1

Macroeconomics: a succinct introduction

For many postgraduates macroeconomics is a totally new subject. For some undergraduates, it is useful to have a bird's eye view of the basics of the subject before plunging into more advanced material. This chapter therefore sketches in some basic elements of macroeconomics as a foundation for what follows. It takes nothing for granted except a knowledge of elementary microeconomics — essentially supply and demand. Macroeconomics is about the behaviour of whole economies but since these are made up of individuals and firms acting in the marketplace together with other institutions like governments that are responsible to individuals, it is natural to think of our understanding coming from microeconomics which analyses how people interact in a particular market and build up the whole economy from there. That is how the 'classical' economists — those who wrote before John Maynard Keynes — naturally analysed the economy. While their thinking was for a long time overlaid by the work of Keynes and his followers, in recent times it has once again become the core of our macroeconomic analysis; we shall begin with the early classical thinking, then sketch in the ideas of Keynes and of his later followers, before building up the modern macroeconomic theory that rests essentially on classical foundations with some extra elements drawn from Keynes' work. This approach, via a short history of macroeconomic thought, is the easiest way, experience shows, to reach an understanding of modern thought. (Students may like to consult a good intermediate textbook such as Parkin and Bade, *Modern Macroeconomics*, any edition, in conjunction with this chapter.)

THE CLASSICAL MODEL

The natural starting point for macroeconomics is the classical model, where it is assumed that prices and wages are flexible as in the familiar microeconomic model of a market with supply and demand. The difference from that model is that supply and demand are for the aggregate of all agents in the economy — it is assumed that one can meaningfully aggregate up or average the behaviour of all in this way; implicitly we assume that when income or activity is redistributed across people or firms the average is not much affected so that we can focus on the average forces affecting the economy rather than on their distribution across agents. Aggregate supply in this theory depends on technology, preferences and relative prices. Demand depends on the quantity of money and on incomes.

Supply

Figure 1.1 shows four quadrants, bottom left is the labour market, top left the production function, top right a 45° graph transferring output through to the bottom right, which is the supply curve between the price level and output.

The production function, PF, which can be thought of as the optimal behaviour of the average firm times the number of firms, takes capital as given and hence shows diminishing returns of output (Q) to increasing labour input (L). From it is derived the demand for labour, DD, where (for the average firm) the real wage equals the marginal product of labour, or equivalently the price = marginal cost, the first-order condition of a profit maximum. The supply of labour, SS, is shown with a flattish slope, indicating that the substitution effect of rising real wages, $\frac{W}{P}$, is substantially larger than the income effect; this could come about because either workers substitute effort across time or unemployment benefits are generous and long-lasting (as discussed below, respectively, in Chapters 3 and 9).

Two aggregate supply, AS, curves are shown. One is vertical, indicating that if all prices (P) rise the supply of and demand for labour will not change since wages (offered and demanded) will rise in proportion, and so with relative prices the same there is no inducement to alter offers. The second curve is upward sloping; the reason is that as prices rise people do not realise that all prices are rising (inflation). Workers in particular do not realise this and in this diagram are assumed to be slow therefore in raising their wages. Hence as prices rise wages lag behind and real wages fall to the right along the demand for labour curve (this

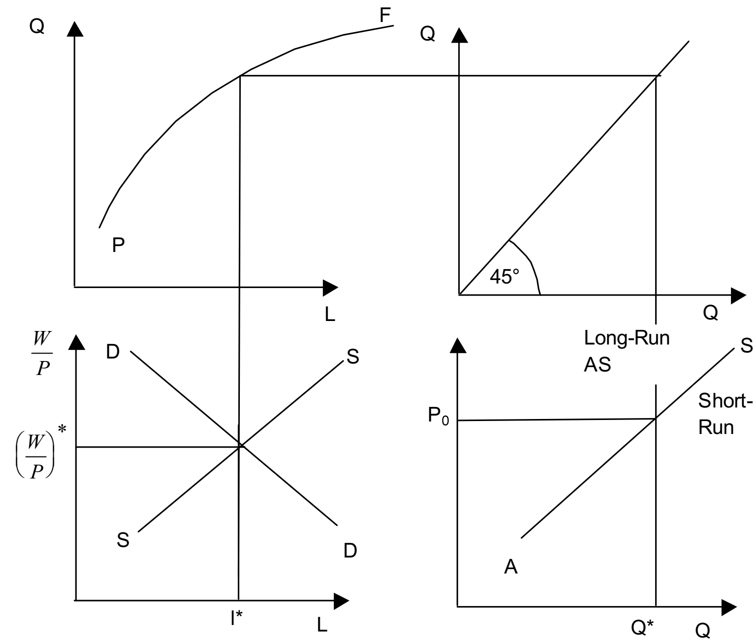

Figure 1.1: Aggregate Supply

can be thought of as a rightward shift of the labour supply curve). This generates more employment and more output (this is developed fully in Chapter 3, first section, 'The New Classical Phillips Curve').

Demand

We suppose for simplicity an economy where there is no alternative to using cash for paying workers and for buying goods. In this 'cash-in-advance' transactions technology, we also assume there is a payment period, one month for example. Then the economy 'turns money over' as illustrated in Figure 1.2).

At the end of the month workers are paid and spend their cash through the succeeding month. As cash moves from workers to firms the total holding of it remains constant — this is the demand for cash for transactions. If we measure money output, PQ, in a year or at an annual rate, it follows that cash required is equal to one month's output and so the demand for money, $M^D = (1/12) \times PQ$ (where 1/12 is the fraction of output required as money, K) or equivalently $M^D \times 12 = PQ$

(where 12 is the 'velocity', V, of money); respectively the Cambridge (UK) and the quantity theory ways of writing money demand.

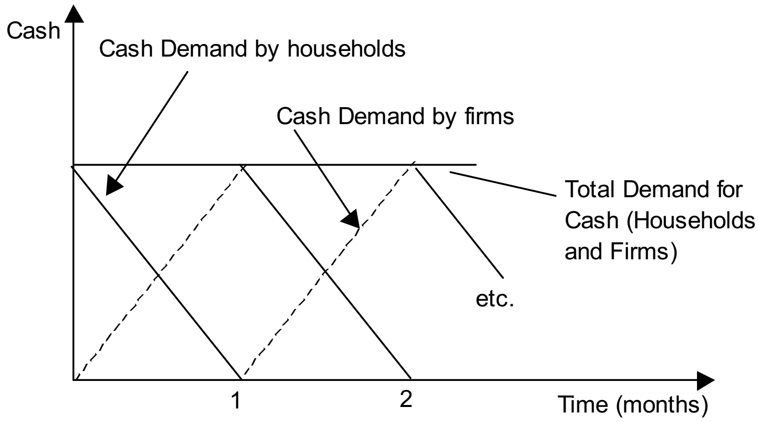

Figure 1.2: Transactions Demand for Money

Begin by supposing that the money supply M equals M^D, that is there is equilibrium in the money market. Under our assumptions this also implies that income equals spending (as in figure 1.2). Now suppose the government increases the money supply by 10 per cent (by printing more) and gives the extra money to workers at the end of the month. At existing $PQ(0)$ the workers plan normal spending of $PQ(0)$ which is their normal income and firms need to pay out only PQ(0) in wages; so neither workers or firms need to hold on to the extra money for future months. The workers will therefore spend the extra money on buying goods, driving up PQ along the AS curve. At this point $PQ(1)$ is higher; firms hold more money to pay for extra work. But when M returns to consumers, still $M^S > M^D$ and again $AD > PQ(1)$. Hence the process will repeat itself and PQ will rise again. Only when PQ has risen sufficiently for $M^S = M^D = \frac{PQ}{V}$ will we have $AD = PQ$. This is where $PQ = MV$. This is the 'aggregate demand' consistent with the stock of money (Figure 1.3).

In classical thinking prices and wages move rapidly to restore the economy to its potential, Q^*; hence interest rates move to ensure that saving equals investment — for example, if at Q^* people decide to save more this would mean more funds would be offered to investors in plant and machinery and interest rates (the cost of these 'loanable funds') would fall until the market for savings equilibrated with investments. So we can say that interest rates are not affected by the demand for or

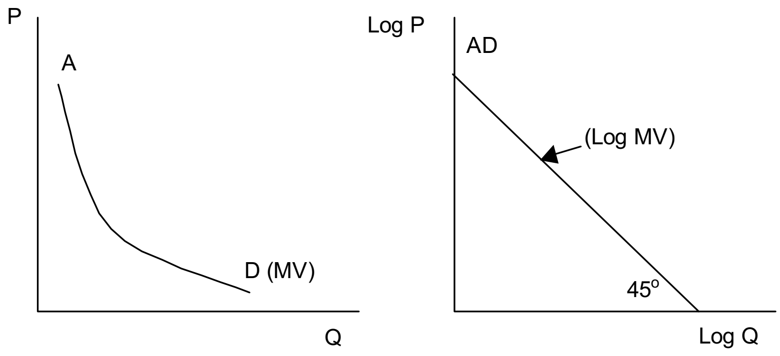

Figure 1.3: Aggregate Demand

supply of money, they are merely a (small) factor in determining the demand for money. Figure 1.4 illustrates the market for loanable funds.

Short- and long-run equilibrium in the classical system in response to a monetary expansion can be summarised by Figure 1.3. In the short run output rises because wages lag behind, but in the long run output is driven back to Q^* as wages rise in response to workers' realization that it is general prices that are rising.

To allow for the effect of government spending, G, and taxation, T, we simply note that $T - G$ is part of savings, viz. savings by government; while private savings are now $PQ - T - C$. So total savings are $T - G + PQ - T - C = PQ - G - C$. Hence for example a rise in government spending, taxes constant, would shift the SS curve in Figure 1.4 leftwards and raise interest rates; it would not raise output even in the short run. (It could do so indirectly via the effect of higher interest rates on the demand for money, a factor we introduce in the next section — by reducing this demand it would shift the aggregate demand curve outwards. However this effect is likely to be quite small, since the transactions demand for money is rather inelastic.)

