] We asked what happens when the money supply increases, say by 10%. In the short run, when prices are sticky, the only way for the money market equilibrium condition (as explained in Chapter 9) to hold is for real GDP to increase by 10%. This is the Keynesian case. If prices are flexible, however, equilibrium can be achieved with a 10% increase in the price level, without any change in real GDP. This is the neoclassical case and its underlying result is monetary neutrality.

If the Keynesian assumption is valid in the short run and the neoclassical assumption characterizes the long run, then it must be the case that an increase in the money supply—which is engineered when central banks cut the interest rate—is followed first by a rise in output that will be eventually dissipated through an increase in the price level. This dynamic adjustment path for output and prices is shown in Figure 12.1.

This can be expressed in terms of rates of change. The inflation rate is the rate of change in the price level, defined as a rate: $\Delta P/P$. Equation (12.1) implies that the price level is given by

$$P = M/(kY). \tag{12.1'}$$

[2] This intellectual debate is presented in Chapter 20.

[3] In the analysis of Chapter 9, this would correspond to the case in which the interest rate is constant—as it is in the steady state.

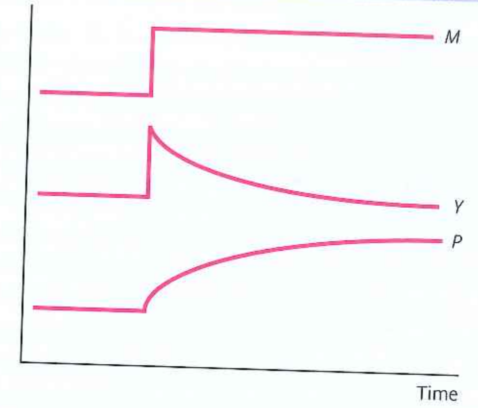

Fig. 12.1 From the Short to the Long Run
A one-off increase in the money supply has short and long effects which are qualitatively very different. The short run corresponds to the Keynesian assumption: prices are sticky and output moves to respond to demand. The neoclassical assumption describes the long run, when prices fully adjust and monetary neutrality implies that output is unaffected.

So for a constant k, the inflation rate, the growth of the money supply $\Delta M/M$, and the growth rate of output $\Delta Y/Y$, the following relationship must hold:

$$\Delta P/P = \Delta M/M - \Delta Y/Y. \tag{12.1''}$$

This is the 'rate of change' version of the classical quantity equation. Later, it will help us to think about the long run.

What causes the money supply to increase without bound? This is not an obvious question, since central banks have long abandoned the practice of setting monetary targets and prefer to use their power in money markets to set short-term interest rates. It turns out that this power does not come for free. Setting the interest rate necessarily means that a central bank must be ready to supply all liquidity the economy demands at that interest rate. As a corollary, it must accept the rate of growth in the money supply which results from that interest rate. How the money supply evolves endogenously is explained in more detail in Box 12.1.

Box 12.1 Where Does the Money Growth Come From?

In Chapter 6, the long-run price level and rate of inflation were related to the money supply and its growth rate. Yet in Chapter 9, we saw that central banks do not set growth rates of monetary aggregates. Rather, monetary policy is executed by setting interbank money market interest rates. In practice, central bank interest rate policy is well-represented by the Taylor rule. It turns out that the Taylor rule automatically implies a long-run rate of money growth. Because it supplies the money demanded given its chosen interest rate, the rate of money growth follows this rate of inflation. In other words: 'Every central bank gets the money growth and inflation rate that it chooses.'

It is easy to see this by re-examining the Taylor rule, which shows how central banks set interest rates:

$$i = \bar{i} + \underbrace{a(\pi - \bar{\pi})}_{\text{inflation gap}} + \underbrace{b\left(\frac{Y - \bar{Y}}{\bar{Y}}\right)}_{\text{output gap}} \tag{B12.1}$$

In the long run, inflation is equal to target inflation $\pi = \bar{\pi}$ and $Y = \bar{Y}$ so it follows that $i = \bar{i}$. Thus, the condition for equilibrium in the money market is given by

$$M/P = k(i)Y = k(\bar{i})Y = kY. \tag{B12.2}$$

In the long run, real output growth $g = \Delta Y/Y$ is given by real factors (the dichotomy assumption) and does not depend on money growth. Since the interest rate is constant, k is constant. Equation (B12.2) is simply another version of (12.1'); the logic of (12.1'') implies that the rate of nominal money growth is equal to the sum of the long run real growth rate and the central bank's inflation rate target:

$$\mu = g + \bar{\pi}. \tag{B12.3}$$

While money growth is neutral, it is not causal for inflation. Instead, the money supply grows endogenously to reflect the underlying trend growth in the economy as well as the choices of the monetary authority.

12.2.2 Supply-Determined Output in the Long Run

Why does price flexibility lead to a situation where demand adjusts to supply? Let us start with the supply side, that is, the long-run output and labour market outcomes studied in Chapters 3 and 5. Figure 12.2(*a*) displays the production function. Panel (*b*) shows the demand for labour implied by the marginal productivity of labour—as determined by the position and the slope of the production function, and the supply of labour—as shaped by labour market institutions. The resulting equilibrium at point *A* in panel (*b*) indicates how much labour is utilized. The corresponding point *A* in panel (*a*) shows how much is produced and supplied. In the classical long run, output equals its trend value $\bar{Y}$ (which is generally growing over time), employment equals $\bar{L}$ and the equilibrium rate of unemployment is $\bar{U}$: These points are labeled as *A* in both panels of the Figure 12.2. We can also see in panel (*b*) the equilibrium real wage, the ratio of nominal wages *W* to the price level *P*.

It is useful to examine an outcome under the Keynesian assumption, that is when wages and prices are sticky. We studied this situation at length in Chapters 10 and 11. In particular, we look at a situation when demand weakens to $Y < \bar{Y}$. Firms will respond by cutting production and move to point *B* in panel (*a*) in Figure 12.2. Panel (*b*), however shows that there is no market equilibrium corresponding to this reduction of production. If wages and prices are sticky, the outcome will be at point *B* is panel (*b*).[4] What is remarkable about point *B* is that it is neither on the labour demand nor on the labour supply schedules. The distance *BA* measures involuntary unemployment, the extent of disequilibrium seen by employees. Firms too are unhappy since being off the labour demand curve means that profits are not optimized. The paradox is that the real wage is at its 'correct' level, the one corresponding to point *A*. The following analysis would also hold if the real wage would change to take us to *B′* (higher than the equilibrium value) or *B″* (lower than its equilibrium value).

[4] We ignore the possibility that firms hoard rather than dismiss unessential labour. In fact, firms often avoid shedding their employees in a brief recession if they can afford to do so. In this case they would not be 'on their labour demand curve', they would employ more labour than strictly optimal.

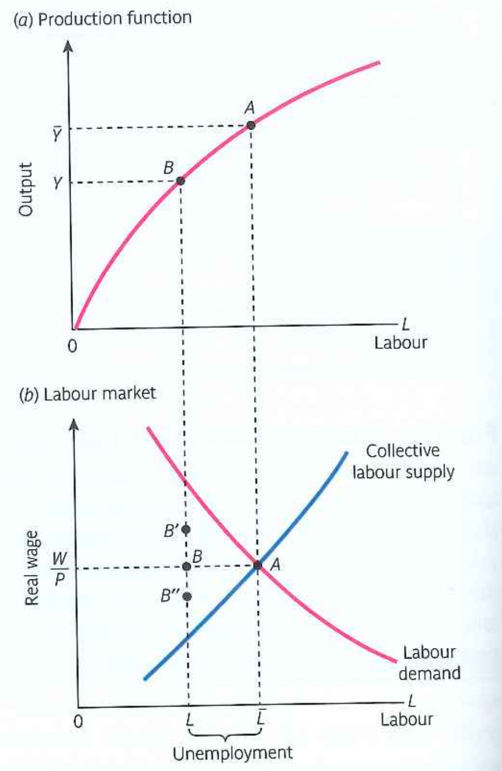

Fig. 12.2 Output and the Labour Market in the Long Run

Long-run output $\bar{Y}$ is produced by firms using long-run labour input $\bar{L}$ (point *A* in panel (*a*)) and the equilibrium of labour demand and collective labour supply (point *A* in panel (*b*)). Point *B* depicts a Keynesian situation in which aggregate demand is insufficient and less than trend level of output, even though the real wage is at its long-run level. (Point *B′* and *B″* show initial situations in which the real wage is too high or too low, respectively.) From point *B*, it would be sufficient simply to raise aggregate demand; in the other cases, the real wage must fall or rise via some appropriate combination of nominal wage and price changes. In the neoclassical case, price reductions (by firms) and wage cuts (by employees) will be necessary to restore equilibrium when demand is low. If exchange rates are flexible, a depreciation can help restore demand to long-run equilibrium at point *A*.

In the long run, the neoclassical principle implies that nominal wages and prices will adjust—rise or fall—until equilibrium is restored in both product and labour markets. This means returning to point *A*. How will they move? The underlying guiding principle is that prices rise when demand for goods is strong, and fall when it is weak. Similarly, wages rise when demand for labour is strong, and decline when demand for labour is weak. In Figure 12.2, firms facing weak demand would reduce their prices. If the nominal wages (*W*) remain sticky, real wages (*W/P*) rise—a falling price level (*P*) means a higher purchasing power of wages. As we move from B toward B′, involuntary unemployment gets even larger. Over time, as nominal wages become flexible, they would decline.

