Chapter 10

The Money Supply in Macroeconomics

Peter Howells*
Bristol Business School

ABSTRACT

The notion that the quantity of money in an economy might be endogenously determined has a long history. Even so, it has never been part of mainstream economic thinking which has remained dominated by the view that the policymaker somehow controls the stock of money and that interest rates are market-determined. However, the need to design and operate a monetary policy that works for modern economies as they are currently constructed, has led to the emergence of the so-called ‘new consensus macroeconomics’ in which it is recognised that the policymaker sets a short-term interest rate and the quantities of money and credit are demand-determined.

This chapter looks at the way in which this ‘new consensus’ is (at last) forcing a recognition, in the teaching of money, that the money supply is endogenously determined. It also shows how we can take this further by adding a banking sector to a model of the real economy in which the money supply is endogenously determined.

1. INTRODUCTION

For many years, the role of money and monetary policy in macroeconomics has been represented by the IS/LM model, developed originally by Sir John Hicks (1937) to capture the essential ideas of Keynes's (1936) General Theory in a simple and tractable form. Its survival as the centrepiece of intermediate macroeconomics for so long is testimony to its versatility: it captures a very large number of simultaneous relationships in a very compact way. There are

* Professor of Monetary Economics, Centre for Global Finance, Bristol Business School sometimes overlooked in discussions of endogenous money and we shall see that it has resurfaced in recent work on the design of monetary policy rules. We conclude in section 6.
few aspects of macroeconomic policy that cannot be explored using the model. Unfortunately, the way in which central banks actually behave and the way in which monetary policy is transmitted to the rest of the economy are foremost amongst them. In the rest of this article we look at the way in which money is represented in the IS/LM model and why this fails to capture the current reality, in which the policymaker sets interest rates and the money supply is endogenously determined. We do this in the next section. In section 3 we look at why the money supply is endogenous in modern economies. In section 4 we review some recent attempts, related to what is often called the ‘new consensus macroeconomics’, to construct a model of monetary policy in macroeconomics which avoid the pitfalls and misrepresentations of the LM curve. In section 5 we look at an issue which is

2. MONEY IN THE IS/LM MODEL

In the IS/LM model, the LM curve traces combinations of the rate of interest and level of real income at which the money market is in equilibrium. This reference to market equilibrium implies independent supply and demand schedules. The supply side is the simpler of the two since the money supply is regarded as fixed by some external agent (the ‘policymaker’) and independent of the rate of interest.

In practice, the exogeneity of the money stock in the LM curve is rarely explained in macro textbooks. However, if we were to press for an explanation the chances are it would resemble the ‘base-multiplier’ model in which the central bank (independently or under government direction) sets the size of the monetary base and this in turn determines the stock of broad money as a multiple of the base.¹ Formally:

\[ Ms = Cp + Dp \]  

where \( Ms \) is the broad money stock and \( Cp \) and \( Dp \) are the non-bank private sector’s holdings of notes and coin and bank deposits respectively. Next:

\[ B = Cb + Db + Cp \]  

where \( B \) is the monetary base and \( Cb \) and \( Db \) are banks’ holdings of notes and coin and deposits with the central bank. If we combine \( Cb \) and \( Db \) and refer to them as bank ‘reserves’ (\( R \)), then we have:

\[ B = R + Cp \]  

and we can express the quantity of money as a multiple of the base:

\[ \frac{M}{B} = \frac{Cp + Dp}{R + Cp} \]  

¹ An interesting account of the origin and development of this model is given by Humphrey (1987)
If we now divide through by $Dp$ then we have:

$$\frac{M}{B} = \frac{\frac{Cp}{Dp} + \frac{Dp}{Dp}}{\frac{R}{Dp} + \frac{Cp}{Dp}}$$ \[5\]

Now let $Cp/Dp = \alpha$ and let $R/Dp = \beta$, then we can write:

$$\frac{M}{B} = \frac{\alpha + 1}{\beta + \alpha}$$ \[6\]

where $\alpha$ is the non-bank private sector’s ‘cash ratio’ and $\beta$ is the banks’ reserve ratio.

Finally, if we multiply both sides by the base, then we have

$$M = B \left( \frac{\alpha + 1}{\beta + \alpha} \right)$$ \[7\]

The here insight is that the broad money supply is a multiple of the monetary base and can change only at the discretion of the authorities since the base consists entirely of central bank liabilities. All of this is assuming that $\alpha$ and $\beta$ are fixed, or at least stable, and above all independent of the size of the base.\(^2\) We can now make this model explicit in the familiar diagram from which we derive the $LM$ curve:

Figure 1. Money market equilibrium.

\(^{2}\) In fact, many years ago, Paul Davidson (1988) introduced a distinction between ‘base-endogeneity’ and ‘interest endogeneity’. The latter arises as a result of $\alpha$ and $\beta$ varying inversely with interest rates. This creates a positive association between the rate of interest, the multiplier and hence the money supply (for a given size of base). The result is a positively-sloped money supply curve (and a flatter $LM$ schedule). The conventional meaning of an endogenous money supply, however, assumes endogeneity of the base as we see below.
The demand for money, however, is more complex in being related (positively) to the level of nominal income and (negatively) to a rate of interest. In Figure 1, we show such a demand curve drawn for each of three levels of income. For each level of income, there is a corresponding rate of interest \((Y_1/i_1; Y_2/i_2; Y_3/i_3)\), enabling us to draw an upward-sloping LM curve in interest-income space.

In the General Theory the interest rate link comes about because agents desire to avoid a capital loss (or benefit from a capital gain) as the rate of interest rises (or falls) and the current rate of interest functions as a guide, albeit a very uncertain one, as to what the next movement in interest rates is likely to be. In these circumstances, all that is needed in Figure 1 is a representative interest rate for which Keynes, reasonably enough, took the long bond rate. In more recent accounts, however, the interest link is often made through an opportunity cost argument. Here the demand for money is negatively related to the rate of return that can be earned on other assets. This poses greater problems when it comes to the choice of interest rate since (if money is non-interest bearing) we have to decide what is an appropriate alternative asset but, more seriously, when money does earn interest, as most deposits do these days, then ‘the’ interest rate has to be a spread term (e.g. bond minus deposit rate). But if money market equilibrium (and the resulting LM curve) require a spread term, it is hard to see how that same spread term can then be used to explain the behaviour captured by the IS curve when we bring IS and LM together.