Allowing money to respond to interest rates: the Wicksellian mechanism

We have assumed up to now that interest rates have a negligible effect on the demand for money, and that therefore we could treat the market for money and bonds as essentially unconnected. But this can only be a stringent simplifying assumption. In fact interest rates do affect the demand for money because people and firms can put their money into

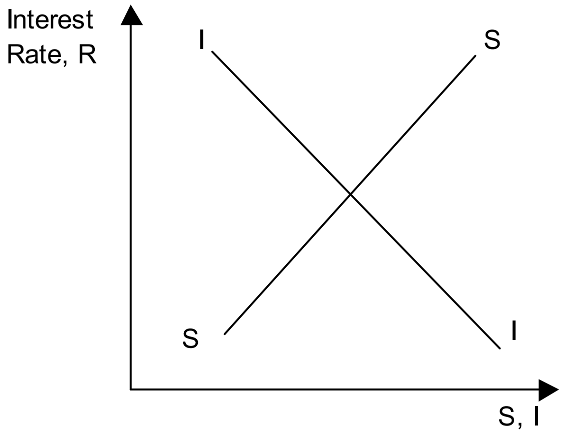

Figure 1.4: Supply and demand for loanable funds

deposits (short-term bonds or loans, in other words) as an alternative to continuously holding the money they will eventually need for transactions. Households can, for example, deposit half their month's wages for the first half of the month and then withdraw them half-way through the month in time to carry out the second half of the month's transactions; firms can in turn deposit their takings from the first half of the month and then withdraw them at the end of the month, just in time to pay the wages. This is illustrated in Figure 1.5. Clearly people decide whether to carry out this deposit strategy by weighing up the cost versus the benefit: the cost is making two trips to the bank or loan market (one to deposit, one to withdraw) per month; the benefit is the interest received. The higher the interest rate the more people may decide to deposit; for example, households could deposit three-quarters of their wages and then withdraw a quarter after week 1, a quarter after week 2, and the last quarter after week 3, while similarly firms could deposit all their first three weeks' takings and withdraw them at the end of the month — making four trips in all for both firms and households. We now find that the demand for money falls with higher interest rates, with the demand for deposits moving in the opposite direction — imagine Figure 1.5 with twice as frequent deposits and withdrawals, implying that the demand for money is half as much again as in the figure.

Now that we have linked the two markets the question arises: what happens when extra money is printed and, say, distributed to households in our previous experiment? Previously we said that households and

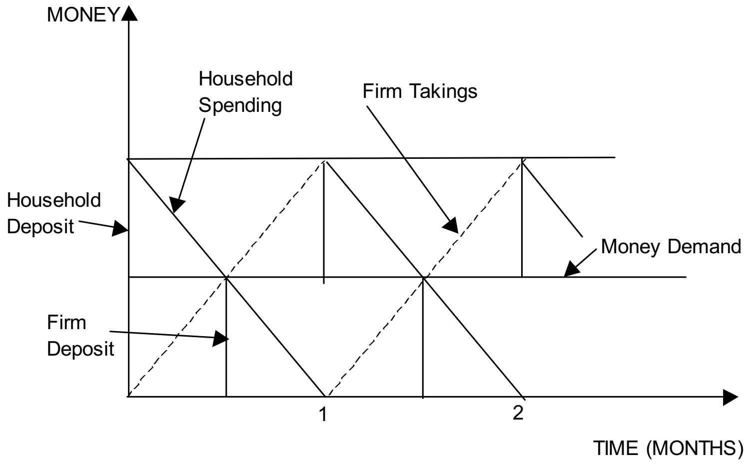

Figure 1.5: Money demand when money is deposited/withdrawn once per month

firms would spend any money surplus to their transactions needs. But now we would argue that they would put the extra money on deposit mainly, perhaps only spending a very small part of it; the reason would be that they would naturally wish to spread out the extra spending over a long period, given that they can now deposit it and obtain interest while waiting to spend it.

If they do this, then as the Swedish economist Knut Wicksell pointed out (Wicksell, 1970), in a way that foreshadowed Keynes' views of asset markets and demand, the depositing of the extra money would act indirectly to create excess demand by stimulating investment and reducing saving (stimulating consumption). Before the money injection, savings were equal to investment and hence demand for output was equal to supply of output: $S = I$ implies that $S = PQ - C - G = I$ so that $C + I + G = PQ$ that is demand equals output (as supplied along the AS curve). He thought of this situation as one where interest rates were at their 'natural' level: namely such as both to clear the loanable funds market and to set demand for output equal to supply of output.

But after the injection interest rates fall below this natural level, stimulating consumption and investment above these levels, and so creating excess demand. This would drive up prices or output in just the same way as we described before when we did not allow people to deposit their money. Aggregate demand would rise until the demand for money rose to the new level of money supply, at which point of course interest rates would return to their natural level as the injection of excess money into the loanable funds market would cease (see Figure 1.6). This mechanism of 'monetary transmission' via interest rates is close to what Keynes proposed and what we think today.

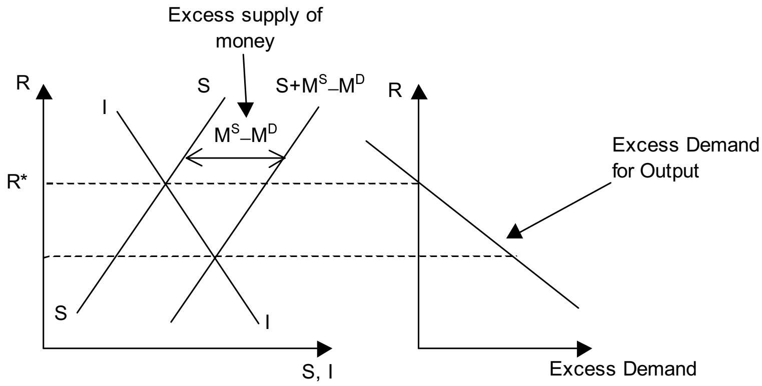

Figure 1.6: Wicksell's monetary transmission

We can think of the central bank carrying out monetary policy by 'open market operations' in the bond market which inject more or less money into the economy; for example, by buying bonds, the bank supplies extra loanable funds to the the bond market, shifting the SS curve rightwards by printing extra money. This drives down interest rates, creates a gap between investment and saving, that is, excess demand for output. In the following period the extra money has entered private sector holdings but being surplus to money demand is put back into the loanable funds market, keeping interest rates down until the excess demand has driven up PQ sufficiently to raise demand for money to equality with the extra supply.

We can also analyse what would happen if demand for money fell — say for the reason discussed above where people decide to take advantage of the interest rate opportunities from depositing (that is lending) their money. The surplus money would go into the loanable fund market like the extra money printed by the central bank just discussed; triggering the same process whereby PQ must rise until the available money supply is willingly held.

The algebra of the Classical system is:

Production function: $Q = f(L)$

Demand for Labour: $W/P = f^{\prime\prime}(L)$

Supply of Labour: $L = g(W/P^E) - P^E$ (the price level expected by workers) in the short run is assumed to lag behind the actual price level.

Aggregate Demand: $M.V(R) = PQ$ (this assumes that monetary disequilibrium has already been eliminated by the Wicksell process).

Loanable Funds: $S(R,\ PQ) = I(R)$

THE KEYNESIAN MODEL

The ideas of Keynes are best understood as a special case, or a degenerative distortion, of the classical model — created by assuming that wages and prices are inflexible. Nowadays we try to rationalise this assumption by referring to nominal wage contracts and 'menu costs' of raising prices. But neither Keynes nor his post-war followers did so: they simply pointed to the facts of wage and price rigidity and said that non-economic factors (such as conventions of fairness, policed by unions among others) could be responsible.

The simplest framework, which we will use here, assumes that both wages and prices are rigid — that is to say, are set by suppliers who thereby state their willingness to supply whatever is demanded at those rates. From time to time they may change them — in response to excess demand or supply, or to general inflation — but that does not alter the point that whatever they have set their supply is driven by the demand they face. It is natural to assume in the Keynesian model that wages are governed by bargaining between workers or unions and firms and that prices are set by firms operating in imperfect competition (where each firm has a different product and some short-run 'monopoly power' — that is a downward-sloping demand curve for its product). These assumptions imply price- and wage-setting behaviour (whereas perfect competition, for example, where wages and prices are given to the individual, would not).

With prices set in this way plainly demand sets output — in microeconomics we would say that the supply curve is horizontal. Keynes (and his co-workers, including Richard Kahn) evolved a theory of demand feedback which he called the 'multiplier'; the idea is illustrated on the 'Keynesian cross' diagram, Figure 1.7. Consumption, C, depends on income (= output); other demand is treated as exogenous — notably investment, I, which is mainly dependent on businessmen's optimism ('animal spirits'). Without government, aggregate demand, $AD = C + I$, therefore rises with income — the AD curve which slopes at an angle less than 45°. However, output is set by producers equal to demand; hence $Q = AD$ — the 45° line. Realistically Keynes thought of this output as responding with a lag to demand (initially orderbooks or inventories would respond). Output would go on moving until $Q = AD$, yet of course AD would respond (feedback) to changing Q. The two would settle down where the two curves intersected (Figure 1.7).

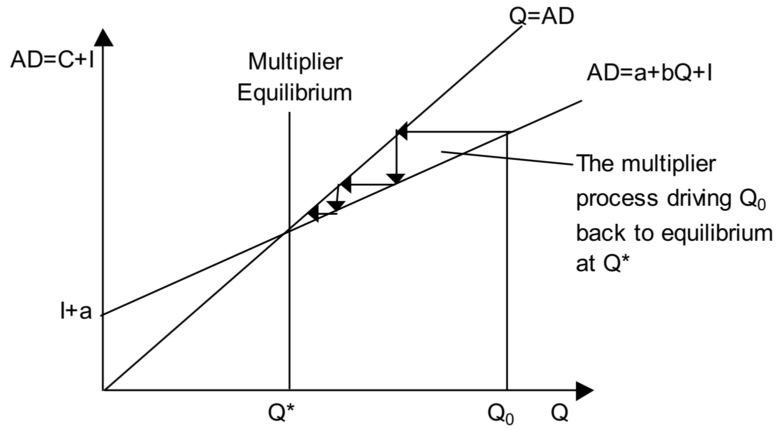

Figure 1.7: The Keynesian Cross

The algebra is:

$$C_t = a + bQ_t$$
$$AD_t = C_t + I$$
$$Q_t = AD_{t-1}$$

The equilibrium of this, where $AD = Q =$ constant, is found by setting Q and AD constant and solving the equations: $Q = (a+I)/(1-b)$. $1/(1-b)$ is the multiplier.

