Towards A Reformulation of Monetary Theory: Competitive Banking*

BRUCE GREENWALD
Columbia University, New York
and
J.E. STIGLITZ
Stanford University, California

I INTRODUCTION

Neo-classical economics is a difficult subject. It requires more than the usual willing suspension of disbelief. One is asked to put aside one's economic intuition, one's experiences of how the world works, to enter the realm of ideas: to begin with seemingly plausible or not terribly implausible assumptions (or assumptions which are implausible, but which are justified as convenient "simplifying" assumptions), and then through a process of deductive reasoning, arrive at surprising conclusions. The more surprising — that is, the less sensible — the conclusion, the greater the mark of the theorist, for he has had to rely on his analytic powers, not his grasp of economic affairs. And the mark of the truly great scholar is a Ptolemaic ability to reconcile observed seeming contradictions with the edifice which has been slowly erected over, by now, almost a century.

*This is a shortened and revised version of an Invited Lecture delivered to the Irish Economic Association in May, 1991 by Stiglitz. Financial support from the National Science Foundation and the Hoover Institution is gratefully acknowledged.
Three examples should suffice: The central theorem of monetary economics is that money does not matter — the classical dichotomy; money supply affects price levels but not relative prices or output. The central theorem of corporate finance is that corporate finance does not matter — the Modigliani Miller theorem. The value of the firm is independent of its debt equity ratio, or any other aspect of its financial structure. A central result in modern public finance theory is the so-called Ricardian equivalence theorem, which in its most general form (Stiglitz, 1987) states that public financial policy is irrelevant. It makes no difference whether the government finances its expenditures with debt or taxes, and changes in the maturity structure of the government debt (operation twist) have no consequences.

A curious twist seems to have been added to the so-called Friedman twist: while Friedman argued that theories should be judged by the validity of their conclusions, not by their assumptions, in many of these cases, neither the assumptions which go into the theory nor the conclusions which follow from it are directly tested. Some of the derivative implications of the theory are tested, or are at least consistent with observations; but a theory is to be judged by all of its testable propositions (see Stiglitz, 1989c).

In this realm of pure reason, there is accordingly great scope for economists to choose their assumptions and conclusions according to a broader Weltanschauung: ideology and politics are the true masters. Those who want to see government play a small rôle advocate the neo-classical model, arguing (i) government is not needed (the economy is Pareto efficient); (ii) even if the economy were not Pareto efficient, government policies are ineffective (the policy irrelevance proposition); and (iii) even if government policies could do something, governments are more likely to make matters worse than better.

The first hypothesis — the formalisation of Adam Smith’s invisible hand — is held in spite of periods of massive unemployment and the accompanying social waste of resources. When these are acknowledged, they are laid down to temporary aberrations, and if this will not suffice, the discussion moves on to the second and third contentions.

The second set of contentions, too, seemingly flies in the face of the evidence — does anybody really believe that Paul Volcker had no impact on the economic events of the late 1970s and early 1980s, or that the US budgetary deficits have had no real consequences! The fact that during some periods in which budget deficits varied very little, the coefficient on deficits in some

1. Barro (1974) is usually given credit for the revival of the Ricardian equivalence proposition. (Whether Ricardo actually held this view seems to be a moot question.) Several years earlier, Robert Hall presented a more general version of that result to a graduate seminar at MIT, and was discouraged not only from pursuing the matter further, but from publishing it, on the grounds that it was both trivial and obviously wrong!
badly designed regression comes out small is hardly evidence on the point.

The third hypothesis brings the economist outside the realm of his discipline, and indeed, largely outside the realm of pure thought in which we have dwelt so far, into the arena of broad historical generalisations, about which reasonable people can, and do, differ.

Those on the other side, advocating more active government intervention, have been equally ideological, even if their theory has not been as finely honed.