But let us assume that money does not pay interest (a reasonable enough assumption in the 1930s). There remain major problems. For example, Hicks (1980) himself drew attention to the problems of combining a stock equilibrium (the LM curve) with a flow equilibrium (the IS curve) as well as the model’s contradictory demand for a real and nominal interest rate while Moggridge (1976) warned students that the model downplayed dramatically Keynes’s emphasis upon uncertainty – as regards the returns from capital spending and the demand for money – by incorporating them into apparently stable IS and LM functions respectively. And it gets worse when we focus on the LM curve itself. If interest rates are market-determined, what is the role of the Governing Council of the ECB (or the MPC at the Bank of England and the FOMC at the US Federal Reserve)? If the transmission of policy effects relies upon the quantity of money why do central banks make no mention of the money stock? If ‘loose’ monetary conditions lead to a fall in interest rates in the IS/LM model, why does the financial press predict a rise in interest rates when the consensus is that monetary policy is too slack? If stocks of money (and credit) can change only at the deliberate behest of the policymaker, why is the relentless growth of consumer debt a recurrent theme in the media? The shortcomings of the IS/LM model are often accepted as the price to pay for a useful teaching device, but these questions are regularly raised by enthusiastic but confused students who try to follow developments as reported in the media. And, as the fashion for policy transparency spreads amongst central banks with impressively informative websites, the student’s confusion can only increase.

The failure of the LM curve to allow a realistic discussion of monetary matters derives from the initial and fundamental assumption that the money supply is exogenously determined in the manner described above and shown in Figure 1. In fact, governments have never shown much interest in the money stock and certainly never in its absolute value. In 1967, when the UK government required a loan from the IMF, a condition of the loan required a restriction on the rate of ‘domestic credit expansion’ (roughly the loans that were the credit counterparts of bank deposits). Notice that the focus was on credit and its growth
rate. Furthermore, when it came to imposing restrictions the UK government relied upon ‘lending ceilings’ and not on any reduction in (the rate of growth of) the monetary base. When, in 1981-85, the first Thatcher government introduced the Medium Term Financial Strategy which included explicit money growth target, the policy instrument was the official rate of interest, intended to operate on the demand for bank loans. Even the Bundesbank, whose public stance on monetary policy involved frequent reference to monetary aggregates, used the management of short-term interest rates as the policy instrument (Clarida and Gertler, 1994; Geberding et al, 2005), a situation that continues under the ECB (Smant, 2002; ECB 2004).

In practice, policymakers set the rate of interest at which they supply liquidity to the banking system and, to maintain that rate of interest, reserves are supplied on demand. In effect, central banks are using their position as monopoly suppliers of liquidity to set the price rather than the quantity. And with the price set and maintained as a matter of policy, the quantity of reserves is demand-determined, determined by whatever banks need to support the deposits created by the demand for net new loans at prevailing interest rates. Two quotations, from different central banks (respectively the Bank of England and the US Federal Reserve), make the point clearly:

In the United Kingdom, money is endogenous - the Bank supplies base money on demand at its prevailing interest rate, and broad money is created by the banking system’. (King, 1994, p. 261)

And from much earlier:

…but in the real world banks extend credit, creating deposits in the process, and look for the reserves later’ (Holmes, 1969, p. 73)

A recent (and topical) illustration of just how important the interest rate is as a policy instrument (as opposed to the money stock) was also shown by a report in the Financial Times in the early days of the current crisis.

Central banks have been forced to inject massive doses of liquidity in excess of $100bn into overnight lending markets, in an effort to ensure that the interest rates they set are reflected in real-time borrowing....The Fed is protecting an interest rate of 5.25 per cent, the ECB a rate of 4 per cent and the BoJ an overnight target of 0.5 per cent. (FT 11/08/07, p. 3. Emphasis added)

Charles Goodhart, an economist who has spent his entire career working with and advising central banks, summarises the process like this (Goodhart, 2002):

- The central bank determines the short-term interest rate in the light of whatever reaction function it is following;
- The official rate determines interbank rates on which banks mark-up the cost of loans;
- At such rates, the private sector determines the volume of borrowing from the banking system;
Banks then adjust their relative interest rates and balance sheets to meet the credit demands;

- Step 4 determines the money stock and its components as well as the desired level of reserves;
- In order to sustain the level of interest rates, the central bank engages in repo deals to satisfy banks’ requirement for reserves.

And most significantly of all, the rate of interest as policy instrument and the consequent endogeneity of money lies at the heart of what is now called the ‘new consensus macroeconomics’.  

It is often supposed that the key to understanding the effects of monetary policy on inflation must always be the quantity theory of money... It may then be concluded that what matters about any monetary policy is the implied path of the money supply... From such a perspective, it might seem that a clearer understanding of the consequences of a central bank’s actions would be facilitated by an explicit focus on what evolution of the money supply the bank intends to bring about – that is by monetary targeting... The present study aims to show that the basic premise of such a criticism is incorrect. One of the primary goals ... of this book is the development of a theoretical framework in which the consequences of alternative interest-rate rules can be analyzed, which does not require that they first be translated into equivalent rules for the evolution of the money supply’. (Woodford, 2003, p. 48. Second emphasis added).

We look next at how we got to this position. Why do central banks set the price rather than the quantity of reserves?

### 3. Why Is the Money Supply Endogenous?

For the money supply to be endogenous, two conditions must be fulfilled. The first is that the causes of monetary expansion (or contraction) must lie with other variables within the economy, as opposed to being at the discretion of some external agency (‘the policymaker’). The second is that, in order to respond to these forces, commercial banks must be able to obtain reserves on demand, or be able to economise on their need for reserves. In either event, reserves must not be a constraint.  

As regards the former, the argument begins with an accounting identity and a behavioural observation. The former is that loans and deposits appear on opposite sides of banks’ balance sheets and thus, ignoring changes in bank capital, a change in loans must be matched by a change in deposits. The latter is that banks respond to demands from the non-bank private

---

3 See, for example, Charles Bean’s (2007) list of defining features of the NCM. Further references to the inability of the money aggregates to exert any independent influence on the economy can be found in Chada (2008), Goodhart (2007), Meyer (2001) and Woodford (2007a, 2007b).