This set-up can be extended to allow for government: its spending, G, is exogenous and its taxation can be represented by a simple proportional tax rate, τ. This means that in Figure 1.7, I now becomes $I + G$ and that the slope of the AD curve is flatter because now extra income is partly taxed and does not have such a large effect on consumption. Algebraically:

$$\begin{aligned} C_t &= a + bQ_t[1-\tau] \\ AD_t &= C_t + I + G \\ Q_t &= AD_{t-1} \end{aligned}$$

The multiplier is now $1/[1 - b(1 - \tau)]$, smaller than before.

The IS curve

We now notice that when $Q = AD$ we also have savings (as planned — that is consistent with the consumption function above) equal to investment (as planned exogenously above) since it implies that $Q - C - G = I$. (Q is income, C is private consumption and G is government consumption; it is usual to include government investment in I.)

In passing, we should not be confused by the fact that in the short run savings can be unplanned (for example, when income rises it may take time to adjust spending) and also investment can be unplanned (for example, when demand falls inventories may increase as it takes time to adjust output; the rise in these inventories is included in actual investment). It is also a confusing fact that by the conventions of national accounting actual savings are always equal to actual investment because output minus actual spending = rise in inventories, DV by definition; therefore Q = actual consumption + G + investment in plant and machinery minus DV = actual $C + G +$ actual I. Hence $Q -$ actual $C - G =$ actual I. This, however, is an identity with no behavioural significance, unlike in the above paragraph.

In the Keynesian model planned savings = planned investment when output has settled down. Thus we would say that output is what drives them to equality, not interest rates as in the classical system. What then do interest rates do? According to Keynes they affect demand and so output — they merely change the output level at which savings are driven to equality with investment. (Unlike the classical system where output is driven to potential by supply via prices moving, and interest rates then drive savings to equality with investment.) Interest rates affect demand straightforwardly: as they rise consumption and investment fall because of the higher cost of borrowed funds and return to saving.

The result is a shift in the AD curve in the Keynesian cross diagram; we can plot the different Q equilibria as interest rates shift the curve. This is the IS curve, that is the curve showing how output changes with interest rates given that the goods (output) market is in equilibrium or equivalently that planned $S=$ planned I (Figure 1.8).

The LM curve

But how are interest rates determined? In the classical system by savings (supply of funds) and investment (demand for funds) clearing the market for loanable funds; but in Keynes' system savings and investment

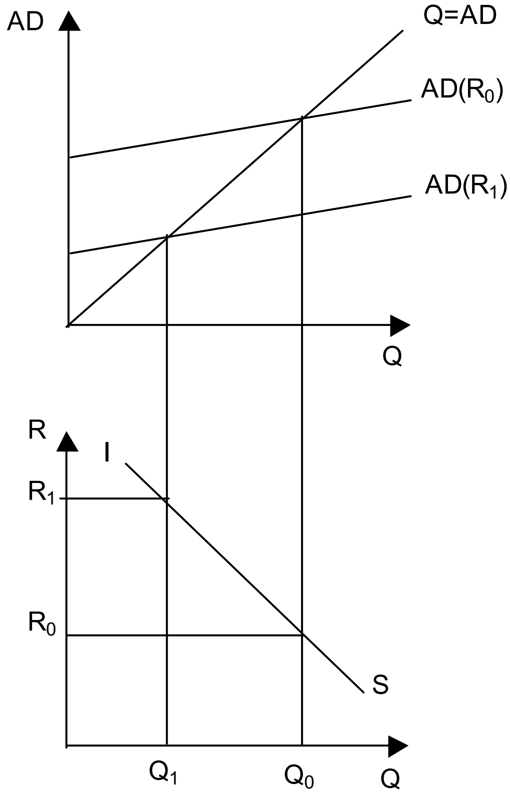

Figure 1.8: The *IS* curve

determine output (also affected by interest rates through their effect on savings and investment). But interest rates are not pinned down — the *IS* curve shows that instead they can vary, and as they do so output will also vary. So what pins them down?

In the classical world the market for money is cleared by income in the short run and then rather quickly by prices. This does not happen in Keynes' world. Prices are fixed or moving slowly in response to price setters' strategies. Output (income) is responding to aggregate demand which has nothing directly to do with the demand for money; it is certainly not moving to make demand for money equal to its supply. We must now ask what is happening in the money market.

Keynes took over the classical analysis of demand for money above but he also emphasised the alternatives to holding money (defined as non-interest-bearing cash): these were short-term or long-term ‘bonds’, that is loans. If we divide the world into private and public sectors, this meant that the private sector holding money (a liability of the government) could also hold public loans as an alternative. The commercial banks are assumed to be subject to government regulation (for example, by reserve ratios) and their deposits are usually treated as equivalent to cash; so cash and non-interest-bearing deposits are ‘money’. Besides holding this money for transactions, private people and firms might hold it because they wanted to get out of bonds for speculative reasons: if interest rates were low (bond prices high), people would expect bond prices to fall back and so hold more cash and vice versa when interest rates were high (bond prices low). Private sector financial wealth in a closed economy is by definition equal to the value of public (and banking) sector liabilities, money, M and government bonds, B.

Holdings of equities are what firms issue to households as paper claims to the capital stock which of course households own; these equities have to be held by someone and so their price moves until they are willingly held by households. But note that as they are a liability of firms they cancel out within the private sector and hence do not constitute a financial asset for the private sector as a whole. So there is no way the private sector can move out of bonds and money into equities; all that would happen if they tried is that the equity price would go up to frustrate them. That equity price may have implications for the rest of the economy (for example, if it is high, it implies that expected returns on the capital stock are high and so investment is likely to increase); but as far as the market for bonds and money are concerned it has no implications of any importance as we shall see.

The total financial wealth of the private sector, consisting as we have seen of money and bonds, has therefore to be held either as money or bonds:

$$M + B = M^D + B^D$$

Hence when $M = M^D$, $B = B^D$. In other words the market for money is the mirror image of the market for bonds, as illustrated in Figure 1.9).

Hence interest rates are determined by the ‘stock markets’ in Keynes’ model: specifically the markets for government stock (or equivalently the ‘market for money’). Here we see that as output is driven higher or lower by the IS curve (the multiplier process) the demand for money moves higher or lower with it for transactions reasons; and this in turn drives interest rates higher or lower as the financial markets equilibrate. The

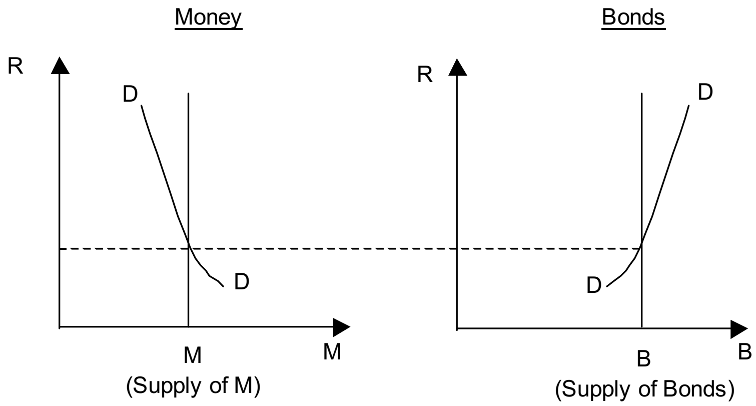

Figure 1.9: Demand for money and bonds

resulting relationship between output and interest rates is the LM curve (Figure 1.10).

So Keynes rearranged the relationships in the classical model, crucially allowing output to be driven far from potential because he cut off the equilibrating effect of flexible prices.

The algebra of the Keynesian model is:

$$(IS) Q = f(G, \tau\ )$$

determining Q

$$(LM) M = m(R, Q; P)$$

with P given; determining R.

THE PHILLIPS CURVE AND THE NEO-KEYNESIAN SYNTHESIS:

We now note that the Keynesian system has a hole — price determination. For a long time after the Second World War economists were not concerned because the assumption that prices moved slowly — at or close to zero inflation — stood up. But systematic policy stimulus keeping economies at low levels of unemployment produced inflationary behaviour that the Keynesian model could not account for. The first reaction of economists was to extend it to include a 'disequilibrium theory' of price change. The idea was simple: output was determined in the short

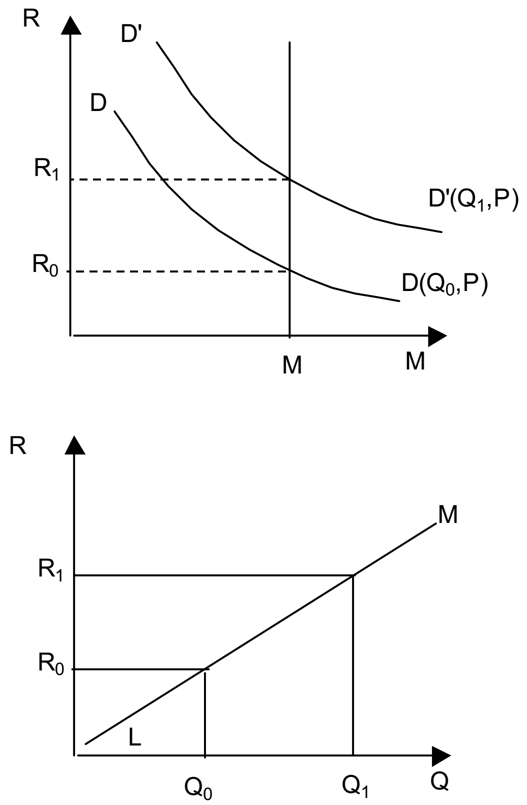

Figure 1.10: The *LM* curve

run by demand as in the Keynesian model but prices would be changed by price setters in response to the gap or disequilibrium between this and their supply as measured by output potential, the equilibrium produced by technology and preferences. Such a relationship was first estimated by A.W. Phillips (1958) for UK data on inflation and unemployment going back a century: hence the Phillips Curve. Since unemployment and output are closely related through the production function assuming some measure of the labour force given exogenously, the Phillips Curve can be drawn either with unemployment along the horizontal axis or with output along it but with a reversed slope.