Are we just going endlessly up and down from *B* to *B′* and back? No, because we must understand what the decline in the price level *P* means for the demand for goods. The short answer is that a decline in goods prices will encourage more demand and that this will bring the economy back to point *A* in both panels. How exactly will this come about? Let us go back to the *IS–TR* analysis of the previous chapter. If the exchange rate is flexible, the Taylor rule implies that a declining price level will be met by a demand-enhancing lower interest rate, a nominal and ultimately real depreciation, a process that will continue until demand is restored to its initial value. If the exchange rate is fixed, the decline of the domestic price level also implies a real exchange rate depreciation, which raises foreign demand. Either way, price and wage flexibility will bring the economy back to point *A*, with lower price and wage levels but at the initial real wage level.

12.2.3 Implications for the Long Run

If we consider that the neoclassical assumption describes the long run, we can conclude that demand disturbances only have a temporary impact on output and employment. Thus, we start from the Keynesian short run where supply passively adjusts to demand and gradually moves to the long run as prices and wages respond to prevailing conditions. Eventually, the economy is back to full equilibrium or, equivalently, GDP returns to its trend level and equilibrium unemployment (as defined in Chapter 5) prevails again. The supply represents what the economy's resources (physical capital, human capital and the state of technology) can produce and how much income can be earned. This description nicely fits the idea of business cycles represented in Figure 10.1 and the idea that price adjustments gradually substitute for quantity adjustments (see Figure 12.1). The task of the following sections is to examine carefully how prices and wages move in response to temporary disequilibria in the goods and/or labour markets.

12.3 The Phillips Curve: Chimera or a Stylized Fact?

12.3.1 A. W. Phillips' Discovery

The short-run Keynesian assumption and the long-run neoclassical assumptions are fairly easy to deal with: prices either do not change at all, or they fully adjust. The medium run, as represented in Figure 12.1, is considerably messier. Prices do move, but only part of the way in any given period of time. The Keynesian assumption always had the major disadvantage of leaving inflation unexplained by assuming it away. Even die-hard Keynesians conceded at the high point of their influence that they had no clue on how to incorporate inflation in their model. They referred to it as the 'missing equation'. The search for this equation turned up the Phillips curve, a negative relationship observed to hold between inflation and unemployment, the twin 'bads' of macroeconomics. It is represented in Figure 12.3.

In the late 1950s, A. W. Phillips plotted the annual rate of growth of nominal wages, i.e. wage inflation, against the rate of unemployment in Britain during the period 1861–1957. He found a remarkably robust negative correlation, which was confirmed for a number of other countries.[5] Figure 12.4 plots actual

[5] Phillips was not the first to discover the Phillips curve. The American economist Irving Fisher published a paper in the *International Labour Review* of 1926 in which he unearthed a similar relationship in the USA.

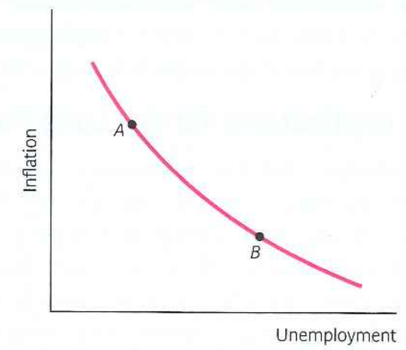

Fig. 12.3 The Phillips Curve in Theory

The Phillips curve implies a negative relationship, or trade-off, between unemployment and inflation: you can get less of one evil only by accepting more of the other. When discovered, it was seen as representing a menu of options from which governments could choose. For example, it could keep unemployment down (point *A*) at the cost of some inflation, or could limit inflation (point *B*) but only by accepting higher unemployment.

Phillips curves—using the rate of price inflation instead of wage inflation—for the UK and the average of 16 advanced economies for the period 1921–1973 (excluding war years). While far less clean than the stylized version (a number of outliers correspond to exceptional events), the actual Phillips curves are consistent with the negative relationship shown in stylized form in Figure 12.3.

The Phillips curve was embraced at the time for two main reasons. First, it provided the missing theoretical relationship that complements and indeed completes the Keynesian *IS–TR* apparatus. Second, even as its durability was increasingly questioned, it remained a practical tool for policy-makers. They can aim at low unemployment but they must accept substantial inflation (point *A* in Figure 12.3), or they may prefer low inflation but at the cost of high unemployment (point *B*), or any intermediate situation. Once they decide, they can use their instruments (monetary and/or fiscal policy) to reach the

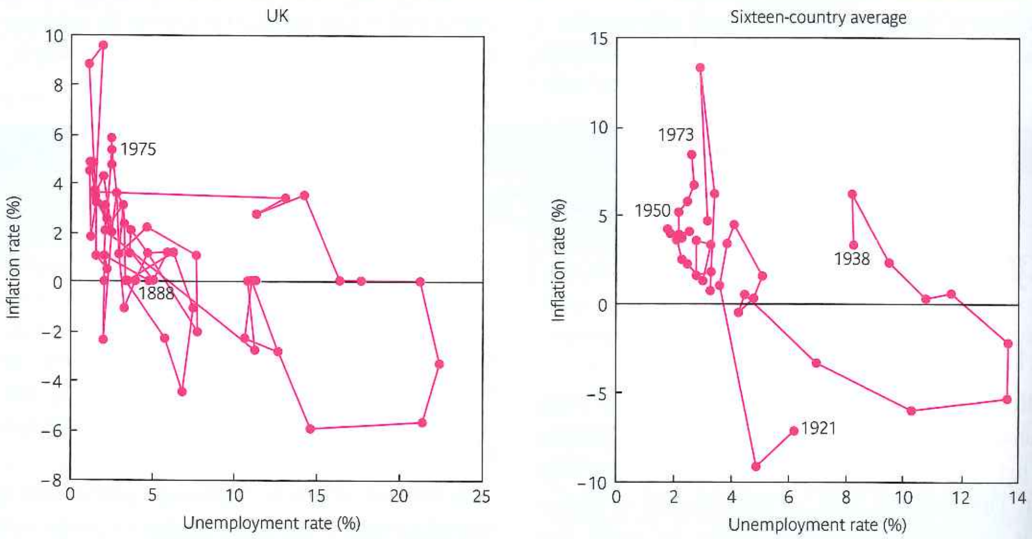

Fig. 12.4 The Phillips Curve in Reality

In reality the Phillips curve was less stable than depicted in Figure 12.3. The two panels depict the evolution of inflation and unemployment in the data for the United Kingdom in the period 1888–1975 and for an average of 16 countries 1921–1973 (excluding 1939–1949). While movements in northwest to southeast directions do dominate the year-to-year movements, there are notable deviations.

Sources: Maddison (1991); Mitchell (1998). Unweighted average of observations for Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Sweden, Switzerland, the UK, and the USA. For some years, some countries are missing.

chosen point on the Phillips curve.[6] This policy dilemma came to be known as the **Phillips trade-off**.

12.3.2 Okun's Law and a Supply Curve Interpretation

The Phillips curve relates inflation to unemployment. A related relationship between inflation and the output gap will be called the supply curve. The link between the two is another stylized fact, known as **Okun's Law**, which is a negative relationship between output and unemployment.[7] More precisely, Okun's Law links the **unemployment gap**, the distance between actual and equilibrium unemployment as defined in Chapter 5, to the **output gap**, the distance between real GDP and its long-run trend, as illustrated in Figure 10.1. Figure 12.5 plots estimates of the unemployment and output gaps for Germany as an example of how unemployment and output gaps systematically move in opposite directions. The logic is intuitive: when the economy grows fast and output is above trend, demand for labour is strong and unemployment falls below its equilibrium level.

A stylized representation of Okun's Law is displayed in Figure 12.6. Formally, Okun's Law can be represented as follows:

$$U_{gap} = -hY_{gap}, \tag{12.2}$$

where $U_{gap} = U - \bar{U}$ and $Y_{gap} = \frac{Y - \bar{Y}}{\bar{Y}}$.

The unemployment gap U_{gap} is defined as the difference between the current unemployment rate U and its equilibrium level $\bar{U}$. Similarly, the output gap is the per cent deviation of real GDP Y from its potential level $\bar{Y}$.[8] Okun's Law means that when

[6] This view of a trade-off was echoed by Helmut Schmidt, the ex-chancellor of West Germany, who stated in a newspaper interview in 1978 that he would prefer 5% inflation to 5% unemployment. How times have changed!

[7] The law is named after the US economist Arthur Okun (1928–1980). In his research, he showed that a 1% decline of the US unemployment rate was associated with a 3% increase of GDP above trend. Referring to equation (12.2), this implies a value for *h* of one-third.

[8] Note that the unemployment gap and the output gap are written in different ways. This is a technical detail. In Figure 12.5, both gaps are shown in per cents. Since the unemployment rate is already measured in per cent (of the labour force), its difference from equilibrium unemployment is measured in percentage points. For the output gap to also be in per cent, we compute the difference between GDP and its potential level—both measured in the local currency—as a proportion of potential GDP.

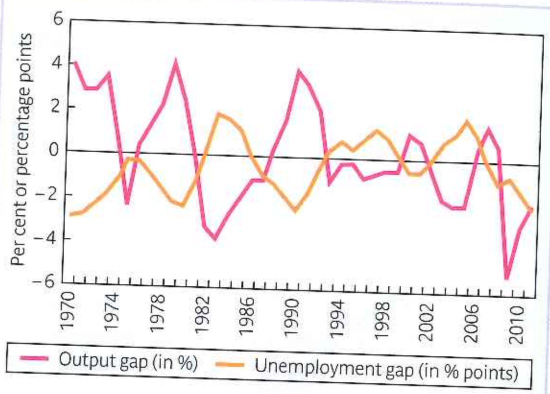

Fig. 12.5 The Output Gap and Unemployment in Germany, 1970–2011

The output gap measures the distance between real GDP and its trend and is computed in per cent of trend GDP. The unemployment gap is the difference between actual and equilibrium GDP, both measured in per cent of the labour force. When business conditions vary, firms adapt the supply of goods and services. To that effect, they adjust their demand for labour. For example, when the economy goes into a recession, firms employ fewer workers and the unemployment rate rises.