4 Which of these applies in modern monetary regimes and to what extent has been the subject of much debate between so-called ‘structuralists’ (banks can innovate to economise on reserves) and ‘accommodationists’ (central banks will always supply reserve on demand). These two positions were originally identified and analysed by Pollin (1991). It seems reasonable to suppose that banks can adjust their need for reserves to some
sector for credit not a demand for deposits. Hence ‘loans create deposits’ rather than the other way round. As an alternative to the base-multiplier model, this focus on the credit counterparts of the money supply can be captured in a simple ‘flow of funds’ model. As with the earlier case we begin with a definition of broad money:

\[ Ms = Cp + Dp \]  \[8\]

In changes:

\[ \Delta Ms = \Delta Cp + \Delta Dp \]  \[9\]

Given the balance sheet identity, then it follows that the change in deposits must be matched by the change in loans which can be decomposed into lending to the private sector (\(\Delta BL_p\)) and to government (\(\Delta BL_g\)).

\[ \Delta Dp = \Delta \text{Loans} = \Delta BL_p + \Delta BL_g \]  \[10\]


\[ \Delta Ms = \Delta Cp + \Delta BL_p + \Delta BL_g \]  \[11\]

Until the present crisis, the UK government deficit has generally been ‘fully-funded’, that is by the sale of government bonds, rather than borrowing from banks. With \(\Delta BL_g = 0\), money growth is explained entirely by bank loans to the non-bank private sector. However, the flow of funds model has its origin in the 1970s when the UK faced very large public sector deficits whose financing posed a potential problem. The fear was ever-present that the government might fail to sell the required volume of bonds, forcing it into residual financing from the banking sector. For this reason, the model was usually presented in a way that spelled out the monetary implications of the public sector deficit. Let the annual deficit (a flow) be represented by \(PSNCR\), then:

\[ \Delta BL_g = PSNCR - \Delta Cp - \Delta Gp \pm \Delta ext \]  \[12\]

where \(\Delta Gp\) is the net sale of government bonds (‘gilts’) to the general public and \(\Delta ext\) is monetary effect of official transactions in foreign exchange by the central bank (and thus equal to zero in a floating exchange rate regime).


\[ \Delta Ms = \Delta Cp + \Delta BL_p + PSNCR - \Delta Cp - \Delta Gp \pm \Delta ext \]

This is then tidied up (the change in notes and coin cancel) and re-ordered to give the conventional presentation:
\[ \Delta M_s = PSNCR - \Delta Gp + \Delta ext \pm \Delta BLp \]  

Once we accept that ‘loans create deposits’ (and not the other way round) it is a fairly simple task to link the demand for credit to the state of the economy, or the ‘state of trade’, as it is commonly described. Assuming normal conditions in which real output is growing, then there will be a demand for net new loans to finance the increasing production and consumption, matched by a corresponding increase in deposits. If we add to this the common condition of positive inflation then there will be further demand for new loans since the demand for credit (like money) is a demand for real magnitudes.

Although the endogeneity of the money supply was recognised many years ago\(^5\) and had powerful supporters in the not so distant past (e.g. Kaldor 1970, 1982, 1985; Kaldor and Trevithick, 1981; Davidson and Weintraub, 1973) it was Basil Moore who did most to promote the cause of endogenous money as a challenge to the monetarist revival of the 1980s. His book, *Horizontalists and Verticalists* (1988) included a chapter in which he tested the hypothesis that it was firm’s demand for working capital which explained the growth of bank lending (and thus the expansion of deposits). This triggered further empirical work which was broadly supportive of the link between the growth of credit and industrial production (e.g. Moore, 1989; Cuthbertson and Slow, 1990; Palley, 1994; Hewitson, 1995).

This notion, that the growth of credit and money reflects changes in nominal output, is important when it comes to the analysis of the role of money in the macroeconomy. For many economists in the post-Keynesian tradition it reverses the causality of the Quantity Theory of Money. Instead of money causing inflation (if its growth rate exceeds the growth of real output), it is the change in nominal income that determines monetary growth. The money stock is no longer the ‘cause’ of anything interesting but merely the passive response to changes elsewhere in the economy. However, the innocence of money in this respect relies fundamentally on the link with production and there are two recent trends, at least in the UK, that raise questions about the uniqueness of this link. The first is that measures of total transactions in the UK economy show a steady and dramatic increase in total transactions relative to GDP between 1980 and 1998, and a slow increase since then. Many of these non-GDP transactions represent transactions between financial institutions as the UK financial sector grew faster than the rest of the economy. But they also include loans to households for the purchase of (largely secondhand) houses. The period in question covers two substantial property booms and one slump. The second is that bank lending to households increases much more rapidly over the period than lending to non-financial corporations with the result that both stocks and flows of bank lending have been dominated by the household sector since 1990. What all this means is that credit (and money) may expand for reasons which may not be closely related to economic activity.

The notion that some variable wider than production or GDP, say total transactions, may be driving loan expansion is in principle testable. In 2001 Caporale and Howells published a

\(^5\) e.g. Wicksell (1898), Schumpeter (1911). It is also of some interest that the exogeneity/endogeneity of money was an issue long before – during the so-called ‘Great Inflation’ in England between 1520 and 1640. Many contemporaries blamed it upon the arrival of gold from Spanish discoveries in the ‘New World’. But there were others who held that the inflation had ‘real’ causes (most commonly population growth) and that the import of precious metals (as well as debasement of the coinage) were endogenous responses. For a detailed discussion see Arestis and Howells (2002) and Mayhew (1995).
paper in which they investigated simultaneously the causal relationship between three variables: total transactions, loans and deposits. The method they used (see Yamamoto and ) also enabled them to explore any direct link between transactions and deposits, by-passing the loan creation process. The study focused solely on the UK, using quarterly data from 1970 to 1998. The findings confirmed again the loan → deposit link but were not strongly supportive of the view that total transactions (rather than GDP) ‘caused’ the loans. Transactions did though ‘cause’ deposits. What the tests also revealed is that there appeared to be some causal feedback from deposits to loans, which has to be interpreted as meaning that the willingness to hold deposits, i.e. the demand for money must also be playing some role here. This is an interesting result in the light of an issue which has just been re-discovered and which we return to in section 5 below.