This relationship worked well for the 1950s and 1960s but clearly broke down in the 1970s as rising inflation was associated with rising unemployment. The reason was suggested by Phelps (1970) and Friedman

(1968) independently: price setters were aiming to adjust relative prices in response to excess demand and so their actual price change would be equal to the response to excess demand, plus expected inflation — only then would relative prices rise by the response to excess demand. This expectations-augmented Phillips Curve is shown in Figure 1.11.

A further development was a theory of expectations formation rooted in rational calculation — 'adaptive expectations'. In parallel with the disregard of inflation had gone a casual attitude to expectations of inflation: it was typical to assume that expected inflation was equal to current inflation ('static' expectations); for some purposes this might be altered to 'extrapolative' expectations where the recent change would be extrapolated into the future. Both were arbitrary. Now economists such as Marc Nerlove (1958) and Phillip Cagan (1956) replaced them with the idea that expectations would be revised in response to past errors:

$$\dot{p}^e_t - \dot{p}^e_{t-1} = k[\dot{p}_t - \dot{p}^e_{t-1}]$$

where $\dot{p}^e_t$, $\dot{p}_t$ = expected and actual inflation. Under certain circumstances it is rational to adjust expectations in this manner — in particular, when people are ignorant about the model and the policy process it is a good approximation to a rational learning procedure.

Finally, in a world of inflationary expectations we must change the IS curve: demand now responds not to (nominal) interest rates but to real interest rates — that is interest rates minus the expected rate of inflation for the period of the loan. The reason is plain: consumers and firms borrowing to undertake purchases pay back their loans in money that is worth less by the rise in prices, hence the true expected annual cost to them of the loan is the rate of interest less the inflation rate they expect. (The same would apply if they were paying for their purchase by running down their savings — then it would be the real interest they expect to sacrifice on the savings deposit.)

We now have a model of the IS and LM curves to determine demand, output and interest rates in the short run; and a Phillips curve augmented by adaptive expectations (Figure 1.11) to determine output, inflation and interest rates in the long run, essentially as in the classical model. This neo-Keynesian synthesis between Keynesian and classical thinking dominated macreconomic thinking in the 1970s and 1980s and is still highly influential today. (For further reading about this model we recommend Laidler and Parkin (1975) and for mathematical back-up on solution methods a textbook on mathematics for economists, such as Chiang, 1984).

The model can be represented by five equations set out in Box 1.1; they are respectively the IS, LM and Phillips Curves, the adaptive expectations process, and the definition of the real interest rate. The equations are deliberately simplified to the maximum — especially, output and prices are all expressed as logs (hence the assumption is made that the relationships are linear, that is additive, in logs) and there are no lags except in expectations formation. Notice that in the IS curve real interest rates employ inflation expected for the coming period, $\dot{p}^e_{t+1}$, rather than inflation expected last period for this, p^e_t, which enters the Phillips Curve.

This model can of course be solved on a computer with numbers assigned to the parameters. But we would like to know its qualitative properties as well. There are two main ones: the nature of its equilibrium and its 'dynamic' properties (that is how it moves from one equilibrium to another when a shock occurs).

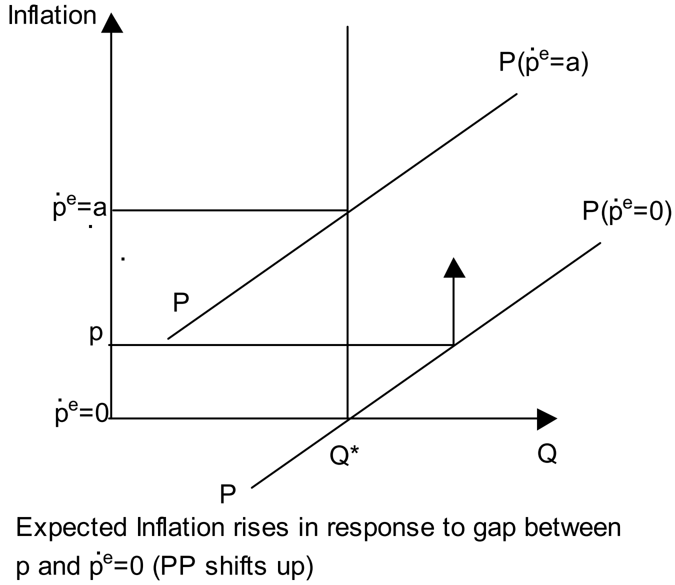

Figure 1.11: The expectations-augmented Phillips curve with adaptive expectations

Box 1.1 shows how the relevant inflation equilibrium — that is resting point — can be found for this model. Box 1.2 shows how the dynamic properties can be found.

Some notes on the equilibrium

Equilibrium is where the model economy comes to rest, that is, its endogenous variables attain a constant or steadily moving value. What this will be in detail depends on the motion of the exogenous variables (it is possible for a model not to have an equilibrium at all if these move in particular ways). In our case we will assume that $\bar{y}$ and $\bar{d}$ are both constant and that $\bar{m}_t$ is growing steadily; this will allow us to look and (as we shall see) find an equilibrium where inflation, expected inflation, interest rates (real and nominal) and output are all constant — and prices are growing steadily (in line with the constant inflation rate).

Box 1.1 shows the detailed steps. The meaning of these steps can be quickly explained. First, the Phillips Curve with adaptive expectations implies that the economy must be at $\bar{y}$, or on the 'long-run Phillips Curve' — by which is meant the relationship between inflation and output when expected inflation is equal to (has caught up with) actual inflation. This relationship can be understood by a policy experiment (a foolish one). Suppose the government decided to stimulate the economy and keep it above $\bar{y}$ permanently: inflation would rise initially along the short-run Phillips Curve, but then expected inflation would rise too, and inflation would rise further. It can easily be checked that inflation will rise indefinitely, hence the vertical curve in the long run — thus illustrating the point that for inflation to settle down, output cannot permanently exceed (or be less than) $\bar{y}$. This point is sometimes referred to as the accelerationist proposition. Figure 1.12.

Note that we still do not know where inflation will settle, only that output is at $\bar{y}$. So secondly we can enquire what inflation will be: this is given by the LM curve. If we ask what the implied rate of change of the variables in the LM curve must be in equilibrium, we can see that since neither interest rates nor output are changing then the rate of change of the money supply must be equal to that of prices, that is inflation equals the growth rate of money, the fundamental proposition of monetarists such as Milton Friedman, Karl Brunner and Allan Meltzer. This gives us the point on the long-run Phillips Curve where inflation will settle.

Given output at $\bar{y}$ we can establish what real interest rates will be from the IS curve: real interest rates must move to equate demand with the available supply — the 'crowding-out' principle (making space for any government demand pressure, $\bar{d}$, by reducing private demand). We can show this in the IS–LM diagram as follows: clearly in equilibrium the IS curve must be located so that output is $\bar{y}$ at the prevailing interest rates, the curve is shifted upwards by expected inflation and rightwards by $\bar{d}$, hence the point on this IS curve above $\bar{y}$ will be the implied nominal

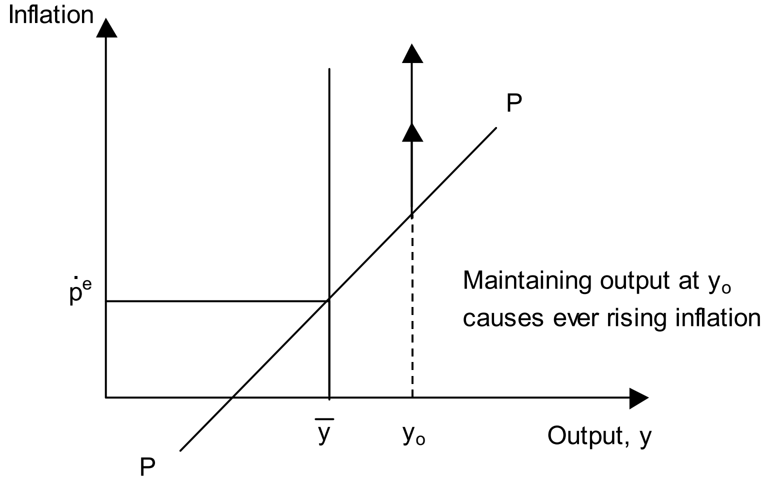

Figure 1.12: The accelerationist proposition

interest rate made up of inflation plus the necessary real interest rate — Figure 1.13.

This leaves us last with the price level. Diagrammatically we can see that the LM curve must intersect the IS curve at the point that lies above $\bar{y}$. We must shift the LM curve to this point: given that the money supply is on its steady growth path, the only way to do this is for prices to change. The LM curve is therefore solved for the price level.