Fortunately, during the past fifteen years, there has been a radical change in our discipline, a change for the better, which holds open the possibility of a truly scientific discourse. This change is the recognition of the existence, pervasiveness, and importance of informational imperfections (and other transactions costs). They affect how households, firms, and government interact in product, labour, and capital markets. While the first phase of the exploration of the consequences of imperfection focused on how none of the standard theorems of neo-classical economics were robust — slight changes in the information assumptions caused drastic changes in the underlying conclusions — the second phase attempted to show how informational problems affected how each of the principal markets of the economy functioned. In the last few years, we have entered the third phase — showing how the insights about how each of the principle markets function can contribute to an understanding of how the economy as a whole behaves. In particular, we have focused our attention at explaining macroeconomic phenomena, with considerable success: we have explained wage and price rigidities; we have explained not only why there may be large fluctuations in hours worked, but why so much of the decrease in hours worked takes the form it does — layoffs, rather than work sharing. We have explained why seemingly small shocks can give rise to large consequences.

In this paper, we wish to turn to an old question — to a re-examination of the role of monetary policy. We want to explore what our new view of how the economy functions has to say about this fundamental question of economic policy. The paper will be divided into three parts. In the first, we will revisit, from our new perspective, the old and familiar IS-LM curve, and the monetary theory upon which it is based. We will explain why that framework — or at least its underpinnings as presented in standard textbooks — is fundamentally flawed.

We are all aware of the rapid changes that the computer revolution has

2. There were problems with existence of equilibrium, with the optimality of market allocations, as well as with the characterisation of market equilibrium. For a survey, see Stiglitz (1985).
brought with it: nowhere are these more significant than in the recording of transactions, one of the central functions of banks. This revolution has potentially profound effects on the structure of financial institutions, some of which are already on the way. In the third section, we construct a model of the banking system, based on the assumption of competition within the financial sector, recognising the changes in that sector which these changes have already set into motion. The competition among banks is sufficiently keen to drive deposit rates to approximate T-Bill rates closely. We examine how and why various policies of the central bank affect bank lending behaviour. The third section shows how this model of banking behaviour can be embedded within a general equilibrium model of the economy. We show how we can construct a modified IS-LM framework, and we identify the salient ways in which the new framework differs from the old.

II REFLECTIONS ON THE CURRENT STATE OF MONETARY ECONOMICS

To understand business fluctuations, we must understand the determinants of investment. To neo-classical as well as Keynesian economists, analysis of investment begins with an analysis of the determinants of the interest rate. But it is here that the two depart: modern neo-classical economics, reflected today most starkly by real business cycle theorists, see the interest rate governed by real factors — by the savings of households (preferences) and the marginal productivity of capital (technology), with monetary factors playing no rôle.

Real Business Cycle Theories

Even if there are periods for which these real factors dominate, the seeming enormous effect that Volcker had in initiating the recession in the late 1970s and early 1980s in the United States, to which we referred earlier, as well as other episodes of credit crunches, makes the view that only real factors matter untenable as a basis at least for short-run macroeconomic analysis. But higher minded
economists have looked derisively at those who made reference to imperfect capital markets, accusing them of, among other sins, "ad hocery".

However, one of the most important developments in economic theory of the past fifteen years has been to explore the consequences of imperfect and costly information for the functioning of the capital market. These studies have shown that those models which assumed imperfect capital markets may have been much closer to the mark than those which, on the contrary, assumed perfect capital markets. These studies have shown that capital markets that are competitive — in the sense which that word is commonly used — may be characterised by credit and equity rationing. These models both provide explanations of institutional details of the capital market, details which are either inconsistent with the perfect capital market models or about which they have nothing to say; but they also provide a basis of an explanation of macroeconomic (aggregate) behaviour which are also inconsistent with the conventional neo-classical model. Some aspects of this will be explored below. For now, we simply note that the imperfections in the capital market have real consequences; they explain why real business cycle theory went wrong, and why monetary institutions and policy matter. But before turning to see why that is the case, we must explain why the alternative view of monetary policy provided by (at least one of the traditional forms of) Keynesian theory also is unpersuasive.

A Critique of the Transactions Based Theory of the Demand for Money

There are several reasons why one might be suspicious of the traditional explanations of how monetary policy works. Keynes was, perhaps, not as clear as he might have been concerning the definition of money that he had in mind. The absence of clarity may have been deliberate: it enabled him to slip from one use to another, without the reader being aware of what was going on. We focus our attention on demand deposits, because these are the part of money which is most directly under the control of the monetary authorities, and slightly more broadly, on M1, which includes currency.