The first condition for the endogeneity of the money supply, namely that the cause of change must lie within the economic system, is satisfied therefore by the notion that it is loans that cause deposits and that, at a given rate of interest, the demand for loans depends upon the current level of economic activity (somehow defined). But this leaves us with the question of why banks are not reserve-constrained in their response to the demand for credit. Why is it that central banks respond passively by supplying the reserves required to accommodate the behaviour of loans and deposits? There are several parts to the explanation and we can usefully divide them into two groups. The first consists of technical difficulties confronting a policymaker who wishes to manage the size of the monetary base within pre-determined quantitative limits; the second consists of undesirable consequences that would most likely follow if such management were indeed feasible.

The base multiplier model is summarised in equation 7 and we said at the time that a fundamental insight it appeared to offer was that the money supply could change only at the discretion of the authorities who would have complete control over the size of the base, since its components were all liabilities of the central bank. The implicit assumption here is that the central bank must know and be able to control its liabilities, much like a household or a firm. But matters may not be so simple. In most monetary regimes, the public sector banks with the central bank. In the course of a normal working day, there will be large spontaneously flows between the public and private sectors. A net flow from the government results in an increase in the bank deposits of the nonbank private sector matched by an increase in banks’ deposits at the central bank. In the notation of section 2, we have an increase in the base since $Db$ is a component (see equation [2]) while government deposits, $Dg$, are not. Recall also that banks’ reserve ratio, $R/Dp$, is a very small fraction. Adding $Db$ and $Dp$ in identical amounts to the numerator and denominator respectively, will lead to a noticeable increase in this ratio and thus to banks’ liquidity. The same will happen in reverse when the non-bank private sector makes net payments to the government. The point is that there will be inevitable fluctuations in central bank liabilities, caused by spontaneous transactions between the public and private sectors. The first step in ‘knowing’ and ‘controlling’ fluctuations in the base requires, therefore, precise predictions of these flows. For the Bank of England, the prediction errors can be seen in the open market ‘fine tuning’ operations that the Bank has to engage in order to offset the effects of what it calls ‘autonomous’ flows in sterling money markets. These operations are reported the Bank of England Quarterly Bulletin. These same autonomous factors are identified by the ECB (2004) and their fluctuating nature is described on pp.88-89.
The difficulties involved in anticipating these magnitudes is implicit throughout the ECB’s discussion of the various open market techniques available to it (2008, ch.3).

Set aside for the moment, the difficulties involved in knowing the path of the base where there are large spontaneous flows between the public and private sectors. Consider now the difficulties of controlling it. Control requires compensating transactions between the public and private sectors. So, for example, reducing the rate of increase in the size of the base requires net sales of government debt to the nonbank private sector. And since the policymaker is aiming at a precise quantity target for the base, this requires sale by auction in order to ensure the precise quantity of the sale. Such auctions would be difficult and costly to organise with the costs and difficulties increasing with the shortness of the period over which the reserve target had to be achieved. For example, a regime which allowed averaging over a month would be more feasible than one which required the target to be achieved at the close of each day. But even so the administrative costs of frequent auctions would be considerable.

The requirement for an auction method of bond issuance is just another way of saying that if the target is a quantity then the price must be market-determined. The price here, of course, is the rate of interest that banks will bid for reserves, effectively the overnight interbank rate. Given that banks’ requirements for reserves are inelastic, the fluctuation in short-term rates could be very severe indeed. Most central banks would find wild fluctuations in interest rates more disruptive than fluctuations in the size of the base. The evidence for this (apart from the fact that it is the choice that central banks universally make in practice) is that bond auctions are invariably accompanied by a minimum price stipulation. Even in the depths of the financial crisis in December 2007, when the Federal Reserve introduced its emergency Term Auction Facility in order to calm money markets, it set a minimum bid rate (see Taylor and Williams, 2009, p. 69). The authorities would rather limit the quantity sold than accept a rise in interest rates beyond a certain point.

By recognising that strict monetary base control would result in very volatile short-term interest rates, we have already acknowledged that the adverse side effects could be considerable. These would include a number of institutional changes. For example, the overdraft system whereby lenders agree a credit ceiling and then charge borrowers on a daily basis for only the fraction of the facility that is used, is widely regarded as a cheap and flexible method of providing short-term credit to firms. But it makes the extent of borrowing (and the resulting deposit creation) a discretionary variable in the hands of banks’ clients. Knowing that they might be reserve-constrained, it seems unlikely that banks would expose themselves to the risk that they could face a substantial surge in loan demand in a situation of reserve shortage.7

Another possibility in a base-targeting regime is that banks would build up holdings of ‘excess’ reserves in periods of feast in order to protect themselves in future periods of famine. In addition to reducing the authorities ability to impose a reserve shortage, operating with a generally higher reserve ratio than is essential to protect against liquidity risk amounts to a tax on bank intermediation. This tax is substantial if reserves pay no interest (as is the case with notes and coin). But even where deposits with the central bank do pay interest, it is invariably

---

7 The conventional wisdom in the UK is that about 60 per cent of overdraft facilities are in use at any one time, meaning that this source of credit could almost double at the discretion of borrowers. Consider now that a reserve shortage and the consequent restriction of other forms of credit would make it almost certain that the demand for overdrafts would surge, the risk faced by banks operating such a system are clear.
at a rate which is less than banks could earn on assets that they could hold if they were not carrying excess reserves.

In modern economies, the money supply is endogenously determined and now we know why. In the next section we turn to recent efforts to incorporate a realistic version of the monetary sector into a simple macroeconomic model.

4. MONEY IN A MORE REALISTIC MODEL

Attempts to develop a ‘macroeconomics without an LM curve’ are now various starting, implicitly, with Clarida et al (1999) and more explicitly with Romer (2000). More recently we have seen a new framework for the teaching of monetary economics developed by Bofinger, Mayer and Wollmershäuser [BMW] (2005) and by Carlin and Soskice [CS](2005) who have since incorporated it in an intermediate level textbook (2006).