Box 1.1

THE NEO-CLASSICAL/KEYNESIAN SYNTHESIS MODEL AND ITS EQUILIBRIUM

$$y_t = \overline{d} - \alpha r_t \text{ (IS)} \quad \text{(a)}$$
$$\overline{m}_t = p_t + \gamma y_t - \beta R_t \text{ (LM)} \quad \text{(b)}$$
$$\dot{p}_t = \dot{p}_t^e + \delta(y_t - \overline{y}) \text{ (Phillips curve, PP)} \quad \text{(c)}$$
$$\dot{p}_t^e = \lambda \dot{p}_{t-1} + (1-\lambda)\dot{p}_{t-1}^e \text{ (adaptive expectations, AE)} \quad \text{(d)}$$

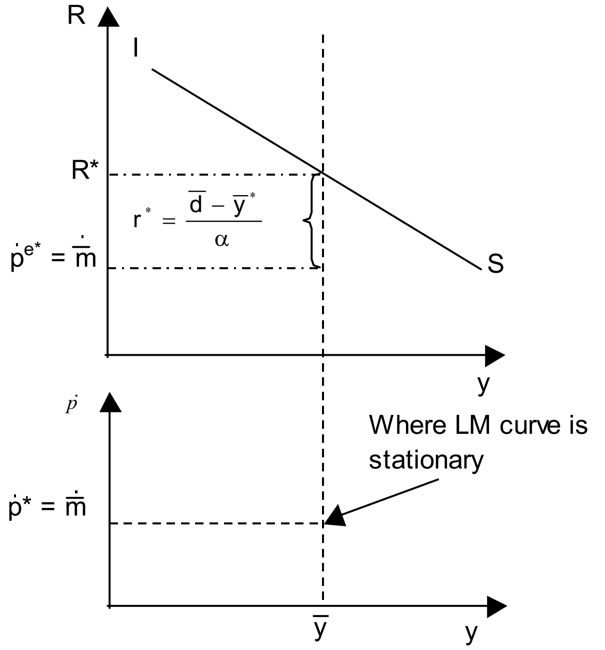

Figure 1.13: Inflation and interest rate equilibrium

$$R_t \quad = \quad r_t + \dot{p}^e_{t+1} \text{ real interest rate (RR)} \qquad \text{(e)}$$

Remembering that $\Delta \ln x_t = \frac{\Delta x_t}{x_t - x_{t-1}}$
y, p and m are respectively ln of real GDP, prices and money supply.
$\dot{x}_t = \Delta x_t = x_t - x_{t-1}$
$\dot{p}^e_t =$ inflation expected last period for this period
$R_t =$ interest rate (fraction p.a.)
$r_t =$ real (inflation-adjusted) interest rate (fraction p.a.)
The model's equilibrium
Suppose $\bar{d}$, $\bar{y}$ are constant; $\bar{m}_t$ is rising at a constant growth rate, $\Delta\bar{m}_t = \dot{\bar{m}}_t = \dot{\bar{m}}$. Then the model has an equilibrium where y_t, r_t, R_t, $\dot{p}_t$, $\dot{p}^e_t$ are all constant (and p_t is growing at a constant rate). Find it by setting all these variables to their constant values (for p_t at its equilibrium value along its constant-growth path) in the above equations, and then solving out for these values:

AE gives $0 = \Delta\dot{p}^{e*} = \lambda(\dot{p}^* - \dot{p}^{e*}) \Longrightarrow \dot{p}^* = \dot{p}^{e*}$
PP then gives $\dot{p}^* = \dot{p}^* + \delta(y^* - \overline{y}) \Longrightarrow y^* = \overline{y}$
IS then gives $r^* = \frac{\overline{d}-\overline{y}}{\alpha}$
RR gives $R^* = r^* + \dot{p}^*$
LM, differenced, gives $\Delta\bar{m} = \dot{p}^* + \gamma\Delta\overline{y} - \beta\Delta R^*$ so that as $\Delta\overline{y} = \Delta R^* = 0$ it follows that $\dot{p}^* = \Delta\overline{m}$
Finally solve for p^* from LM as

$$p_t^* = \overline{m}_t - \gamma\overline{y} + \beta[r^* + \dot{p}^*] = \overline{m}_t - \gamma\overline{y} + \beta[\frac{\overline{d} - \overline{y}}{\alpha} + \Delta\overline{m}]$$

This final step tells us an important feature of this model: that when the growth rate of the money supply rises the price level will rise by more than the money supply level — and vice versa when the money growth rate falls, prices fall by more than money supply. The reason for this is that the demand for money responds to interest rates: when money supply growth increases, interest rates rise and reduce money demand, so causing the excess money to be spent until prices have risen enough to restore the money demand to equality with money supply. This is sometimes referred to as 'price overshooting'. Consider, for example, a hyperinflation caused by a very large rise in money supply growth: this causes a 'flight from money' as above, which greatly adds to the inflationary pressure in the economy, driving prices to levels much higher than the rise in the money supply itself. The contrary holds when the hyperinflation is cured, say by a currency reform: then deflationary pressure sets in as people rush back into holding money, so cutting demand for goods — this implies that a currency reform should (oddly, it might seem) make liberal amounts of the new currency available in order to stave off a sharp deflation. Figure 1.14.

Notes on the dynamic properties

To understand the nature of the dynamic path we can conduct a policy experiment — a simple one-off money supply shock, for example — and examine how it affects the economy period by period, simply using the model's equations in a diagrammatic way. So in period 1 the *LM* curve shifts rightwards, triggering higher inflation; this raises expected inflation for the next period which shifts the *IS* curve upwards, giving us the period 1 effect marked as 1 in Figure 1.15. In period 2, the Phillips

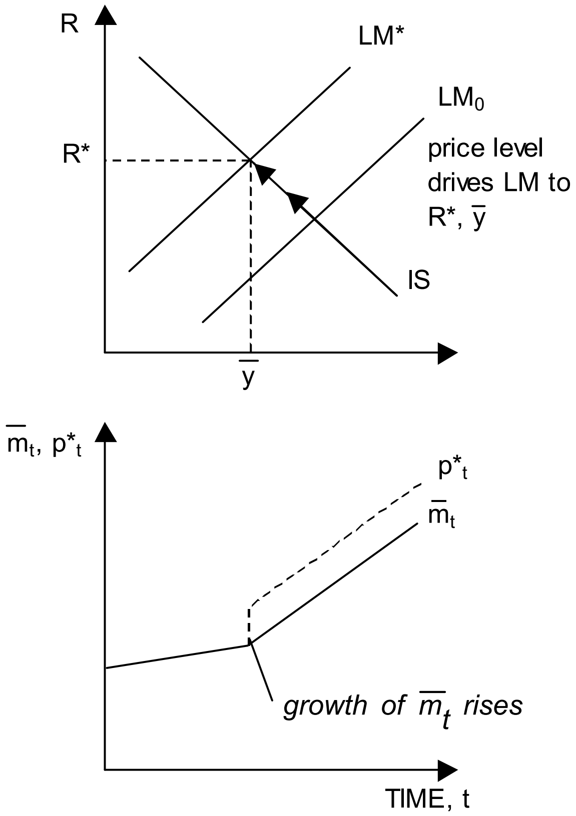

Figure 1.14: Price overshooting

curve shifts up with higher expected inflation; by now the LM curve is shifting back to the left as inflation exceeds the growth of money supply. The IS curve also continues shifting upwards with still higher expected inflation — giving us point 2 in Figure 1.15. It is easy to continue into further periods and show that the effects circle as shown. (This particular model in fact requires most unusual parameter values to behave in any other, non-circular, fashion, though it can do so.)

Figure 1.15 also illustrates that all the variables share the same motion. This follows from the simultaneity of the equations in the model — because all the variables react on one another they must obviously all move consistently, rather like the waves in a sea must all have the same sort of motion since they are all connected to one another. To discover

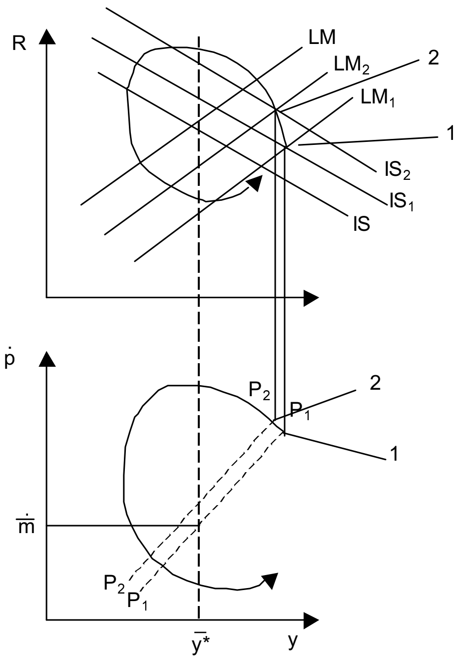

Figure 1.15: One-off rise in $\bar{m}_t$ — dynamics

the model's 'equation of motion' or 'characteristic equation', we follow a series of steps shown in the next box. What we are doing there is to reduce the model — by simple substitution — to a single equation in one variable (as deviations from its equilibrium value) and its lagged values: a difference equation (in this case 'homogeneous' because it equals zero, all the exogenous variables having been eliminated by expressing the endogenous variables in deviations from equilibrium). We can then readily show (see Box 1.2) that all the other variables have the same difference equation as the variable we have chosen.

Having obtained this equation we can then examine its behaviour, using the standard analysis of difference equations — see Chiang (1984) for example.

Comments on the model and its policy implications

The type of model we have been examining was the work-horse of macroeconomic modelling in the 1970s (clearly the versions used were often much more complicated, especially with the introduction of lags in the relationships). It has 'monetarist' properties as we have seen, in that inflation depends on money supply growth, and output equals its 'potential' or 'natural' level, in the long run. But it was also used by Keynesians, such as James Tobin, who stressed its short-run properties — such as taking a long time to get to long-run equilibrium and being prone to generate large business cycles. In a number of interchanges during the 1970s Keynesians and monetarists agreed to disagree about the likely size of different parameters in such a model and on appropriate policies but acknowledged that on the nature of the model itself they were essentially in agreement.

By the end of the 1970s the model was under serious challenge. At least as used by most forecasters it had failed to predict the serious inflation of the 1970s (associated with the tripling of the oil price in 1973) or the recession that went with it. The heart of the objection was the mechanism of adaptive expectations which gives the model its circular behaviour (as can readily be checked by reworking the model on the assumption that, for example, inflation expectations are determined by the growth of the money supply); the new 'rational expectations' school (discussed next) argued that people could do better. By 'adapting' they regularly made easily-preventable forecast errors: for example, if there were a one-off rise in the money supply of the sort we examined above, they would first underpredict, then overpredict, inflation and then repeat this cycle of errors indefinitely as the economy circled around its equilibrium inflation which of course remains precisely unchanged. Such expectations behaviour was unintelligent, irrational — hence unlikely — and a poor basis for predicting the behaviour of the economy, sensitive as this is to expectations.