Notes: Trend real GDP and equilibrium unemployment are estimated by the OECD.

Source: OECD, Main Economic Indicators.

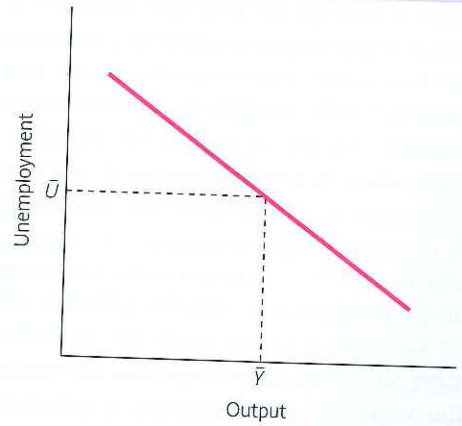

Fig. 12.6 Okun's Law in Theory

Okun's Law implies that, when the economy slows down, unemployment increases; when output rises relative to trend, unemployment declines.

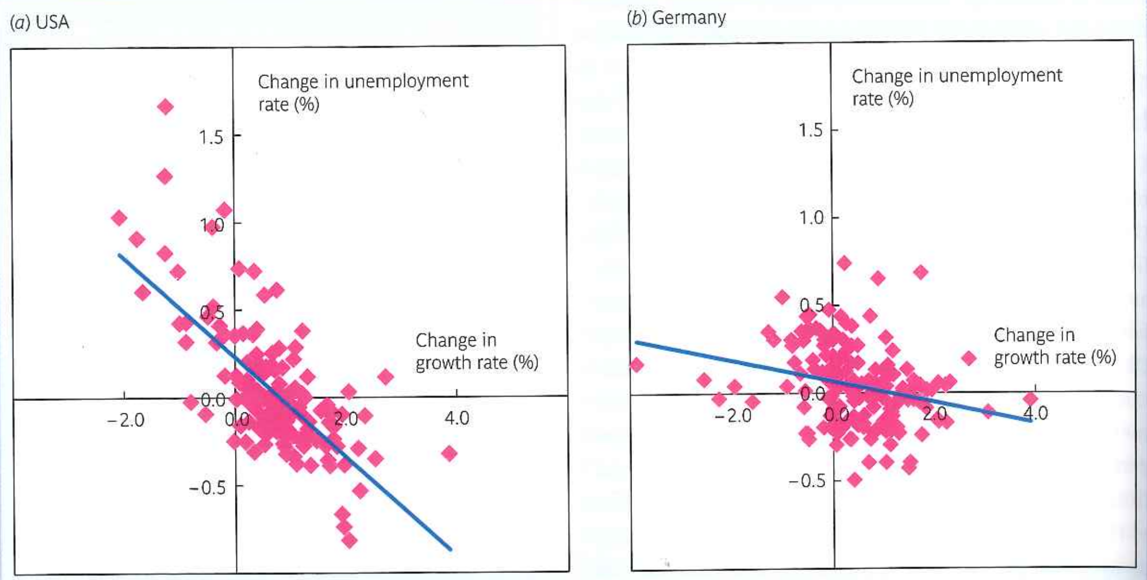

Fig. 12.7 Okun's Law in Reality

Okun's Law is not identical across countries. The two panels show how Okun's Law accounts for changes over time in the unemployment rate (in %) against percentage changes in changes in output, in the USA and Germany. (This is slightly different from the depiction in Figure 12.6.) In the USA, a one per cent increase in output is associated with roughly a one-third percentage point decline in the unemployment rate. In Germany, the response to unemployment change is about a fifth of that in the USA. Okun's Law cannot explain all the variation in unemployment but it does explain a fair amount. It accounts for much more overall variation of unemployment in the USA than it does in Germany.
Source: OECD.

output is above trend, the output gap is positive and the unemployment gap is negative; when output is below trend as in Figure 12.2., the output gap is negative and the unemployment rate is above trend. The parameter h captures the response of unemployment to GDP fluctuations measured as the output gap.

Figure 12.7 plots changes in the unemployment rate against changes in output for the USA and Germany. The negative relationship is very clear in both countries, but the slope of the Okun relationship is not constant across countries. In Germany, an increase in growth of 1% yields a much lower reduction in unemployment than in the USA. This reflects labour laws that are more protective of workers in Germany. Of course, Okun's Law does not explain all the fluctuations in unemployment, which in reality are affected by many other factors.

When we combine the negative relationship between inflation and the unemployment rate from the Phillips curve with the negative relationship between the unemployment rate and the output gap from Okun's Law, we obtain the positive relationship between the output gap and inflation shown in Figure 12.8. This relationship is called the **aggregate supply curve**. Section 12.4 provides a complete interpretation of the

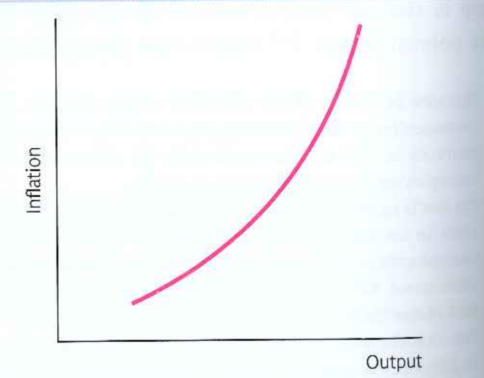

Fig. 12.8 The Aggregate Supply Curve

Combining the Phillips curve and Okun's Law, we obtain the aggregate supply curve. The curve says that inflation rises when output increases.

supply curve, but the intuitive logic that lies behind it is straightforward. The supply curve answers the following question: under which conditions will an increase in aggregate demand lead firms to supply more output and employees to work more to produce that extra output? The short answer is: inflation must increase, to boost wages and profits.

That the aggregate supply curve is upward sloping reminds us of the supply curve in microeconomics. While this similarity is intuitive, it can also be misleading. In microeconomics, the horizontal axis corresponds to the production of a particular good and the vertical represents that good's relative price. Here we deal with *aggregate* output or real GDP on the horizontal axis and the *overall* price index—or here its rate of increase, which is the inflation rate—on the vertical axis. This highlights the difference between micro- and macroeconomics.

12.3.3 Hard Questions about the Phillips Curve

Weak foundations

Not long after its discovery, the interpretation of the Phillips curve was perceived as standing on weak theoretical legs. In the late 1960s, Milton Friedman, the Nobel laureate in 1976, and Edmund Phelps of Columbia University,[9] who received the Nobel Prize in 2006, independently attacked the Phillips curve implication of a permanent trade-off between inflation and unemployment or output. They both asked the following question: how could the rate of change of nominal variables, such as wages or prices in pounds or euros, be related to real variables, such as employment, unemployment, and output in the long run? If the monetary neutrality principle is valid in the long run, then rates of change in the price level and other nominal variables should be unrelated to the real economy.[10] In this case, they noted, the Phillips curve should be observed temporarily. Only if workers and firms suffer from **money illusion**—i.e. if they act on increases in their own prices or wages without taking contemporaneous increases in all other prices into account—will they raise output.

To understand the critique of Friedman and Phelps, it is helpful to think about the long-run behaviour of labour markets and output as presented in Section 12.2. The principle of monetary neutrality asserts that the economy is dichotomized in the long run: the real and nominal sectors of an economy stop influencing each other. If the long run is defined as when output level is on its trend growth path $\bar{Y}$ and unemployment is at its equilibrium rate $\bar{U}$, then the rate of inflation is determined by the rate of money growth, not by the level of output or the unemployment rate. These arguments were presented in detail in Box 12.1. Graphically, if unemployment returns to its equilibrium level in the long run, the **long-run aggregate supply curve** and the Phillips curve must be vertical lines, as displayed in Figure 12.9. This conclusion is in agreement with the neoclassical view, developed in Section 12.2, which implies that output is determined by the supply side in the long run.

Wobbly evidence

The critique of Friedman and Phelps was largely ignored in the late 1960s, but it proved to be right on target in the 1970s when, as Figure 12.10 shows, the Phillips curve simply vanished. Quite spectacularly, in the mid-1970s and early 1980s, both inflation and unemployment started to rise, a phenomenon which came to be known as **stagflation**. Stagflation was strictly incompatible with the Phillips curve and its trade-off. A number of consequences followed. Policymakers, now grappling to beat back inflation, became sceptical of 'Keynesian activism'. They started to put greater emphasis on long-run monetary neutrality as a guiding principle for monetary policy. Macroeconomics also underwent a profound transformation, which can be thought of as a massive effort to rehabilitate and reconstruct the Phillips curve.

Over nearly a century, the inverse relationship between inflation and unemployment had seemed relatively robust. Why did it suddenly break down in all countries, and at about the same time? Because it was unlikely to be simply a case of bad luck, the systematic breakdown of the Phillips curve is now much better understood than 3–4 decades ago, and even helped economists better understand the true forces behind it.

9 This is the same Phelps who formulated the golden rule of economic growth in Chapter 3.

10 In his address to the American Economic Association in 1967 Friedman argued that 'there is always a temporary trade-off between inflation and unemployment, there is no permanent trade-off. The temporary trade-off comes not from inflation *per se*, but from a rising rate of inflation', Friedman (1968: 10).