The flavour of all these attempts is best understood by looking at Romer (2000) who basically took the IS part of the IS/LM model, and dispensed with the LM curve by simply treating ‘the’ rate of interest on the vertical axis as an exogenously-determined policy instrument. Subsequent developments are essentially refinements and extensions of this approach. What follows is based, largely, on what Carlin and Soskice call the IS/PC/MR model in their 2006 textbook. The C-S book is doubly interesting since it represents one of the first attempts to introduce a more realistic treatment of money into a mainstream textbook. This requires the treatment to provide not just a sensible framework for the discussion of money and policy but also to be consistent with the modelling of the external sector and economic growth and a wide range of topics covered later in the book. It is also interesting because it starts from a position which embraces more wholeheartedly the essence of the new consensus than, for example, Romer (2000) whose discussion of the policy (interest) rate still relies upon the central bank controlling the stock of narrow money with a view to setting this rate.

As the name of the model implies, it is based on three equations. The first is the familiar IS equation:

$$Y_{t+1} = A - \phi r_t$$  \[14\]

where $A$ is autonomous demand and $r_t$ is the real rate of interest, in the previous period.\(^8\)

The second is a conventional short-run Phillips curve:

$$\pi_{t+1} = \pi_t + \alpha(Y_{t+1} - \bar{Y}_{t+1})$$  \[15\]

wherein inflation in the next time period depends upon current inflation (the inertia is due to price stickiness rather than inaccurate expectations) and the pressure of aggregate demand.

\(^8\) Notice that the real rate of interest determines output with a one-period lag. Realistically, in the following equations we should introduce a further lag from output to inflation. However, we have omitted this lag for convenience of exposition.
We then require a third equation, a ‘monetary rule’ equation, which sets the interest rate $r_t$. This could take the form of a Taylor rule or it could be written more generally as the rate of interest that minimises a loss function of the kind:

$$L = (Y_{t+1} - Y_{t+1}^*)^2 + \lambda (\pi_{t+1} - \pi_T)^2 \quad \ldots$$  \[16\]

Note that with $\lambda=1$ the policymaker is giving equal weight to output and inflation gap losses and that the effect of the quadratic term is to make overshoots and undershoots equally objectionable.

Next, we substitute the Phillips curve \[15\] into the loss function \[16\] and differentiate with respect to $Y$:

$$\frac{\partial L}{\partial Y} = (Y_{t+1} - Y_{t+1}^*) + \alpha \lambda (\pi_t + \alpha (Y_{t+1} - Y_{t+1}^*) - \pi_T) = 0 \quad \ldots$$  \[17\]

Substituting the Phillips curve back into this equation gives:

$$(Y_{t+1} - Y_{t+1}^*) = -\lambda \alpha (\pi_{t+1} - \pi_T) \quad \ldots\ldots$$  \[18\]

This shows the equilibrium relationship between the level of output (chosen by the policymaker in the light of preferences and constraints) and the rate of inflation.

If we wish to see this in diagrammatic form, then the starting point is Figure 2. The policymaker is assumed to have an inflation target ($\pi_T$) of 2 per cent. Initially, the economy is in equilibrium at $A$, with inflation running at that level. Output is at its ‘natural’ level (on a long-run vertical Phillips curve) so there is no output gap to put positive (or negative) pressure on inflation. An inflation shock is introduced which moves the economy to...
At which inflation is 6 per cent. In order to return to target, the central bank raises the real interest rate\(^9\) and pushes output below its natural level and we move down the short-run Phillips curve (drawn for \(\pi_t = 6\)) to the point labelled \(F\). Notice that \(F\) is selected because the central bank is at a point tangential to the best available indifference curve at that combination of output and inflation. The indifference curves are shown by the dashed lines. The indifference curve represents the output/inflation trade-off (the degree of inflation aversion) for that particular central bank. (A more inflation averse central bank would have a different indifference map and would move the economy to a point on PC \((\pi_t = 6)\) to the left of \(F\)).\(^{10}\) As the inflation rate falls to 5 per cent, the short-run PC shifts down to \((\pi_t = 5)\). The central bank can then lower the real interest rate, allowing output to rise, so the economy moves to \(F'\) and by this process (described as following a monetary rule) the central bank steers the economy back to equilibrium at \(A\).

The next step is to introduce the \(IS\) curve and the real rate of interest. This is done in the upper part of figure 3. To begin with, the economy is in equilibrium, shown in both panels by the point \(A\). Notice that in the upper panel, this includes a real rate of interest identified as \(r_s\) (a ‘stabilising’ rate of interest which maintains a zero output gap). In the lower part, we then have a replay of figure 2. There is an inflation shock which takes the economy from equilibrium at \(A\) to a rate of inflation of 6 per cent (at \(B\)). In figure 2a, the central bank now raises the real rate of interest (to \(r'\)) which has the effect of moving us up the \(IS\) curve to \(C\) at which the level of output is reduced. (In the lower panel we move down the SRPC \(\pi_t = 6\) curve to a point, corresponding to \(F\) in figure 2, at which the reduction in demand pressure lowers inflation to 5 per cent). As inertia is overcome, contracts embrace 5 per cent and the Phillips curve shifts down to SRPC \((\pi_t = 5)\), the real rate is reduced allowing some expansion of output. We are now at point \(D\) on the \(IS\) curve (and at a point corresponding to \(F'\) in figure 2) but since we are still to the left of \(Y^*\), inflation continues to fall. For as long as we remain to the left of \(Y^*\), the Phillips curve will continue to shift (and the real rate of interest can be lowered further) until inflation comes back to target at 2 per cent.

The next step is to incorporate the banking sector. A summary of the system we are trying to model is provided by Goodhart (2002):

- The central bank determines the short-term interest rate in the light of whatever reaction function it is following;
- The official rate determines interbank rates on which banks mark-up the cost of loans;
- At such rates, the private sector determines the volume of borrowing from the banking system;
- Banks then adjust their relative interest rates and balance sheets to meet the credit demands;

\(^9\) Carlin and Soskice (p.84) make the same point as Romer, that the central bank strictly speaking sets the nominal interest rate but does so with a view to achieving a real interest rate. Since it reviews the setting of this rate at regular, short, intervals, and the behaviour of inflation is a major factor in the decision, it is reasonable to see it as setting a real rate.