Box 1.2

DYNAMIC PROPERTIES OF THE NEO-CLASSICAL/KEYNESIAN SYNTHESIS MODEL — REDUCING THE IS-LM-PHILLIPS CURVE MODEL TO CHARACTERISTIC EQUATION

1. Subtract equilibrium values from each equation, to write model in deviations from equilibrium e.g. IS curve:

$$\delta y_t = -\alpha r_t + \overline{d}$$

$$-y^* \quad = \quad +\alpha r^* - \overline{d}$$

implying $y_t - y^* = -\alpha(r_t \ - r^*)$

2. Model in deviation form becomes (all variables in deviations from equilibrium):

$$\begin{aligned} y_t &= -\alpha r_t \text{ (IS)} && \text{(a)} \\ 0 &= p_t + \gamma y_t - \beta R_t \text{ (LM)} && \text{(b)} \\ \dot{p}_t &= \dot{p}_t^e + \delta y_t \text{ (Phillips Curve)} && \text{(c)} \\ \dot{p}_t^e &= \lambda \dot{p}_{t-1} + (1-\lambda)\dot{p}_{t-1}^e \text{ (Adaptive expectations)} && \text{(d)} \\ R_t &= r_t + \dot{p}_{t+1}^e && \text{(e)} \end{aligned}$$

3. Derive Aggregate Demand function from (a), (b), (e). From (b) and (e).

$$r_t = +\frac{1}{\beta}p_t + \frac{\gamma}{\beta}y_t - \dot{p}_{t+1}^e$$

Substitute into (a) and collect terms, implying:

$$y_t = -\frac{\alpha}{\beta + \alpha\gamma}p_t + \frac{\alpha\beta}{\beta + \alpha\gamma}\dot{p}_{t+1}^e \text{ (AD curve)}$$

Write (c) as AS curve $y_t = \frac{1}{\delta}(\dot{p}_t - \dot{p}_t^e)$ and equate to AD

$$\Longrightarrow \frac{1}{\delta}(\dot{p}_t - \dot{p}_t^e) = -\frac{\alpha}{\beta + \alpha\gamma}p_t + \frac{\alpha\beta}{\beta + \alpha\gamma}\dot{p}_{t+1}^e$$

Now (Koyck transformation) subtract $(1-\lambda) \times$ lagged value from both sides. (Equivalently, use the lag operator, L, where $Lx_t = x_{t-1}$. So L is a symbol instructing us to lag a variable; it can be treated algebraically in the same way as a coefficient.
Rewrite $\dot{p}_t^e = \lambda L\dot{p}_t + (1-\lambda)L\dot{p}_t^e$ or $\dot{p}_t^e = \frac{\lambda L}{1-(1-\lambda)L}\dot{p}_t$. Now substitute this expression in the above equation and multiply through by $1 - (1-\lambda)L$.)
Through (d) this yields

$$\frac{1}{\delta}(\dot{p}_t - \dot{p}_{t-1}) = -\frac{\alpha}{\beta + \alpha\gamma}(p_t - [1-\lambda]p_{t-1}) + \frac{\alpha\beta}{\beta + \alpha\gamma}(\lambda\dot{p}_t)$$

Since $\dot{p}_t = p_t - p_{t-1}$, substitute all through for p_t, p_{t-1}, p_{t-2} in place of $\dot{p}_t$, $\dot{p}_{t-1}$ and collect terms to get:

$$p_t + ap_t + bp_{t-2} = 0$$

where

$$\begin{aligned} a &= \frac{-(1-\lambda+2s-r)}{1+s+r}; \\ b &= \frac{s}{1+s-r}; \\ s &= \frac{\beta+\alpha\gamma}{\alpha\delta}; \\ r &= \beta\lambda \end{aligned}$$

Notes on difference equations

Solutions to second order equation $x_t + ax_{t-1} + bx_{t-2} = 0$ is

$$x_t = A\lambda_1^t + B\lambda_2^t$$

where λ_1 and λ_2 are solutions of quadratic

$$\lambda^2 + a\lambda + b = 0, \lambda_1, \lambda_2 = \frac{-a \pm \sqrt{a^2 - 4b}}{2}$$

In case of complex roots (where $4b > a^2$), the solution can conveniently be written:

$$x_t = A\lambda^t \cos(\theta t - \varepsilon); \lambda = \sqrt{b}; \cos\theta = -\frac{a}{2\lambda}$$

The length of the cycle is $\frac{360^\circ}{\theta}$.

The sufficient and necessary conditions for stability are:

$$\begin{aligned} 1+a+b &> 0 \\ 1-a+b &> 0 \\ 1-b &> 0 \end{aligned}$$

Against this critique, the model's users claimed that adaptive expectations were a reasonable approximation when, as was usually the case, people had poor knowledge either of the workings of the economy or the processes driving policy or both — a neat demonstration of this point was made by Benjamin Friedman (1979), no relative of Milton Friedman. Then adaptive expectations is close to an optimal learning procedure. True, if there were to be just one shock such as a money supply rise, then the model would break down as people would soon realise what had happened; but in practice a lot of shocks are continuously hitting the economy and so people cannot easily work out what is going on: the model would be a good approximation then.

If we accept for purposes of argument that the model does make sense in some such terms, what policy implications can we draw from it? We can identify three main issues:

1. Rules versus discretion in monetary policy.

Monetarists argued that the money supply should be kept to a steady growth rate as far as possible (rules) rather than being varied in response to the perceived state of the economy (discretion). The argument was based on the difficulty of knowing the parameters, or even the economy's actual state, accurately — the slogan used by Friedman was that 'the lags were long and variable'.

If one reacted excessively to the economy out of such ignorance, then the cycles one was trying to dampen would get worse, possibly much worse. One can examine this point by adding into the model a money supply response function, in place of the exogenous money supply we have assumed; this will change the dynamic properties of the model and depending on how well it is chosen, the model can become more or less cyclical. Monetarists argued that the prudent (risk-averse) government would prefer to avoid the chance of making things much worse to enjoying the possibility of making them much better — and so would opt for a rule.

Against this Keynesians argued that such fears were exaggerated — we did know enough from our improving econometrics to steer the economy along a less boom–bust course.

Clearly this argument was empirical; it could only be settled by seeing how well or badly discretionary policy did. The experience of the 1970s, when there was a strong boom in the early 1070s, followed by the oil crisis bust, then again followed by a sharp expansion leading to bad inflation — did not inspire confidence in the highly discretionary policies pursued in that era.

2. Gradualism

Friedman argued that when far away from the appropriate rule, policy should be shifted back in small repeated steps because a large shift would be highly destabilising for the economy. For example, faced with high inflation due to past excessive money supply growth, a government that suddenly cut money supply growth back to a much lower rate would precipitate first a huge recession, followed eventually by a large boom. This sequence would lead to the rejection of the policy. If, however, money supply growth is slowly cut, each step would induce a small recession but after a few steps this would come on top of revival from the first steps; while the economy would be making steady long-run progress back to low inflation, the side-effects in lost output and unemployment would be more tolerable politically.

This argument is obviously a strong one if this model is correct. However. it is not clear that even gradual steps would be easy to undertake politically because the length of time an economy would be suffering from (mild) recession would be very long. Since bad inflations have frequently been ended, this raises questions about the appropriateness of the model in examining the curing of inflation: the rational expectations school have argued that it is too pessimistic because governments can express the national willingness to cure inflation and this message can be understood by people in the economy — hence their inflation expectations will adjust rapidly downwards, and much less recession is required to reduce inflation therefore.

3. Indexation

Friedman argued that contracts (especially for wages — but also for government bonds) should be indexed to prices, for two reasons. First, this would reduce the economic damage in markets; there is a loss of consumer surplus due to taxes whose true rate is often raised by inflation — indexing the objects of taxation, levying the tax on the indexed rather than the nominal amount, would effectively lower the tax. Second, it would stabilise output around its natural or equilibrium level by speeding up the response of inflation to deviations from this level. In other words the Phillips Curve would shift more rapidly upwards in response to excess demand and drop more rapidly in response to excess supply. Friedman was most concerned in practice about curing inflation and so about the response under excess supply (brought about by cutting money supply growth). He saw indexation as a way to speed up the cutting of inflation without slump.

We can illustrate the point under the extreme case where indexation of wages is immediate and prices are a simple margin mark-up on wages adjusted for productivity: in this case the Phillips curve will become

$$\dot{p}_t = \dot{p}_t + q(y_t - \overline{y})$$

In other words expected inflation disappears from the equation (because wage bargainers do not need to forecast inflation any more), entirely replaced by actual inflation under indexation. It is easy to see that this gives us a vertical Phillips Curve — any excess supply raises inflation which then raises wages and so inflation ad infinitum. In this case, there is no need for gradualism; to cure high inflation, the government merely needs to cut money supply growth at once to its target inflation rate, and virtually overnight inflation will drop to it with a minimal cost in recession.

This seems an attractive argument; and yet of course indexation cannot in practice be so immediate. If so, it generates instead a Phillips Curve with lagged inflation on the right-hand side:

$$\dot{p}_t = \dot{p}_{t-1} + q(y_t - \overline{y})$$

for example, where the lag is one period. But as this example readily shows this may worsen the instability of the economy. In this particular case, the one-period lag is equivalent to adaptive expectations with an adaptive parameter of unity; this, as we saw in Box 1.2 above, may produce unstable behaviour. The cycle is faster but may be explosive.

Experience with indexation in the 1970s did indeed reveal that it increased the violence of the economy's responses to shocks — increasing inflation in the upswing and lowering it in the downswing. For example, the Heath government of the UK in 1970–74 indexed wages almost universally and presided as a result over an inflation explosion in response to a programme of policy stimulation. It was not clear from this experience that it was an advantage in counter-inflationary policy to have indexation in place. During the late 1980s when counter-inflationary policy became general in the Organization for Economic Cooperation and Development (OECD) several countries — including Italy and Belgium — abandoned indexation at the same time.

We now turn from this model to discuss rational expectations, the hypothesis which spearheaded the challenge to this way of understanding the economy.