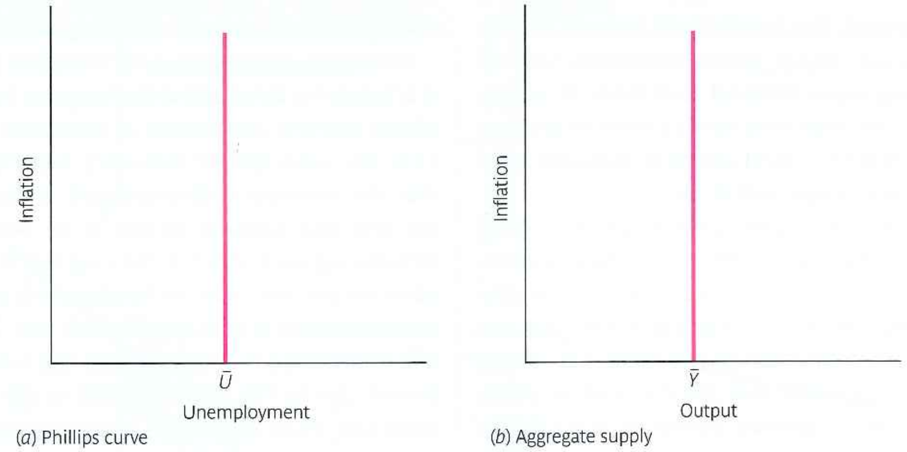

Fig. 12.9 The Long Run

In the long run, unemployment is at its equilibrium rate and output is on its trend growth path. Both the Phillips curve and the aggregate supply curve are vertical. Inflation is determined by the rate of money growth or by the central bank's target rate of inflation, independently of output or unemployment.

The challenge is to explain both the existence of a Phillips curve and its disappearance, as well as the striking similarity between different countries' experiences. Nearly everywhere inflation and unemployment increased sharply, at first around 1973–1974, and then around 1979–1980, precisely at the time of sharp increases in the price of petroleum, the so-called **oil shocks**. Oil prices increased fourfold in 1973–1974 and then doubled again in 1979–1980. Interestingly, in between the oil shocks, and after the second oil

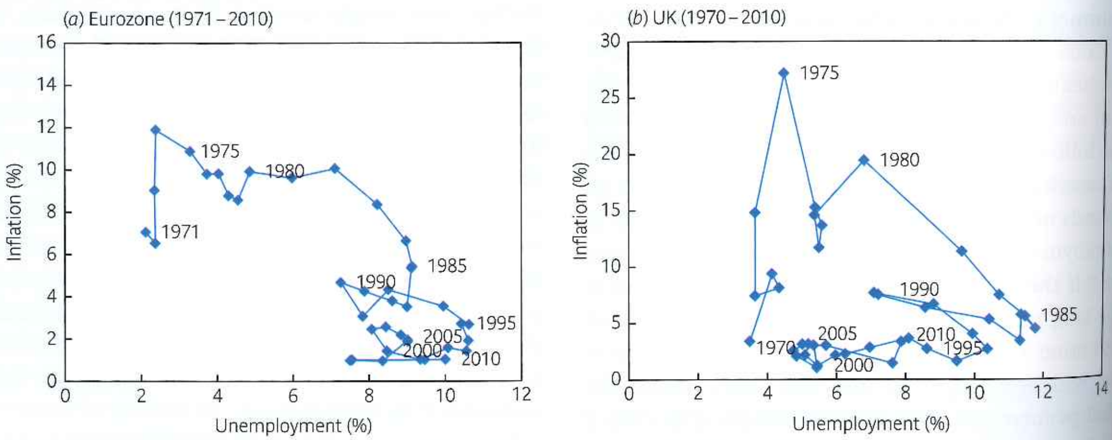

Fig. 12.10 Phillips Curves: Recent Experiences in the Euro Area and the UK

The Phillips curve broke down at the end of the 1960s. In the euro area, the sharp movements of the early and late 1970s coincided with the two oil shocks. Then the great deflation of the 1980s is compatible with the traditional Phillips curve, as are the periods 1990 – 1994 and 1997 – 2007. In the UK, the old Phillips curve survives over 1960 – 1967 until the upward jumps in 1973 – 1975 and 1979 – 1981 that coincided with sharp increases in oil prices—the so-called 'oil shocks'. As in the euro area, three different Phillips curves can be identified: during the deflation period of the 1980s, over 1986 – 1993, and in 1994 – 2007.

Source: OECD, *Economic Outlook*.

shock, Phillips curves re-emerged, each time further above and to the right of the previous one. Two examples of these episodes are presented in Figure 12.10 and further dissected in the following section.

The weak theoretical foundations of the Phillips curve and its successive mutations over time are two sides of the same coin. Confronted with uncompromising data, economists indeed had to go back to the

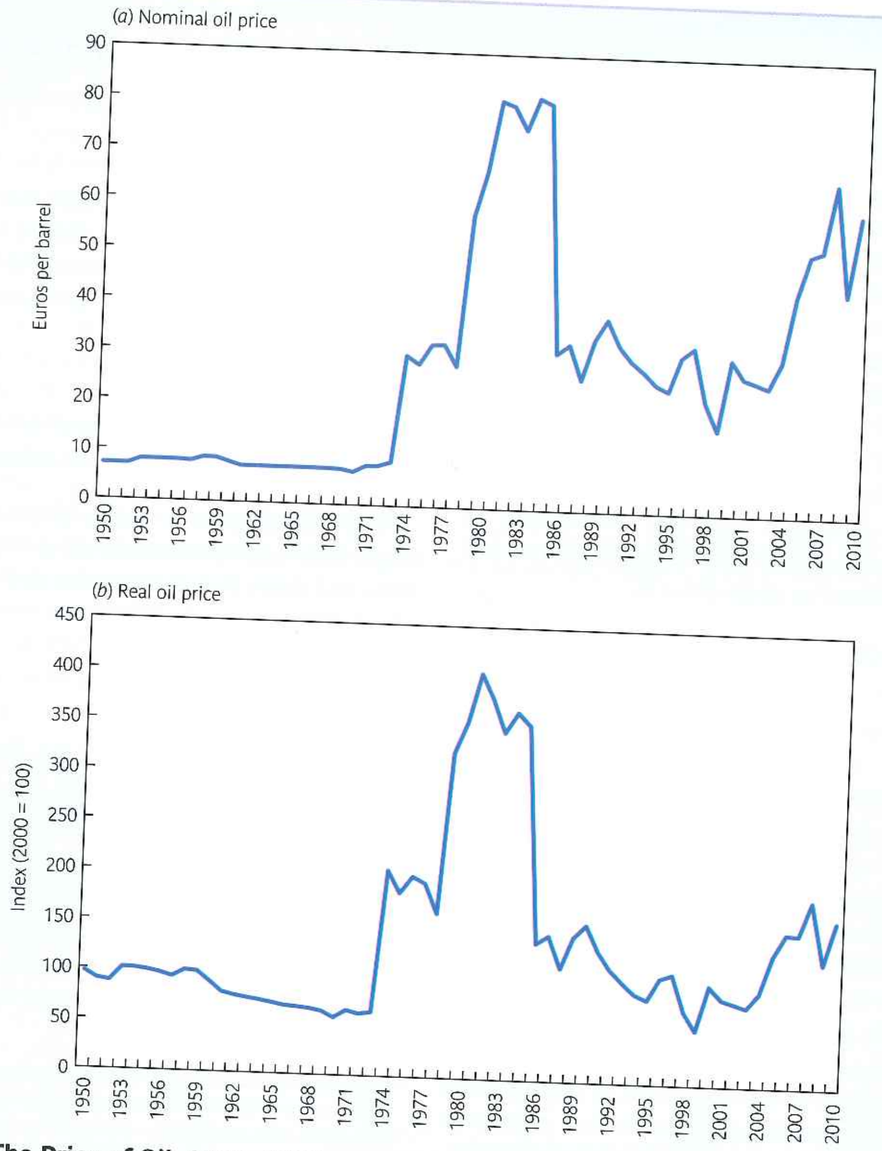

Fig. 12.11 The Price of Oil, 1950 – 2011

The price of oil is generally quoted in US dollars; in panel (a) it has been converted into euros. The first two oil shocks (1973 – 1974 and 1979 – 1981) and the two next ones (1999 – 2001 and 2005 – 2008) are clearly visible. In 1986, a favourable oil shock occurred—a sharp decline of oil prices. Panel (b), which deflates the oil price by the consumer price index, shows that, while oil prices rose sharply in 2007 in euros, in real terms that increase hardly compares in severity with the oil shocks of the 1970s.

Source: British Petroleum, IMF.

drawing board. Yet, as already noted, Friedman and Phelps had already provided the answer to the puzzle even before the puzzle emerged. The next section presents an interpretation of inflation that can account for these facts.

12.4 Accounting for Inflation: The Battle of the Mark-ups

Why do we care about the fate of the Phillips curve in the first place? Recall that it deals with the medium run. In the short run, prices are sticky and we can ignore inflation. In the long run, the Phillips curve is vertical and has nothing to say about inflation. If we want to understand the inflation phenomenon, therefore, we need to understand the time in between. To that end, we will follow a pragmatic, 'divide and conquer' strategy. By breaking inflation into a series of its most important components, we can better account for its immediate sources and will be better equipped to think about it.

12.4.1 Prices and Costs

Price setting

The price level is the average of individual prices, but who sets the prices of goods and services and how?[11] Looking at price tags in stores and price lists in catalogues, you may correctly conclude that producers and retailers decide on the prices of their goods and services. Obviously, each producer prefers higher to lower prices, but their enthusiasm is held in check by the need to attract enough buyers. This is the process that we now study. Box 12.2 recalls a few principles. It explains how competition limits the ability of firms to set prices, and that firms try hard to reduce these same limitations. It describes firms as setting prices as a mark-up above nominal production costs, aiming at a margin consistent with maximizing profits.

Costs

Now that we understand the price mark-up—the relationship of price to cost—it will make sense to look at those production costs, called **average** or **unit costs**, measured in pounds or euros.[12] Unit cost (UC) is simply the total cost of production divided by the number of units produced. It is convenient to break down unit costs into two main categories: labour and non-labour costs:

Unit costs in euros
= total costs in euros/number of units produced
= unit labour costs + unit non-labour costs.