\(^{10}\) The indifference curves in figure 1 are segments of a series of concentric rings centred on \(A\). If the central bank’s loss function gives equal weight to inflation and output (as in the loss function [16]), the rings will be perfect circles. If the central bank puts more weight on inflation, the rings will be ellipsoid (stretched) in the horizontal plane. Hence greater inflation aversion on the part of the central bank would create a tangent ‘further down’ the PC, to the left of \(F\).
- Step 4 determines the money stock and its components as well as the desired level of reserves;
- In order to sustain the level of interest rates, the central bank engages in repo deals to satisfy banks’ requirement for reserves.

Figure 4, based on Fontana (2003, 2006), Howells (2009) and Bain and Howells (2009), embraces these requirements in four quadrants.

In QI the central bank sets an official rate of interest, \( r_0 \).

\[ r_0 = r_0 \ldots \] [19]
This official rate determines the level of interbank rates on which banks determine their loan rates by a series of risk-related mark ups. We make two simplifications. The first is that interbank rates are conventionally related to the official rate so that the mark-ups are effectively mark-ups on the official rate. The second is that we can represent the range of mark-ups by a single, weighted average, rate. This is shown as \( m \).

\[
    r_L = r_0 + m .
\]

In QII banks supply whatever volume of new loans is demanded by creditworthy clients at the loan rate \( r_L \). Notice that the loan supply curve, \( L^S \), denotes flows, consistent with what we have said about the flow of funds being positive at the going rate of interest. This is further confirmed by the downward-sloping loan demand curve, \( L^D \), showing that the effect of a change in the official rate is to alter the rate of growth of money and credit. At \( r_0 \), loans are expanding at the demand-determined rate \( L_0 \).

\[
    L^S = L^D \quad \text{(21)}
\]

\[
    L^D = f(\Delta \ln P, \Delta \ln Y, r_L) .
\]

QIII represents the banks’ balance sheet constraint (so the \( L=D \) line passes through the origin at 45\(^\circ\)). In practice, of course, ‘deposits’ has to be understood to include the bank’s net worth while ‘loans’ includes holdings of money market investments, securities etc. At \( r_0 \) the growth of loans is creating deposits at the rate \( D_0 \).

\[
    L^S = L^D = L_0 = D_0 \quad \text{(23)}
\]

The DR line in QIV shows the demand for reserves. The angle to the deposits axis is determined by the reserve ratio. In most developed banking systems this angle will be very narrow, but we have exaggerated it for the purpose of clarity.

\[
    DR = \frac{R}{D} \quad D .
\]

In a system, like the UK, where reserve ratios are prudential rather than mandatory, the \( DR \) line will rotate with changes in banks’ desire for liquidity. Even in a mandatory system, the curve may rotate provided that we understand it to represent total (ie required + excess) reserves. Thus one of the model’s strengths is that can show changes in banks’ liquidity preferences either induced by changes in central bank operating procedures (as in the UK in April 2006), or as an autonomous response to changed market conditions (see section 5).

---

Finally, in QI again we see the central bank’s willingness to allow the expansion of reserves at whatever rate (here R0) is required by the banking system, given developments in QII-QIV.

\[ R_0 = \frac{R}{D} (D_0) \]  \hspace{1cm} [25]

\[ R_S = R_D. \]  \hspace{1cm} [26]

How do we combine this with the analysis of Carlin and Soskice (or BMW) in figure 3? The key lies in QI. Recall that the rate of interest in QIV is the official rate, \( r_0 \) (usually a repo rate). We have already agreed that \( r_0 \) can reasonably interpreted as a real rate of interest which is what is required by the IS curve. All that we have done in QI is add a mark-up, \( m \), in order to convert \( r_0 \) into a loan rate, \( r_L \). Since the IS curve represents an equilibrium between investment and saving, there should be no objection to showing changes in equilibrium output to be dependent upon changes in the loan rate. This is directly relevant to investment spending and while one may object that the rate paid to savers is different, this objection could be made to any single rate of interest on the vertical axis. We are bound at accept that any single rate is a proxy for a spread term. In figure 4, therefore, we show (in QI-QIV) a

---

12 As we noted above, it was a widespread criticism of the IS/LM model that while the behaviour summarised in the IS curve required a real rate, the relationships in the LM curve depended upon a nominal rate.

13 Although the LM curve was traditionally drawn for a single rate of interest (usually the bond rate), this was strictly correct only if money’s own rate was zero. Strictly, the rate should have been a spread term incorporating the rate on money and the rate on non-money substitutes.
banking system in flow equilibrium (loans and deposits are expanding at a rate which satisfies all agents at the current level of interest rates and banks can find the appropriate supply of reserves to support this expansion).

Constraints of space prevent us from detailed demonstrations of the way in which the model(s) in these six quadrants can be used to illustrate the conduct of monetary policy. But two examples may be possible. First of all, consider the case that we had in figures 2 and 3 where there is an inflation shock and the policymaker raises interest rates in order to steer the economy back to $\pi^T/Y^*$. In QIV, the official rate ($r_0$) is raised. With a constant mark-up, this raises the loan rate, $r_L$ in QI. Transferred to QV, this moves the economy up the IS curve and the sequence of events that we saw in figures 2 and 3 begins. If we return to the monetary sector, the rise in interest rates raises the cost of credit and reduces the flow demand for new loans, and so deposits grow more slowly, accompanied by a slower rate of growth in required reserves which the central bank accommodates. As the rate of inflation (in QVI) falls, the policymaker can reduce the rate of interest and the expansion of money and credit returns progressively to normal levels as the real economy converges on the policymaker’s $\pi/Y$ target. This sounds like a reasonable description of how the monetary sector and the real economy interact in normal circumstances in modern economies where the money supply is endogenous and the policymaker is targeting the rate of inflation but is mindful of output losses.