Rational expectations: Introductory Ideas

Expectations are fundamental in economics. Every economic decision is about the future, given the existing situation. Even decisions apparently only involving the current situation such as choosing to buy apples rather than pears, implicitly involve a view of the future; in this case it is a view both about future income which underlies the budget one is prepared to allocate to apples and pears, and about future prices of apples and pears, which dictates whether it is worthwhile storing them in place of future purchases.

By definition the future is unknowable. Economics, the study of economic decision making, therefore is concerned with how people deal with the unknowable.

The Liverpool University economist, George Shackle, made this a constant theme of his work. In his view (fo example, Shackle, 1958) each individual constructs in his imagination different scenarios or possible outcomes following on different actions he is contemplating. He knows now the pleasure given by the consequences he imagines for each course of action. The course which gives him the most pleasure when its consequences are imagined spurs him to action. So Shackle envisages a hedonistic calculus of the imagination as the basis for economic choice.

For Shackle the pleasure involved in a course of action will depend not only on the 'expected pleasure' if the consequences anticipated were to occur, but also on the 'chances' of them occurring. A fine outcome will normally give less pleasure when imagined if it is fraught with uncertainty. But this will not always be true. For example, for some people the same outcome may seem the sweeter if it is accompanied by the challenge of possible failure. Too easy a victory may not inspire a sense of triumph at all and may not attract the decision taker.

Shackle therefore regards economic decisions as entirely subjective, because they are the product of the individual's imagination interacting with the known facts of the past and present.

Nor does he feel that we can predict how people will act except in matters where nature reproduces itself with regularity. For example, in inventory control where the product has a reliable distribution, the manager can and will stop production when a sample has a proportion of defective products higher than some percentage. In such cases Shackle says there is 'uncertainty'; the rules of probability can be applied by people and we can infer what they will do from a knowledge of their objectives and of the probability distributions.

For many decisions the major elements in the future outcome are not subject to reliable distributions. For example, what the voters of Warrington would do in July 1981, faced for the first time in Britain with a Social Democratic candidate, could not be described by a probability distribution. It was not a regular event. No sense could be given to 'the' probability distribution. Such elements Shackle terms sources of 'potential surprise'. In evaluating future outcomes with such elements, each person will make his individual assessment and we cannot say what he will do without a complete knowledge of his psychology, even if we have access to exactly the same facts as he has.

Shackle's perceptions appear to be quite devastating to the use of statistical models in economics (that is econometrics). For these models assert that economic behaviour is predictable and regular. Their implication is that economic agents take decisions about future outcomes which are predictable and subject to well-defined probability distributions. The models are supposed to be derived from assumptions about agents' preferences, technology and the probability distributions they face.

The keystone of econometrics is the implicit assertion that, in the mass, individual decisions exhibit regularity, even though each individual decision will be quite unpredictable. Suppose we could define the probability distribution of future outcomes. Then econometrics asserts that there is a 'typical' individual who, faced with this distribution, will decide in a certain way; we might in particular say that he 'maximizes an expected utility function', where the utility function is a mathematical representation of his preferences. Individuals, when aggregated in a large sample, will behave like many typical individuals.

This assertion is supported up to a point by the central limit theorem, according to which the distribution of the mean of a sample of n random variables each of variance σ^2 will have a variance equal to $\frac{\sigma^2}{n}$ and as the sample n gets larger will tend to normality regardless of the distributions of each variable in the sample (see, for example, Hoel, 1962, p. 145). But although the assertion is supported in this way, it is still required that an individual decision can be regarded as a random variable with a defined mean, generated by a systematic decision process. Shackle denies this, while econometrics asserts it. Furthermore, it was supposed in the above argument that the future outcomes had a probability distribution. But for this to be the case it is necessary that individuals do behave in the aggregate in the regular manner just described. For the distribution of future outcomes is the product of the interaction of individuals' behaviour — for example, the distribution of future prices and quantities for sugar results from the interaction of individuals' supplies and demands for sugar.

Thus the linchpin of the whole edifice of econometrics is the postulate of regularity in economic nature. That postulate in turn justifies the assumption that econometricians face regularities. Whether this postulate is 'true or false' can only be settled empirically, by evaluating the success and failure of econometrics in attempting to apply this basic assertion. But many economists regard it (as did Keynes — see, for example, Keynes, 1939) as an unfortunate development which has perverted policy on many important issues because of a false perception that numerical estimates of likely outcomes from different policies can be generated.

Some of the early hopes of econometricians were that econometrics would enable complex problems of decision making by governments, firms and individuals to be reduced to the mere application of known techniques to models of the relevant economic environment. But these hopes have been cruelly dashed. Nowhere has this been more so than in macroeconomics. Macroeconomic models of the Keynesian (and later neo-Keynesian synthesis) type were built in great profusion in the 1960s for most major countries and used, especially in Britain and the United States, as the basis for forecasting and policy. Crude forms of 'optimal control' were used by governments: typically governments made forecasts, using their models, of inflation and unemployment over some horizon and then varied policies to obtain better outcomes according to these models. By the mid-1970s disillusion with these methods was widespread, as the Western world grappled with a combination of high inflation and high unemployment which, in general, neither had been predicted by econometric models, nor had responded to policy in the manner predicted by these models.

The reaction to this disillusion has been varied, with the result that the consensus in macroeconomics that had appeared in the late 1960s had abruptly disappeared by the late 1970s, as we have seen in the previous section. To some this was a vindication of their scepticism about econometrics. Others, however, have searched for reformulations of the models which failed in the 1970s. Rational expectations has been one major result.

Rational expectations takes one step further the basic assertion of econometrics that individuals in the aggregate act in a regular manner as if each was a typical individual following a systematic decision process. The step is to assert that in this decision process the individual utilizes efficiently the information available to him in forming expectations about future outcomes.

By 'efficient utilization' is meant that the typical individual's perception of the probability distribution of future outcomes (his 'subjective distribution'), conditional on the available information, coincides with the actual probability distribution conditional on that information.

It cannot be stressed too heavily — since this is a source of repeated misunderstanding — that this is an assertion about the 'typical' individual; in other words, it is to be used operationally with reference to the aggregate behaviour of individuals. It cannot be falsified by examples of behaviour by any actual individual. Clearly particular individuals may behave in ways that exhibit no systematic rational behaviour or expectations formation according to normal criteria, and yet if the behaviour of other individuals, who contributed a dominant proportion of the variability in aggregate behaviour, were rational in this sense, this would be sufficient to generate aggregate behaviour that exhibited rationality.

A further point is that, like all assertions in econometrics, it is to be judged by statistical criteria which will be relative. Whether a relationship asserted to operate for the behaviour of a group 'exists' or not will depend upon how reliable it is statistically, relative to the uses to which it may be put. Such uses in turn may themselves evolve from the discovery of new relationships.

There is therefore no such thing as an objective criterion for judging whether an econometric assertion is valid. Rather there is a joint and interactive evolution of models and their uses. The 1970s crisis in macroeconomic policy and modelling described earlier arose because the uses to which models were being put required properties that those models turned out not to have. Therefore the question that the rational expectations research programme has to answer is: can models based on rational expectations be developed that are useful in application? If so, what are these uses, and how do they compare to the uses available for other modelling approaches?

The assertion (or 'hypothesis') of rational expectations in one sense vindicates Shackle's basic perception that expectations are at the centre of economic behaviour. For on this view behaviour reacts to expected future behaviour which in turn depends on current behaviour; the capacity therefore exists for changes in the environment to affect current behaviour sharply, as individuals react to their perceptions of the changed environment and its implications for the future. Expectations are therefore completely integrated into behaviour.

In previous theories of expectations, this integration was incomplete. Econometric models used 'proxies' for expectations of particular types of outcomes. These proxies were based on what had been found in the past to correlate well with behaviour in a way that could reasonably be attributed to the operation of expectations rather than actual events. Hence for any particular problem in hand a proxy would be sought *ad hoc* — the most satisfactory formulation on which we have dwelt was adaptive expectations. However it is obvious enough that if changes in the environment disturbed the previous relationship between the proxy's behaviour and expectations, this would affect behaviour in a way that would be inaccurately captured by the effect of changes in the proxy. Hence only under the restricted set of circumstances where the previous relationship is unaffected will the proxies be useful. Unfortunately, the restricted set excludes most of the policy experiments that are of interest to governments.

In another sense the rational expectations hypothesis conflicts with Shackle's vision in that it is an attempt to use econometrics to capture the integration of expectations into behaviour. As such, it could be rated as even more foolhardy than basic econometrics. For whereas in well-settled times, patterns of behaviour might conceivably evolve which could give rise to some econometric relationships, in times of change when expectations are being disturbed in an unfamiliar way, then surely to model the effect of the changes on expectations and on their interactions with behaviour must be a mad attempt. For there is the capacity for immense diversity of imagined outcomes from the changes taking place; the 'typically' imagined outcome will be a useless construct in this diversity, where different individuals will be behaving in unrelated and possibly conflicting ways.

This reaction correctly identifies the way in which rational expectations is a programme for pushing econometrics to the limits of its possibilities in the prediction of behaviour. At the same time the reaction correctly notes the ambition of this attempt. It is possible that the attempt is hopelessly over ambitious. If it is proved to be so, then at least the limits of econometrics will have been clearly defined. For it will be the case that econometric relationships can at best only be useful in the restricted circumstances where the environment shows considerable stability. This would imply that they could be used to forecast existing trends in an unchanged environment, but not to predict the effects of changes in the environment, especially policy changes. If this were so, then it would equally invalidate attempts to use econometric models to design 'optimal' policy rules. Hence the implications flowing from a proper exploration of the potentialities of rational expectations are substantial and important.

We wrote in our first edition: 'While our own view is that these potentialities are considerable, it will only be a decade or so before we will know with reasonable clarity just what can be delivered from this approach.' The view of the economics profession today, judging from the widespread application of rational expectations, is that much has been delivered from it that is useful; the hypothesis has evolved its own set of uses as well as revealing the limits of some traditional ones.