For the economy as a whole, labour costs are the single largest component of unit production costs. Table 12.1 shows that at the macro level, the share of labour costs (wages, benefits, plus contributions to social insurance) ranges from 50% to 70% of value added in developed countries, and is usually higher in labour-intensive services than in capital-intensive industries.[13] For this reason, we will focus on labour costs for the next few sections. We will consider non-labour costs later in Section 12.4.6.

Labour costs and wages

A firm's total labour costs are simply the product of the number of hours worked in the firm (L) and the average gross hourly labour cost (W). Gross hourly labour costs include not only wages and salaries per

[11] Chapter 2 presents various ways of measuring the price level.

[12] In theory, the relevant concept is the marginal production cost, the cost of producing another unit of output. In practice, marginal costs are difficult to measure, so we approximate them by unit costs. This is an acceptable approximation under conditions of constant return to scale in all factors.

[13] In interpreting these numbers, it is important to remember the distinction between value added and turnover or total sales, which was stressed in Chapter 1. As a percentage of total sales, wage shares are much lower because total turnover in an economy includes the costs of intermediate goods. The figures reported in Table 12.1 have netted out payments for intermediate inputs produced by other firms.

Box 12.2 When and How Firms Set Prices

Economists speak of perfect competition in goods markets when firms are unable to set prices. A good example is the fruit grower who sells his apples in the town market. With many other sellers around, the farmer cannot set his price very far from some average price for apples of the same variety. If he raises his price just a few cents, he loses all his customers; if he lowers his price, he will sell everything, but regret the forgone profits. Producers of standard products, such as agricultural products or raw materials, have relatively little choice. We say that they have no **market power**.

Yet most firms do have market power, and some firms have a lot of it. In fact, they go to extreme lengths to establish market power because this helps them achieve higher profits. This is called **product differentiation**, which means making each product special and different from the competitors. They do so through design (similar cars always differ in many subtle ways), through advertising and marketing to win consumer loyalty (some people like Volkswagen, others Renault), or product quality and reputation. The pay-off is a higher willingness to pay by their customers, and a greater degree of market power. Firms can charge higher prices above the bare minimum that they need to survive, and still keep a large share of their customers.

Mark-up pricing is a simple description of how firms with market power set prices. Obviously, when setting a price, a firm wants to cover its production costs, and then it tries to do more. The margin over the cost of producing one more good, the marginal cost, is the mark-up. The mark-up depends on the sensitivity of the particular product's demand to price—its demand elasticity. If the market is highly competitive, the demand elasticity is near-infinite, because customers have lots of alternatives: any price increase will drive customers to other sellers. In that case the mark-up is zero and the price is equal to marginal cost. This is the case of perfect competition. If competition is weak, because the firm has been able to differentiate its good and build a 'niche', or because other firms cannot enter the market, the demand elasticity is low and the mark-up will be higher.

Table 12.1 Wage Share of Value Added by Country and Selected Industries, 2009

	Total Economy	Manufacturing	Construction	Finance and Business Services	Basic Metal Industries	Wholesale/ Retail Trade
Belgium	59.1	70.5	54.2	35.2	79.9	62.7
Czech Republic	49.3	52.7	42.6	40.5	56.4	53.5
Denmark	68.3	72.8	84.5	45.3	72.7	81.5
Germany	57.3	78.8	57.1	31.7	n.a.	68.9
Italy	47.7	65.0	43.3	25.5	71.6	43.1
Japan	52.3	56.8	73.5	n.a.	66.8	62.2
Netherlands	58.5	62.9	60.0	51.7	84.2	62.0
Poland (2008)	42.4	49.8	34.6	27.8	n.a.	31.2
Spain	52.8	62.6	52.2	35.3	68.0	49.0
United States	55.4	54.2	70.4	38.7	63.6	57.1

Source: OECD STAN Database

hour, but also paid vacations, direct labour taxes, social security contributions, and other benefits paid by employers on behalf of their workers. (In many European countries, these additional costs can be nearly equal to the net pay received by the worker.) A country's aggregate unit labour cost is computed in the same way. It is the ratio of total labour costs WL, also called the **wage bill**, to total real output, i.e. real GDP Y:

$$\text{(12.3)} \quad \text{nominal unit labour costs} = \frac{\text{wage bill}}{\text{real GDP}} = \frac{WL}{Y}$$

Nominal unit labour costs must be distinguished from the *real* labour cost of producing a unit of real GDP or, equivalently, the ratio of the nominal wage bill to nominal GDP:

$$\text{(12.4)} \quad \text{real unit labour costs} = \frac{\text{nominal unit labour costs}}{\text{nominal GDP}} = \frac{WL}{PY} = \frac{\text{wage bill}}{\text{nominal GDP}}.$$

The real unit labour costs measure the share of GDP (PY) that goes to labour.

12.4.2 The Battle of the Mark-ups: A Simple Story

Prices as mark-up on labour costs

Mark-up pricing means that firms set the price of goods as much as they can above their nominal unit costs. We have agreed to ignore non-labour costs, so we now consider that firms will aim for a price as much above their unit labour costs as it takes to maximize its profits. This can be written as follows:

$$\text{(12.5)} \quad P = (1+\theta)\frac{WL}{Y},$$

where $\theta \geq 0$ is the price mark-up. A simple way of thinking of the mark-up is if $\theta = 0.3 = 30/100$, the price is set 30% above nominal unit labour costs.[14]

Wages as a mark-up on prices

Firms mark prices above labour costs, but what determines labour costs and nominal wages? A good starting point is Chapter 5, which notes that wages are set through negotiations between employees and employers. These negotiations deal with two different issues.

First, the employees want to secure as large an income as possible. Yet both wages and salaries as well as employment must originate in value added (GDP) of ongoing firms. Thus, while labour is likely to aim at as large a labour share of output as possible, as we saw in Chapter 5, its efforts are limited by firms' demand for labour.

Second, wage negotiators can only bargain over *nominal* wages. They do not know for sure what tomorrow's price level will be. Typically, wage agreements cover a period of one or more years, and the future evolution of the price level is unknown. Naturally, employees want to protect their nominal wages from inflation. Employers normally agree to incorporate the likely evolution of the price level in wage settlements but worry about overestimating it and paying their workers 'too much'—paying workers too much means making low profits or even losses. This is why the expected evolution of the price level, which is incorporated into the wage agreement, is a central part of negotiations. We denote the resulting expected price level by P^e.

In the end, both sides bargain directly over the nominal wage W and indirectly over the expected labour share WL/P^eY. A simple description of the outcome of the negotiations is that wages determine unit labour costs as a mark-up γ over the expected price level:

$$\text{(12.6)} \quad \frac{WL}{Y} = (1+\gamma)\overline{s_L}P^e,$$

where $\overline{s_L}$ is the 'normal' share that wage negotiators accept as representing an appropriate division of income in normal times, i.e. when the demand for labour is neither too strong nor too weak.[15] The wage mark-up γ shows how the agreed-upon share varies over the cycle. The mark-up can be positive or negative, and it is zero on average. In some years, $\gamma > 0$, and the agreed-upon wage is above its normal level. In other years, $\gamma < 0$ and the wage is lower.

[14] Formally, $\theta = \frac{P - WL/Y}{WL/Y}$.

[15] Box 12.3 provides a formal explanation.

Putting it all together

We have described prices as a mark-up over unit costs, agreed to ignore all non-labour costs and therefore to focus on unit labour costs, and found that the expected labour share is a mark-up over its normal level. This in turn means that wages are a mark-up over prices. In the end, prices depend on wages and wages depend on expected prices.

Isn't this a bad case of circular reasoning? Quite to the contrary, this description captures the fundamentally conflictual relationship between employers and employees as they divide up the GDP cake. It has earned the title '**battle of the mark-ups**'.[16] The price mark-up, equation (12.5), sets the price level as a mark-up on wages, while the wage mark-up, equation (12.6), sets wages as a mark-up on the (expected) price level. Firms increase profits by reducing labour costs relative to the prices that their products can fetch. Employees increase real wages by pushing them up relative to expected prices.

Note carefully the distinction: *actual* prices depend on wages, while wages depend on *expected* prices. The apparent circularity really links actual prices to expected prices, as can be seen by combining (12.5) and (12.6), resulting in

$$\text{(12.7)} \quad P = (1+\theta)(1+\gamma)\overline{s_L}P^e.$$

We see that the circular process has an anchor: P^e, the level of prices expected by wage negotiators to prevail over the course of the contract. We will later find that this expectation is one central determinant of inflation in the medium run.

12.4.3 Productivity and the Labour Share

Before we move on to the next and final step, a clarification may be helpful.[17] We know from Chapter 3 that labour productivity (Y/L) and real wages do tend to grow systematically over time. Productivity does not appear at all in the final outcome of the price- and wage-setting process as represented by equation (12.7). This may be surprising but a bit of reflection shows that it is not.

To see this, recall that equation (12.4) defines real unit labour cost (or the labour share) as WL/PY. When real wages grow at the same rate as labour productivity, real unit labour costs remain constant, as shown formally in Box 12.3. Put differently, when productivity is rising, real wages can increase without raising production costs. This is the base case, which defines the 'normal' labour share $\overline{S_L}$. The battle of the mark-ups studies fluctuations in real unit labour costs that occur when real wages temporarily depart from this principle.

The base case is precisely what we expect to see following the principles developed in Chapter 5 and one of the stylized facts in Chapter 3. Indeed, Figure 1.3 in Chapter 1 showed that, in four countries over half a century, labour shares fluctuate from year to year, but with little discernible trend. This is just another way of saying that technical progress continuously generates higher incomes, but gross hourly wages (including all employee costs) and hourly productivity track each other pretty well.