Furthermore, we can use the model to illustrate abnormal circumstances of the kind that we have experienced recently. In QI, for example, we can show the effect of an increase in banks’ mark-up over the policy rate. This corresponds to recent experience whereby banks becoming concerned about each other’s creditworthiness raise interbank rates, from which many other bank products are priced. The effect in the rest of the model is as if the policymaker had increased the official rate and we can follow through the deflationary effects. We can also show the recent reductions in policy rates by the ECB, The Fed and the Bank of England, as an attempt to hold the market rate, $r_L$, down to an appropriate level in the face of the increase in $m$. The fuller discussion in Howells (2009a and 2009b) shows how the model can be used to illustrate other aspects of the credit crunch.

5. THE DEMAND FOR ENDOGENOUS MONEY

At the beginning of section 3, we described the flow of funds model of money supply determination as being more helpful than the base-multiplier model in understanding the money supply process since it focused upon (a) flows and (b) the credit counterparts of the money supply. However, the model suppresses one, potentially, important issue. 14 This is the question: ‘what has become of the demand for money?’ Consider: the model explains changes in the money stock as the sum of the flow demand for net new loans. But the demand for loans originates with a subset of the non-bank public who have an income-expenditure deficit while the resulting deposits must be held by a wider population who are making a portfolio

14 By way of comparison, standard representations of the base-multiplier model suppress discussion of the determination of the key ratios $\alpha$ and $\beta$. These are complex portfolio decisions and failing to consider them as the outcomes of maximising behaviour on the part of the non-bank public and banks respectively makes the model profoundly ‘uneconomic’.
decision. We appear to have two decisions being made by (partially) different groups of agents and with clearly different motives. And yet, as if by magic, they must coincide. The answer, as pointed out by Cuthbertson some years ago is that ‘There is an implicit demand for money in the model, but only in equilibrium ... the FoF model delivers an implicit equilibrium demand for money function’. (Cuthbertson, 1985 p. 173. Emphasis in the original).

This debate (the missing demand for money) received a boost a few years later when Basil Moore (1988) published what was for many years the seminal work on endogenous money, *Horizontalists and Verticalists*. Moore’s position was quite simply that in regimes where the money supply is endogenous, there is no independent demand for money. Money will always be accepted, in whatever quantity, because of its special role as a means of payment. Hence, whatever deposits loans may generate, they will always be willingly held. This gave rise to a lively debate (Goodhart, 1989, 1991; Moore 1988b, 1991, 1997; Howells, 1995, 1997) in which Moore was accused of misunderstanding the demand for money (accepting money in exchange for goods and services is not the same as the portfolio decision to go on holding it) and of denying that agents have preferences about how they hold wealth. But Moore was not alone in thinking that in an endogenous money environment the demand for money was irrelevant. A similar argument had appeared in Kaldor and Trevithick (1981) and was implicit in Kaldor (1985). The main target was the naive monetarism of the first UK Thatcher government. Their interest in the supply of money, therefore, was to show that it could never be in excess supply in a way that threatened the stability of the price level. After all, if it were possible for the demand for credit to result in a stream of new deposits which were in some sense ‘excessive’ in relation to demand, then this opened the troublesome possibility that the desire to run down these deposits would result in an increased demand for goods and services and the whole monetarist sequence could re-emerge whereas if the money supply were endogenously determined (let us say by passively responding to the growth of nominal income) then the causality is reversed. Thus Kaldor’s purpose was an attack on the Quantity Theory and all its works rather than a thorough discussion of the dilemma we have posed here. Nonetheless a mechanism was required that would ensure the permanent equilibrium referred to by Cuthbertson. The device that Kaldor envisaged for the reconciliation of deposit creation with money demand was the automatic use of excess receipts of money for the repayment of overdrafts. Thus, the individual actions of borrowers taking out new loans (or extending existing ones) could threaten an ‘excess’ creation of deposits *ex ante*, but the actions of other (existing) borrowers in immediately repaying some of their debt would mean that the net deposits which resulted *ex post* would be only what people wished to hold.

‘Automatic’ is the keyword. It is the way that overdrafts work that the size of the debt is automatically reduced by the receipt of payments and this will (‘automatically’) reduce the quantity of new deposits that are actually created. The problem is - not everyone has an

15 The title was chosen to emphasise the difference between an exogenous money supply, conventionally represented by a vertical money supply curve in interest-money space (as in figure 1 above) and an endogenous money supply which could be represented by a horizontal money supply curve. Unfortunately, the contrast is misguided and has led to much confusion and error in attempts to represent an endogenous money supply in a simple diagram. (See Howells, 2001, pp. 159-167 and the references cited therein).

16 The fact that the demand for money does play some role in determining an endogenously created money supply is suggested by causality tests that suggest some feedback from the change in deposits to the flow of new lending. It is not simply the case that ‘loans create deposits’.
overdraft, an observation made by Cottrell (1986) and by Chick (1992, pp. 204-5). And it is not sufficient to argue that some people somewhere (e.g. virtually all firms) do have overdrafts. Once it is accepted that the first round recipients of ‘new’ money may not wish to hold it, then the genie threatens to leave the bottle. The question remains: how are the ‘excess’ balances to be disposed of?

It is significant that many of the contributors to this debate regarded themselves as ‘post-Keynesians’ since the endogeneity of money has been a cornerstone of post-Keynesian economics for many years (Fontana, 2003, p. 291). And for many of them, the significance of this endogeneity, as it did for Kaldor, lay in its reversal of the classical notion that changes in the quantity of money were causally responsible for changes in the price level alone (at least in the long-run).

In post-Keynesian circles, the debate has subsided somewhat in recent years. This may hint at a consensus, and if it does then the consensus is probably based on two foundations. The first is the notion that money does have special characteristics which mean that the willingness to hold it is to some degree elastic, even with unchanged values in other variables. Ironically, there are echoes here of Laidler’s (1984) ‘buffer stock’ notion: the demand for money is not a point demand but a range.