The early intellectual History of the Hypothesis — an Overview

As so often in economics, it turns out that early economists propounded ideas at different stages that bear a striking resemblance to the hypothesis. For example, Marshall (1887), in his evidence to a Royal Commission, argued that an increase in the supply of money would affect economic activity by lowering interest rates, increasing loans, expenditure and finally prices. However, he also added that if the increase in the supply of money was well known, then individuals would anticipate the consequent expansion in demand and the effect on prices would be much faster[1].

[1] This reference was pointed out to us by Richard Harrington.

Modigliani and Grunberg (1954) are credited by Robert Shiller (1978) with the earliest post-war promulgation of the ideas behind the hypothesis. However, it was one of Modigliani's collaborators, John Muth, who truly created the hypothesis (Muth, 1961) in the sense that he set it down in precise form and showed how it could be applied to specific problems.

Muth's article was written partly in order to defend the prevailing flagship of the 1960s in expectations modelling, adaptive expectations, according to which expectations of a variable are an exponentially weighted average of past values of this variable. It turns out that under certain circumstances this is the same as the rational expectation (see the appendix to chapter 2). These were the circumstances to which Muth seemed to draw attention. His other work published at the time was exclusively devoted to the use of exponentially weighted averages (Muth, 1960; Holt et al., 1960).

The use of a particular modelling technique depends as much on its perceived tractability by economists as on its inherent plausibility. It was in fact the best part of a decade before economists started to use Muth's concept in its own right in applied and theoretical work. Adaptive expectations seemed in the 1960s the best tool, partly perhaps because it was still new and relatively unexplored (it made its first journal appearance in Nerlove, 1958), partly because of its most convenient econometric transformation (due to Koyck, 1954) into a single lagged value of the dependent variable. By contrast, the techniques for using rational expectations (RE) were not widely available; the solution of models with RE presented difficulties, overcome by Muth in very simple models, but not readily dealt with in the larger models being developed in that decade. As we shall see, solution of larger models (and also estimation) required substantial computing power; and it may well be that the rapid quickening of interest in RE modelling from the mid-1970s has been due to the explosion in the capacity of the electronic computer.

The earliest published work in macroeconomics using Muth's concept seriously, if only to a limited extent, is that of Walters (1971) and Lucas (1972a). Walters showed that the effect of money on prices would be substantially quickened by RE. Lucas argued that under RE (unlike adaptive expectations) monetary expansion could not raise output permanently above the natural rate of average (the 'natural rate hypothesis'), although the responses of money to lagged output and prices (feedback responses) could affect the time-path of output in response to shocks. Note that neither author in these articles argued for the ineffectiveness of monetary policy feedback responses for influencing the time-path of output.

Another early paper was by Black (1973), who applied rational expectations to the foreign exchange market, an area which has seen many further and productive applications of the concept. Some of this subsequent work used Muth's original model in a partial equilibrium treatment of the market, treating macroeconomic variables exogenously (for example, Minford, 1978; Bell and Beenstock, 1980). Other work (for example, Frankel, 1979) has estimated reduced-form models of the exchange rate derived from monetary models with rational expectations. Shiller (1973) applied it to the term structure of interest rates; here too a separate but closely related body of work has been extremely fruitful.

This has taken the form of tests of the 'efficient market hypothesis'. This is a combination of the rational expectations hypothesis and some hypothesis about market behaviour, which thus makes it possible to test rational expectations through the behaviour of market prices. Eugene Fama (for example, 1970, 1976) and his collaborators at Chicago were prolific pioneers in this area, and covered not only financial markets but also a variety of commodity markets, with results which at the least substantially revised the popular notions of the early post-war period that markets were irrational and highly inefficient.

Work on general macroeconomic applications (that is to inflation and output) in the early 1970s was substantially that of Lucas (1972a, b) and Sargent and Wallace (1975). Lucas' concern was to develop the rationale for fluctuations in money supply to affect output; his problem being that if information on money is available, then movements in money should immediately be discounted in prices, as everyone seeks to maintain relative prices unchanged. He developed a theme due to Milton Friedman (1968) that individuals perceive economy-wide data such as money and the general price level with an information lag, and are forced to estimate whether the price changes they currently perceive at the market level are relative price changes or merely reflect general price movements (inflation). A positive relationship between output and money or price movements (a 'Phillips curve') can occur because of mistakes made by individuals in estimating current inflation; they supply more output mistakenly thinking that the relative price of their output has risen.

This 'surprise' supply function was an essential component of the small-scale macro models used by Sargent and Wallace to illustrate the potential implications for policy making of RE. They showed that in a 'Keynesian' model, if the Phillips curve (a standard Keynesian construct) is interpreted as a surprise supply function, then only monetary surprises can affect output; monetary plans, whatever they are, can only affect prices. Hence, in particular, sophisticated feedback rules (which, for example, raised money supply in response to a poor past output performance) would be ineffective in their output stabilization objective. However, Lucas and Sargent (1978) have pointed out that this work was not intended to imply that monetary policy was in general ineffective for output, merely that in a particular 'standard' model (which they did not in any case endorse) to which 'optimal' monetary control techniques were routinely applied, such techniques were in fact useless. The lesson they drew was cautionary: monetary policy rules should carefully allow for the reactions of private agents to the rules themselves.

Lucas (1975) and Sargent (1976a) proceeded in the later 1970s to develop an alternative to the standard model, the so-called 'equilibrium business cycle' model. In this model the information lag story is maintained, but all markets are treated as if they are auction markets in which all agents are atomistic competitive units. Households are consumers and suppliers of labour and, period by period, compute optimal intertemporal plans for consumption and work based on the price sequences they perceive themselves as facing. Firms, on the other side of the labour and goods markets from households, similarly compute (given their technology) optimal plans for hiring labour and producing goods based on the same price sequences. The price sequences that are perceived are the rational expectations equilibrium sequences that clear the markets today and are expected to clear them in the future. Because firms and households, once they have made a plan, incur 'sunk costs', there is an adjustment cost in changing plans as these past decisions are unwound. This imparts the correlation of prices and quantities over time that is a feature of business cycles. But the impulse to the cycle comes from shocks to the economy, whether from monetary policy, technological innovations or surprise shifts in household preferences.

Such a model is by no means mandatory in the rational expectations research programme. The 'surprise' supply function is hard to sustain when information about prices is as up to date as it is in the modern economy. Also, while the 'as if' assumption of auction markets and automatic agents may work well for economies with highly competitive structures, monopoly power, in labour markets particularly, is widespread in Western economies, markedly so in Europe, and could require explicit modelling. Furthermore, the role of long term contracts — in goods, labour and financial markets — may be inadequately captured by this 'as if' assumption; this assumption would imply that contracts were approximately 'fully contingent' (that is, such that prices and quantities altered in response to shocks in just the manner that optimal plans would call for if there were no contracts), yet this apparently is not generally the case, for reasons that are still not well understood but may well be entirely consistent with rationality. Later, 'menu costs' (the cost of changing prices, as on a printed menu) were introduced to justify the setting of prices in nominal contracts (Parkin, 1986; Rotemberg, 1983).

Work by Phelps and Taylor (1977) and Fischer (1977 a,b) took non-contingent contracts in goods and labour markets as given and set up simple models to explore the implications. The influential model of an 'overshooting' foreign exchange market in an open economy, by Dornbusch (1976), centred on the interaction of 'sticky' contracts in goods and labour markets with an auction foreign exchange market. The effect of changes in financial asset values due to the presence of financial contracts (such as long-term bonds) similarly played an important role in the models of Blanchard (1981) and Minford (1980). Integration of labour monopoly power into these models has been carried out by Minford (l983). These developments implied a model different in many aspects from the narrowly defined equilibrium model of Lucas and Sargent, although it should be stressed that a model with contracts and monopolies where agents have rational expectations could perhaps most naturally be described as an equilibrium model, in that contracts have been voluntarily entered into and monopolies have agency rights on behalf of their members (and if the state legitimizes a closed shop, their non-members too). The policy properties differ substantially; there is in general more scope for stabilization, although whether it would be beneficial is another matter, requiring careful welfare analysis (see Chapter 5).

There were yet other strands in the ongoing RE research programme. B. Friedman (1979) and Brunner et al. (1980) attempted to introduce the modelling of learning about the model, notably about the evolving 'permanent' elements in the model's structure. Research on learning and whether it converges to rational expectations equilibrium continued to be active (Frydman and Phelps, 1983), although tractability was (and remains) a problem.

Long and Plosser (1983) suggested that business cycles may be modelled without an appeal to information lags or the price supply function; they may stem from real shocks to consumer wealth which generate equilibrium cycles in consumption, production and relative prices. Correlations between money and output may be explainable by implication in terms of reverse causation (King and Plosser, 1981).

This 'real business cycle' approach has the attraction that it relies on fewer *ad hoc* institutional restrictions than the New Classical approach (where information on money and prices, though in principle available rapidly, is slow to arrive in practice) or the New Keynesian (where contracts have unexplained nominal rigidities). Kydland and Prescott (1982) showed that a model of this type, calibrated to US data from cross-section studies, could replicate the time-series properties of the US post-war economy.

An influential school led by Sims (1980) despaired of modelling the economy's structure, but was interested in time-series analysis for predictive purposes within stable structures using the vector autoregression, or VAR (the regression of a group of variables on the past of all members of the group). McNees (1986) and Litterman (1986) found that such VAR models could predict as well as the main US commercial forecasters; however, in the UK during the 1980s, VAR models were outperformed by the main forecasters (Holden and Broomhead, 1990). Finally, a large and growing theoretical literature deriving from the work of Lucas (1978) and Wallace (1980) explored the behaviour of artificial economies peopled by representative agents. These models give us insights into the role of money, taxes and government bonds in enabling exchanges between groups that may be separated by geography (as in Townsend's, 1980, tribes travelling in opposite directions along a turnpike) or time (as in Samuelson's, 1958, overlapping generations models, for example, Wallace, 1980).

In the rest of this book we develop these themes in macroeconomics — in its modern guise macroeconomics is based entirely on the idea that agents are rational. Hence rational expectations is central to the subject today.