12.4.4 Cyclical Behaviour of Mark-ups

Now we turn to the determinants of the two mark-ups. The brief answer is that they tend to move over business cycles. To see which side would 'win' the battle of the mark-ups, we need to separate out the two, distinctly different mark-up decisions.

Start with the price mark-up θ. The price mark-up is like a profit margin—the excess of revenues over variable costs. Firms naturally would like profits to be as large as possible, but have to worry about competition.[18] Competition increases in good times, when there is 'more money in the marketplace'. While firms may try to raise price mark-ups when demand is high, high business volumes increase total profits so much that new firms enter and steal competitors' customers. The effect here is uncertain.

The wage share mark-up γ is the outcome of wage bargaining at the level of collective negotiations, but

[16] The battle of the mark-ups approach to understanding inflation has found empirical support in OECD countries in groundbreaking work by researchers at the London School of Economics: Richard Layard, Steven Nickell, and Richard Jackman, among others.

[17] This section may be skipped without loss of continuity.

[18] A natural tendency is for competitors to agree on large mark-ups. Such collusion is usually illegal and competition authorities seek out and prosecute anti-competition agreements.

Box 12.3 The Wage Mark-Up and the Labour Share

Equation (12.4) defines the labour share as $\frac{WL}{PY}$. We describe wage negotiators as proceeding in two steps. First, they agree—in a process that can in fact be highly conflictual—on a 'normal' level of this share:

$$\frac{\overline{W}L}{PY} = \overline{s}_L,$$

where $\overline{W}$ is the corresponding 'normal' real wage. Once this is done, wage negotiators will recognize that the labour market is tight (strong demand because the economy is booming) or weak (in a recession). They then agree—an equally controversial step—on the expected real wage (W/P^e) as a markup γ on the normal wage, delivering equation (12.6).

What drives the normal labour share? This is a complex social issue. Here we just note that the share remains constant where workers get their appropriate or 'fair' share of productivity gains. To see this, note that the labour share can be rewritten as $\frac{W/P}{Y/L}$. Using the rate of change rule presented in Box 6.3, we see that the percentage increase in the labour share is $\frac{\Delta(W/P)}{W/P} - \frac{\Delta(Y/L)}{Y/L}$. This shows that the labour share is constant when the real wages (W/P) grow at the same rate as labour productivity (Y/L).

could also reflect individual outcomes of a more competitive labour market. During boom periods, rising employment generally improves the bargaining position of unions and workers in general, which is reflected in a higher wage mark-up. In addition, firms may offer higher real pay to motivate employees to work harder or longer hours, or even to encourage others to join the labour force.[19] More likely than not, the wage mark-up is—or is likely to be—procyclical.[20]

The battle of the mark-ups, expressed by equation (12.7), shows that the expected price level drives wages via the wage mark-up, and that wages drive labour costs, which drive prices via the price mark-up. Both mark-ups tend to move over the business cycle. It is a well-established fact that, combined together (the *product* $(1+\theta)(1+\gamma)$ of the two mark-ups in equation (12.7)) is procyclical. It tends to push the actual price level above its expected level in boom times, and to pull it down in lean years.

This procyclical outcome of the battle of the mark-ups is the foundation of the Phillips curve. We now use (12.7), which describes how the price *level* P is set, to conclude that the inflation rate, the *rate of increase* $\Delta P/P$ of the price level, is driven by two factors: (1) the combined change of the two mark-ups, and (2) the expected inflation rate. Using the symbol π to stand for the rate of inflation, and π^e to represent the expected rate of inflation, we represent this result as:[21]

$$\pi = \Delta(\text{mark-ups}) + \pi^e. \tag{12.8}$$

We are now just two short steps away from reaching the Phillips curve. First, the change in individual mark-ups will be different for different wage-bargaining situations, even at the same stage of the business cycle. Yet only the *average* bargaining situation in the economy is relevant for aggregate inflation—an average of many price changes. It follows that the aggregate mark-up moves with the business cycle. It rises when the real GDP Y moves above its trend $\overline{Y}$, i.e. when the output gap Y_{gap} is positive, and it declines when the gap becomes negative. Alternatively, we can use Okun's Law to describe mark-ups as rising when the unemployment rate U declines below equilibrium $\overline{U}$, i.e. the unemployment gap U_{gap} is negative, and declining in the opposite situation.

[19] This reasoning is developed in detail in Chapter 5.

[20] The wage share tends first to fall in an expansion, catching up with a significant lag and continues to rise past the peak of the usual cycle, as can be seen in a Burns–Mitchell diagram.

[21] For the mathematically minded reader: using the results of Box 6.3, (12.7) provides the rates of change of all variables, with the following approximation: $\pi = \frac{\Delta P}{P} = \frac{\Delta\theta}{1+\theta} + \frac{\Delta\gamma}{1+\gamma} + \frac{\Delta s_L}{1+s_L} + \frac{\Delta P^e}{P^e}$. Assuming that the normal labour share is constant, equation (12.8) follows, where Δ(mark-ups) is formally equal to $\frac{\Delta\theta}{1+\theta} + \frac{\Delta\gamma}{1+\gamma}$.

The second step recognizes that wage bargaining and price setting do not occur all at once, but are staggered over time. It is simply too costly for firms and workers to discuss wages all the time; after all, there's other work that needs to be done! Similarly, firms are not in a position to change their prices every day. Customers expect a certain stability in pricing, and may have even entered contractual agreements to guarantee that stability. As a result, while decisions are taken rationally, they are often based on older information. There is no single expected inflation rate π^e, but rather a complex mix of many expected rates, some relevant for the present, some coming from the past and looking to the future. For the aggregate economy, we can capture this inertial inflation in the economy as $\tilde{\pi}$, which we call the **underlying inflation rate**. As we will see, this is not only a realistic detail, it also turns out to be crucial for understanding the Phillips curve.

We have now reached our final destination. The process captured by equation (12.7) is turned into a relationship between actual and something which resembles expected inflation. Indeed, taking into account the cyclical behaviour of the combined mark-ups, the relationship (12.8) can be written as

$$\pi = \tilde{\pi} + aY_{gap} = \tilde{\pi} - bU_{gap} \tag{12.9}$$

The first part of the equation is an augmented aggregate supply curve. The second is an augmented version of the original Phillips curve. Both are augmented by consideration of underlying inflation.

The relationship (12.9) offers a simple but very effective accounting of what drives inflation. It asserts that inflation is driven by two main forces:

(1) What people expect or have expected inflation to be, now and in the future.

(2) Cyclical conditions, with wages and prices tending to rise faster in years of rapid economic growth and slowing down in harder times.

The positive parameters a and b simply describe how the mark-ups jointly respond to cyclical fluctuations, represented alternatively by the output and unemployment gaps. They say that a high level of activity leads to higher mark-ups.

12.4.5 More on the Underlying Rate of Inflation

In the last section we introduced a new concept, the underlying rate of inflation.[22] Because this concept is so important, we will spend a little more time discussing it here.

Recall that the expected price level in equation (12.6) is one important component of wage bargaining. It describes how wage negotiators expect the price level to evolve over the course of wage contracts under discussion. In that sense, the underlying inflation rate is forward-looking. In practice, it also includes a backward-looking component for several reasons. First, workers may use past behaviour of inflation to forecast the future. This may be a clever way of learning what the inflation rate is doing, or it may reflect a cheap alternative to hiring some economist to do the job. Second, cost-of-living clauses in labour contracts may actually mandate increases in nominal wages based on past inflation. Third, inflation forecasts are inevitably inaccurate *ex post*, and workers and firms will try to correct their past mistakes—if they can.

Consider the following examples of the last mechanism at work. Suppose a worker's wage contract incorporates a nominal wage increase of 4% in the year to come, reflecting the expectation of 3% inflation plus a 1% gain in labour productivity. After the fact, productivity indeed does rise by 1% but inflation turns out to be 5%! Workers, having lost 1% of purchasing power, will feel 'cheated' of not just 2% inflation forecast error, but also by the fact that their employer made off with 1% productivity. They are likely to try to recover these losses in negotiations for the following year.

Conversely, suppose inflation is only 1% after the year has passed. Then real wages are likely to have risen too fast and profits squeezed. Employers will no doubt use this fact to slow down wage increases in the future. Correcting for past forecast errors is an

[22] Sometimes the underlying rate is simply called 'inflationary expectations', relying more on the interpretation that it incorporates the anticipated inflation of firms and workers. Others use the term 'core inflation'—which is sometimes also used to describe the more stable inflation rate for non-food, non-energy items.

integral part of wage negotiations, and indeed this is why the logic of the purely forward-looking price level P^e in (12.8) has to be extended with a combination of forward- and backward-looking considerations. This is what underlying inflation is designed to capture.

Forecast errors are unavoidable. When inflation is reasonably low, the errors are small and the combination of catch-up and forward-looking forecasts encoded in the underlying inflation rate is acceptable. When inflation is substantial, say, around 20% per year, forecast errors can easily be much higher. In periods when inflation is high, it also tends to be more variable, leading to even larger forecast errors. This is why various mechanisms designed to cope with high inflation tend to emerge when the price level is rising fast. For example, explicit or implicit **indexation** clauses can link wages automatically to the evolution of the consumer price index. Another response is to shorten the contract length for wages and prices to shorter periods, like a quarter, or a month. In periods of extremely high inflation, wages and prices may even be reset weekly or daily! Avoiding such situations is why price stability, i.e. low inflation, is so desirable.

12.4.6 Completing the Picture: Supply Shocks

It is now time to look at the non-labour costs of production, studiously ignored until now. These costs correspond to the other factors of production—e.g. capital and land—as well as to intermediate inputs such as unfinished goods, primary commodities, and energy. In general, these costs follow the general trend of final goods prices. Their effect on inflation is then adequately reflected in the underlying inflation rate, adding nothing new to the analysis.