But this leaves the question of what happens in those circumstances (which maybe exceptional) when the *ex ante* change in deposits resulting from loan demand, differs so far from the willingness of agents to hold this extra liquidity that it breaches the limits of the buffer? The consensus here appears to involve an adjustment in relative interest rates that has a distinctly Keynesian ring to it. Take the case where the demand for credit creates new deposits in excess of those demanded in present circumstances. Agents, individually, attempt to run down their deposit holdings by buying assets. Collectively, this is self-defeating - causing only a redistribution of deposits. However, the redistribution is accompanied by a rise in asset prices and a fall in their yields. The return on bonds falls, relative to money’s own rate. This change is the well-known mechanism traditionally cited in the textbook account of how changes in money supply are reconciled with money demand. Its effect is relevant here, in so far as a fall in the rate on non-money assets moves us down the money demand curve and yields a one-off increase in the demand for our excessively growing deposits. However, non-money assets are the liabilities of non-banks. They are liabilities issued by non-banks as a means of raising funds. To some degree, therefore, they are substitutes for bank loans. As the rates on corporate bonds and short-term paper (for example) fall relative to the rate charged on bank lending, so there is a fall in the price at which the economic units whose liabilities these are can raise new funds. If the yield on existing corporate bonds falls, new bonds can be issued with these lower yields and bond finance becomes cheaper, at the margin at least, relative to bank finance. With the cheapening of a partial substitute for bank finance, the demand curve for bank lending shifts inward and the demand for bank credit falls. It is this change in relative interest rates that brings the *ex post* demand for bank lending (and the ongoing flow of new deposits) into line with the community’s increasing demand for money.

Ultimately, the flow of new loans is matched by the willingness to hold the new money (as it must be). But the process by which the excess growth (in this example) involves agents individually trying to divest themselves of excess money balances and changes in interest rate spreads, both of which may have some effect on aggregate demand. It is no longer clear that changes in the money supply are entirely passive.
However, while this debate has subsided in post Keynesian circles it has resurfaced recently, with interesting echoes of the earlier discussion. Towards the end of section 2 we noted that the endogeneity of the money stock is now widely accepted as part of the new consensus macroeconomics and that one consequence of this is an ambivalence about the role of monetary aggregates in the determination of output and inflation. (See also footnote 3). So far as the conduct of monetary policy is concerned, the ECB is unusual in including a ‘reference value’ for the growth of M3 as part of its ‘two-pillar’ strategy. But even the ECB has expressed recent doubts as to whether the evolution of M3 provides any useful information, over and above that contained in the variables that it monitors as part of the second pillar (Atkins, 2007).

The current revival of interest in the information content of monetary aggregates has its origins in recent upheavals in credit markets where conventional interest rate differentials have broken down. The best known and most dramatic example is the jump in LIBOR relative to the UK policy rate (and similarly in the USA and eurozone) in August 2007. But a recent paper by Chada et al. (2008) looks at the behaviour of a different spread, one which has some affinity with the bank loan – deposit spread that we have just seen playing an important role in the reconciliation of the demand for loans and the demand for money.

The spread in question is described as the ‘external finance premium’ (EFP) and is defined as ‘...the difference between the opportunity cost of internally generated finance and the cost of issuing equity or bonds’ (Chada et al., p. 3). Looking at US data from 1992 to 2007, the paper shows that the growth of real M2 and the EFP are positively correlated until about 1995 whereafter the correlation turns negative. In other words, increases in real money balances seem to lead to a compression of the EFP and they interpret this as evidence of money supply shocks dominating the market from the mid 1990s. As possible causes of such shocks they offer changes in the value of collateral (offered for bank loans) and costs of screening applications for such loans. The argument in brief, therefore, is that changes in bank lending behaviour can cause increases (or decreases) in liquidity which in turn cause changes in interest rate spreads which ultimately have an impact on aggregate demand (and may diverge from what was intended in the setting of the policy rate). For this reason, policymakers do need to take account of what is happening to the monetary aggregates as well as to the policy rate. In section 4 of their paper they show how the policymaker needs to be able to offset changes in the EFP and how a rule, incorporating changes in money can be formulated to achieve that.

Although this particular perspective on why money aggregates might matter even when the money supply is endogenous, has evolved quite independently of the earlier debate in the post Keynesian literature, the similarities are quite striking. The proximate cause of new deposits is net new lending. If the demand for credit (at a given rate of interest) depends solely upon the evolution of nominal income, then the money stock is a passive reflection of events elsewhere in the real economy. But if the demand for loans is subject to shocks which are independent of the path of nominal income then there is a change in interest rate differentials which have an effect on aggregate spending. Furthermore, the sources of the shocks are not so very different. In Chada et al (2008) it is changes in the value of posted collateral and/or changes in banks’ screening of loan applications. The possibility that asset price bubbles might influence the demand for credit independently of the requirements of the real economy must strike any observer of recent events as an obvious possibility. But this is not so very different from the earlier debate in the post Keynesian literature as to whether
loan demand is driven solely by the ‘state of trade’ or whether it is better explained by recourse to some broader range of transactions including asset purchases.

6. CONCLUSIONS

The debate as to whether the money supply is exogenously or endogenously determined goes back a long way. Almost certainly, it is impossible to answer this without reference to a particular context. One might imagine, for example, that a money supply consisting solely of coins minted from precious metals is more likely to be exogenous than one that consists almost entirely of bank deposits. But even in the former case, as we have seen, there is room for debate.

While one can certainly trace the argument that the money supply must be endogenous in a modern economy back to the end of the nineteenth century, it has been the post Keynesian economists like Kaldor, Davidson, Moore (but others too) who have done most, in the last forty years, to develop a monetary economics founded on the interaction of banks’ commercial interests with the needs of their customers. And throughout this period, central banks have, as a matter of practice, set interest rates and allowed banks and their customers to negotiate their preferred outcomes.

In the circumstances, it is difficult to know why macroeconomic textbooks have persisted for so long with the fiction that the money supply is exogenously determined and in so doing have exposed generations of students to misinstruction. For a graduate student interested in methodology and/or the history of economic thought, there is a thesis waiting to be written.

The imperative to design and conduct an optimal monetary policy in the real world, however, has finally forced a reappraisal in the form of the ‘new consensus macroeconomics’ and this, at last, is beginning to force a realistic treatment of money in the latest textbooks. It is curious though that what is hailed as a ‘consensus’, appears to make no reference to almost two generations of earlier work, even when that work touches on issues that are now coming into focus again. There is another thesis waiting to be written.

REFERENCES