For this reason, these costs can be neglected most of the time. Now and then, however, special circumstances arise when non-labour costs do not behave so innocuously and significantly affect the short-run evolution of inflation, independent of the state of the economy, the market power of workers and firms, or what they have expected inflation to be. These events are treated as exogenous because they happen elsewhere or are special in the sense that they are not explained by the framework that we develop. Because they affect the costs of production, they are called **supply shocks**.

The leading example of a supply shock is a sharp increase in the price of oil, commonly referred to as an oil shock. Petroleum is a primary energy source and is an especially important raw material for the electrical generation, transport, chemical, and construction sectors. These sectors supply all sectors with inputs, so changes in oil prices have an important impact on production costs across the economy as well as on household budgets. Oil shocks have occurred at various points of time in the 1970s, early 1980s, and mid-2000s. Figure 12.11 gives a historical perspective. Another example of a supply shock is a deep depreciation, which renders imported good prices more expensive when evaluated in the domestic currency. Supply shocks may also be favourable. Oil prices, for instance, fell very fast in late 2008 after having risen considerably in 2007. The strong appreciation cheapens imports and for an open economy can have a significant *negative* impact on the short-term evolution of inflation, even while possibly hurting international competitiveness (raising σ) at the same time.

In the end, we think of non-labour costs as generally following underlying inflation $\tilde{\pi}$, but need to remember that supply shocks can and will occur—the sharp energy price increases seen in 2007–2008 are a clear reminder. To that effect, we allow for an occasional occurrence of a supply shock, a catch-all for exogenous disturbances affecting production costs. We denote this supply shock term by s and extend the aggregate supply curve and Phillips curve (12.9) in an important way:

(12.10) $\pi = \tilde{\pi} + aY_{gap} + s$ (aggregate supply curve)

(12.11)

$$\underset{\text{actual inflation}}{\pi} = \underset{\text{underlying inflation}}{\tilde{\pi}} \underset{\text{cyclical effect}}{- bU_{gap}} + \underset{\text{supply shock}}{s} \quad \text{(Phillips curve)}$$

As long as non-labour production costs are in line with the underlying rate of inflation, the supply shock term is zero. Occasionally, in unmistakable circumstances, it can be positive (an adverse supply shock) or negative (a favourable supply shock). On average it should be nil.

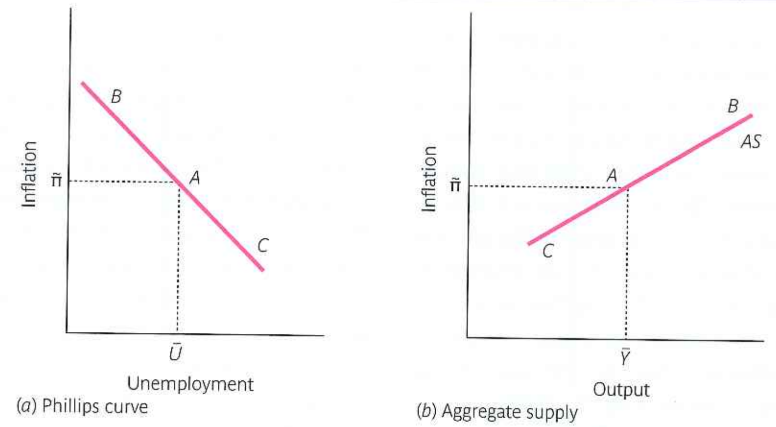

Fig. 12.12 The Expectations-Augmented Phillips and Aggregate Supply Curves
By definition, point *A* represents the case where actually observed inflation π is at its underlying rate $\tilde{\pi}$ and where unemployment *U* is at its equilibrium rate $\bar{U}$, and output *Y* is at its trend value $\bar{Y}$. When unemployment is low and output high, actual inflation is above the underlying inflation rate (point *B*). When unemployment is higher and output is lower than their respective equilibrium levels, actual inflation is below underlying inflation (point *C*).

Beyond changes in non-labour costs, two other types of supply shocks can be important. First, we assumed that the labour share is constant. A quick glance at Figure 1.2, however, shows that it is approximately trendless, but certainly not constant. Occasionally, it rises or declines, a process that can last several years. Such changes are usually the outcome of deep socio-political events that strengthen or weaken trade unions or employers' associations. We can take care of these events by interpreting changes in the labour share as supply shocks. An exogenous increase of the labour share, for example, adds to labour costs and can be treated as $s > 0$.

Second, supply shocks may also be caused by the government. An increase in certain types of taxes (for example, the value added tax) or prices for regulated utilities can raise the inflation rate even though the underlying rate and labour market conditions remain constant. The variety of taxes borne by firms relate directly to production and affect the final selling price—value added or excise taxes, profit taxes, establishment and property taxes, and so on. Other costs imposed by governments are implicit, but may have a significant impact (e.g. various labour regulations, environmental or consumer protection legislation).

12.5 Inflation, Unemployment, and Output

12.5.1 The Phillips Curve Rehabilitated

The inflation-accounting framework summarized by equation (12.11) effectively solves the puzzle of the Phillips curve. We can explain its persistence over decades, its instability over particular periods of time, and the impression that it shifts now and then. The original Phillips curve claimed that inflation depends only on the level of unemployment. The inflation account shows that cyclical labour market conditions do indeed matter, but so do underlying inflation, equilibrium unemployment, and occasional supply shocks. For a Phillips curve to be visible, these latter factors must be stable so that changes in inflation are mostly driven by cyclical conditions.

When these factors are not stable, the Phillips curve seems to vanish. This is precisely what happened during the 1970s, when price and commodity shocks became a major source of instability (see Figure 12.11). As inflation rose, underlying inflation rose as well and became more variable, reflecting rapidly changing expectations. The Phillips curve's demise reflects the emergence of underlying inflation and supply shocks. In addition, equilibrium unemployment rose in many countries during this period, for reasons explained in Chapter 5. These additional determinants of inflation, above and beyond cyclical factors, have helped to rehabilitate the Phillips curve.

How should we think of the Phillips curve if it proves to be unstable? Much as with the *IS* and *TR* curves, the answer is to keep clearly in mind which variables are endogenous and which ones are taken as exogenous. As long as the exogenous variables remain constant, the curve does not move. When they change, the curve will shift and apparently 'vanish'. This is the way the reconstructed Phillips curve explains the apparent puzzle. The modern Phillips curve is sometimes referred to as the expectations-augmented Phillips curve. This name emphasizes the central role of underlying inflation. Yet, it is more than that; it also recognizes that equilibrium unemployment $\bar{U}$ may change. It also allows for supply shocks.

Locating the Phillips curve requires a proper understanding of the underlying determinants of inflation. According to (12.10), when supply shocks are zero ($s = 0$) and when actual unemployment equals its equilibrium level ($U = \bar{U}$) and output is on trend ($Y = \bar{Y}$), the unemployment and output gaps are nil and actual inflation equals the underlying inflation rate ($\pi = \tilde{\pi}$). The pair ($\bar{U}, \tilde{\pi}$) serves as an anchor which pins down the position of the Phillips curve at any point in time. Without knowing the equilibrium rate of unemployment $\bar{U}$ and the underlying rate of inflation $\tilde{\pi}$, it is difficult or impossible to pin down the inflation-unemployment trade-off. And because both $\bar{U}$ and $\tilde{\pi}$ may change over time, it essential to understand the determinants of these factors.

This situation corresponds to point *A* in either chart of Figure 12.12. Point *A* uniquely determines the position of the Phillips or aggregate supply curves. The other points on the curves simply follow from allowing the unemployment rate to vary around its equilibrium level, holding underlying inflation constant and setting the supply shock to zero. This is the key intuition from the old Phillips curve that inflation varies with the business cycle. For instance, at point *B* unemployment is below equilibrium and output is above trend, so the demand pressure pushes inflation above its current underlying rate. Conversely, point *C* corresponds to the case where inflation is below its underlying rate because the unemployment rate is above equilibrium and output is below trend. In the left-hand chart of Figure 12.12, this traces out a downward-sloping schedule that resembles the old Phillips curve shown in Figure 12.3. In fact, it is the same but with a crucial difference: its position is determined by point *A*, that is, by the underlying inflation rate and equilibrium unemployment $\bar{U}$.[23]

12.5.2 Underlying Inflation and the Long Run

Underlying inflation captures the rate of inflation agreed upon during wage negotiations, both now and, for some bargaining parties, at points in the past. It has both a backward-looking (it reflects old forecasts of inflation today, as well as efforts to fix past errors of inflation forecasts used in previous contracts) and a forward-looking component (what will inflation be in the future). Somehow, it must be related to the actual rate of inflation. How this is so in the short run is considered in the next chapter. In this section we deal with the long run. We already saw in Section 12.3.3 that the **long-run Phillips curve** is vertical. We now look at how this comes about.

As negotiators consider the amount of inflation to be factored into wage settlements, they strive to guess it accurately. Of course, employees have an incentive to overstate the underlying rate of inflation, but employers have precisely the opposite incentive. If there were no uncertainty and both sides always knew *ex ante* what inflation would be over the lifetime of the wage contract, underlying and actual inflation would just be equal. Uncertainty therefore means forecasting (or guessing) underlying inflation. More often than not, the guesses are wrong. Yet, the principle of rational expectations from Chapter 7 means that wage negotiators do not make systematic forecast errors. Although forecasts are almost always incorrect, the errors are largely unsystematic and average to zero: in some years inflation is overestimated, in others it is underestimated.

[23] This is why the equilibrium rate of unemployment is sometimes called the NAIRU: the non-accelerating inflation rate of unemployment. At point *B* inflation accelerates above its underlying rate. At point *C* it decelerates. Only at point *A* is inflation stable.

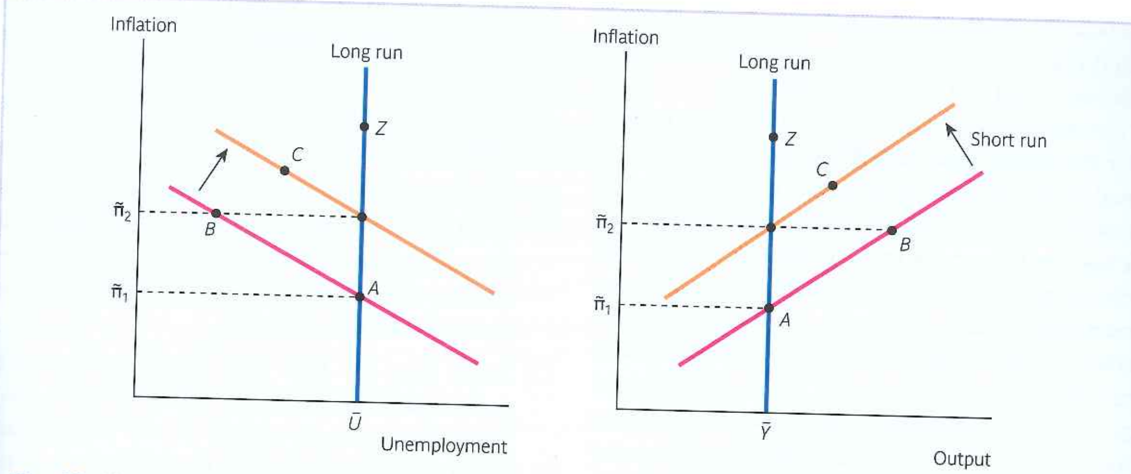

Fig. 12.13 From the Short to the Long Run

For a given underlying rate of inflation, as it moves from point *A* to point *B*, the economy can sustain lower unemployment and higher output but at the cost of higher inflation. This trade-off is not permanent, however. When underlying inflation rises to track higher actual inflation, the short-run Phillips curve shifts up. In the long run when the actual and underlying inflation are equal (for example, point *Z*), there is no longer any trade-off at all.

These observations have two important implications. First, there must be a link between actual inflation π and underlying inflation $\tilde{\pi}$. Underlying inflation must track, albeit imperfectly, actual inflation. The backward-looking component implies that underlying inflation *lags* behind actual inflation, but the forward-looking component implies that underlying inflation *leads* actual inflation. The interplay of both components is bound to be rather murky and difficult to detect precisely. As a result, it is far from obvious how to interpret Figure 12.9. There exist statistical techniques to do so, however, and they are routinely used by economists.

The second implication relates to the long run. The rational expectations hypothesis means that actual and underlying inflation cannot remain different in any systematic way. Any discrepancy between π and $\tilde{\pi}$ must be temporary. This fits well with other views of the long run. Obviously, over the years to come, we expect booms and recessions, possibly as consequences of supply shocks, positive and negative. We cannot really guess what the situation will be five or ten years from now. The best bet is the agnostic one, that there will be no shocks and that the economy will not be far from trend, with actual unemployment equal to underlying unemployment. For this reason, it makes sense to *define* the long run as the situation where the economy is back on its trend. This means that the Phillips curve is vertical. From equation (12.10), we know that when $s = 0$ and $U = \bar{U}$, then $\pi = \tilde{\pi}$. In a way, the long run is atemporal; the backward- and forward-looking components of underlying inflation have worked themselves out.

Thus, for the third time, we conclude that the long-run Phillips curve is vertical. We first encountered this result in Section 12.2.2 when we found that long-run price flexibility implies that output is supply determined. Next, in Section 12.3.3, we argued that long-run neutrality—which really follows from long-run price flexibility—implies a vertical Phillips curve. Now, we see that a vertical Phillips curve is also a consequence of the battle of the mark-ups, once we admit that there cannot be any permanent deviation between actual and underlying inflation.

The vertical long-run Phillips curve only tells us that the actual and underlying inflation rates are equal. It leaves entirely open the question of what these rates are. We will see in the next chapter that we need to bring back the demand side of the economy to pin down long-run inflation, actual and underlying. In fact, we already know that. The neutrality principle established in Chapter 6 relates long-run inflation to money growth, i.e. to monetary policy. In Box 12.1, we saw that this principle holds even if central banks set interest rates, because the neutral interest rate contains the central bank's target inflation rate, and nothing else. We will see that it takes time for underlying and actual inflation to catch up with each other and stabilize at whatever rate monetary policy allows for. Views vary about how quickly this happens, and herein lie some of the most fundamental controversies in macroeconomics, already encountered in Chapter 10 and to be studied further in Chapter 16.

The vertical Phillips curve carries a crucial implication: there cannot be a long-lasting trade-off between unemployment and inflation. Demand policies cannot move the actual unemployment rate permanently away from its equilibrium level. But the equilibrium level can very well shift over time, e.g. as labour markets undergo structural changes. Indeed, one of the implications of Figure 12.10 is that unemployment equilibrium must have shifted over the last decades. It massively increased in the euro area over the 1970s and 1980s—shifting the Phillips curve to the right—and then declined in the late 1990s. In the UK, the decline of the equilibrium unemployment rate started earlier and has been more pronounced—a legacy of Mrs Thatcher's supply-side policies aimed at weakening the bargaining power of unions, lowering unemployment benefits, and reducing the equilibrium unemployment rate.

12.5.3 Aggregate Supply

All of the previous reasoning applies to the aggregate supply curve shown in the right-hand chart in Figure 12.12 since it corresponds—via Okun's Law—to the Phillips curve. Its position is determined by the output trend and the underlying inflation rate, and it shifts when any of the exogenous variables ($\bar{\pi}, \bar{Y}$, or s) changes.

The aggregate supply curve conveys two important messages:

(1) In the short run, as GDP fluctuates about its trend growth path, the actual rate of inflation can differ from the underlying rate. In the absence of supply shocks, output and inflation move in the same direction.

(2) In the long run, GDP must return to its growth path, regardless of what the inflation rate is. Real forces determine the growth of real activity and the growth of money supply determines inflation.

The long-run aggregate supply schedule is vertical. It will, however, shift continuously to the right since, as a consequence of long-run economic growth, trend output keeps rising.

12.5.4 Factors that Shift the Phillips and Aggregate Supply Curves

The original position of the Phillips and aggregate supply curves in Figure 12.13 is determined by point *A*. This point corresponds to the initial underlying inflation rate, to the equilibrium unemployment rate, and trend GDP. Thus, the underlying inflation rate and the equilibrium unemployment rate or trend GDP are taken as exogenous when we draw the curves. The fact that they determine the position of the curve, giving us the location of point *A*, provides two reasons why the curves can shift. The first is a change in the underlying inflation rate: an increase in underlying inflation shifts the curves up. The second reason is that equilibrium unemployment and trend GDP may change. The supply shock s is the third exogenous variable that shifts the curve.

There is no presumption whatsoever that either underlying inflation or equilibrium unemployment are constant over time. If either changes, point *A* moves, and so does the whole Phillips curve. This implies that, potentially, there exists an infinity of Phillips curves, corresponding to the infinity of values that the underlying inflation rate or the equilibrium rate of unemployment can take. It just so happened that, over the hundred years surveyed by Phillips, the underlying rate of inflation and the equilibrium rate of unemployment

Table 12.2 Equilibrium Unemployment Rates

	1970	1980	1990	2000	2010
Austria	1.9	2.3	3.6	4.0	4.3
Belgium	2.9	6.3	7.9	8.0	8.2
Denmark	n.a.	5.2	6.7	5.1	4.9
Finland	3.7	3.4	6.2	9.8	7.8
France	n.a.	5.7	8.6	9.0	8.7
Germany	3.3	4.0	6.9	8.0	8.1
Hungary	n.a.	n.a.	n.a.	6.1	7.9
Ireland	n.a.	11.4	14.6	5.8	8.0
Italy	5.3	5.7	9.1	7.6	7.1
Japan	1.6	1.8	2.7	3.8	4.1
Netherlands	2.4	4.4	6.8	4.2	3.7
Norway	1.6	2.2	4.4	4.1	3.4
Portugal	n.a.	7.0	6.3	8.9	8.1
Spain	n.a.	6.4	14.5	12.4	12.6
Sweden	2.7	3.3	5.2	6.6	7.0
Switzerland	n.a.	0.6	1.8	3.3	3.8
United Kingdom	2.7	7.7	8.9	5.2	5.8
United States	5.3	6.6	6.2	7.2	5.3

Source: OECD, *Economic Outlook*.

did not change much, so there seems to have been just one Phillips curve.[24] This changed in the early 1970s when unemployment became much more stable—as the result of active Keynesian policies—while more volatile commodity prices became the dominating factor affecting inflation.

Shifts in the equilibrium unemployment rate occur occasionally. Table 12.2 presents some estimates of the equilibrium unemployment rate obtained by asking what rate would keep the Phillips curve unchanged when the actual and underlying inflation rates are equal and in the absence of a supply shock.[25] The table shows that the equilibrium rate rose in most countries in the 1970s and then declined in some countries two or three decades later, as the result of labour market reforms. This is one reason for the rightward shift of most Phillips curves in the 1970s, and for the subsequent leftward shift in those countries that managed to bring unemployment down in the 1990s or 2000s. Two examples are presented in Figure 12.10.

Trend output continuously rises as the outcome of long-run growth. To avoid dealing with a curve that constantly moves to the right, in later chapters we will draw the aggregate supply curve with the output gap $(Y - \bar{Y}) / \bar{Y}$ on the horizontal axis.

[24] There are good reasons for this. The period corresponds to the time of the gold standard and the Bretton Woods system, both of which constrained inflation from rising too much and kept underlying inflation in check.

[25] This is precisely the NAIRU as explained in footnote 23.