Keynes spoke of three motives for holding money: the precautionary motive, the speculative motive, and the transactions motive. Given our definition, only the third is relevant. The other two motives are related to the use of money as a store of value; and as a store of value, money is dominated by treasury bills and money market mutual funds, which yield higher rates of interest. Though some economists have suggested that Keynes really had in mind a definition of money that includes these assets ("L" in the standard terminology), surely this broad aggregate is not under the control of the monetary authorities; even when monetary authorities set the money supply as a target, they never focus their attention on this broad measure. Empirical
studies have concentrated their attention on narrower definitions, such as M1 (or, as M1 has done increasingly poorly, on M2).

Thus, an analysis of the demand for money must focus on the transactions demand for money. (Of course, in a general equilibrium model, demands for all assets are interdependent. Still, it is useful to think of the individual as first deciding on how much he wishes to hold in short-term dollar denominated government insured assets, and then to ask, of that total, how much should be held in the form of money, for transactions purposes.\(^4\))

The past fifteen years have witnessed remarkable changes in transactions technologies. Computers enable the velocity of circulation to become virtually infinite, for instance, in the use of money market accounts. The changes in the relationship between conventionally measured money and income, while they have not been stable in recent years, have not moved in a way that a transactions-based theory would have predicted. These events have forced us to reconsider the transactions-based theory of the demand of money. Upon reflection, it becomes clear that the theory was — and should have been recognised to be — badly flawed.

The traditional theory argued that the demand for money depended on the level of national income and the interest rate. Let us consider the arguments relating to each of these variables.

First, it is not apparent that money should be closely related to income. There are two basic reasons for this.

(a) Most transactions are exchanges of assets, and are not related directly to the production of income. And there is no a priori reason to expect that the relationship between the two should be stable. Indeed, there is reason to believe that in periods of rapid change, such as when the economy is going into a recession, there will be large differences in opinions concerning future prospects of the economy, as well as large changes in relative net wealth positions, leading to corresponding larger than normal exchanges of assets.

Of course, many, if not most, asset exchanges may not entail money (or may entail money with a much higher velocity than transactions involving labour services.) But that is exactly our next point:

(b) Transactions do not, in principle, require money, only credit. Just as once the economy "discovered" money and one no longer had to have the double coincidence of wants to enable exchanges to occur, so too, once the economy discovers credit, an individual no longer has to have money to acquire goods. Of course, individuals cannot always use credit to pay for

---

4. This approach would not even run into problems if the amount of money wanted to hold for transactions purposes exceeded the amount of short-term government insured assets an individual would wish to hold; in that case, he would simply sell short the T-bills.
goods. Understanding when goods can be acquired with credit should be an essential aspect of any analysis of the demand for money. But certainly, theories which begin by assuming that cash is required (which simply posit a cash in advance constraint) are, at best, *ad hoc* and incomplete, at worse, wrong; and at the very least, cannot be relied upon as a basis of policy: for the set of transactions which may rely upon credit can clearly change with changes in economic circumstances.

Nor does it do much good — as several models currently fashionable in some circles have done — to explain away the presence of credit by constructing models in which individuals meet only once, and in which there are no financial institutions. While true enough, in such a world, a cash in advance constraint can be endogenously explained, that world has sufficient little semblance to our world — in the key respects in which we are interested — that there is little that we can learn from such finger exercises.

The second set of objections relate to the dependence of the demand for money on the interest rate. The standard argument was simple: the interest rate represented the opportunity cost of holding money. With the recent growth of interest bearing accounts, the *opportunity cost* of holding money has, by and large, become the *difference* between the interest rate paid on money market accounts which allow checking accounts and those paid on other short-term financial instruments — a difference which is, for economic purposes, minuscule. It is this interest rate differential which should appear in the money demand equation (and in the LM curves of macroeconomic analysis). It stretches the credulity of even the most hardened Keynesian to believe that monetary policy operates through changes in the real value of this interest rate differential. Of course, in a general equilibrium system, any exogenous change will have effects throughout the system. Were the government to buy up peanuts and burn them, it would have real effects on output, investment, and employment. But though there is a long tradition in macroeconomics on focusing on third and fourth order effects — dating at least back to the real balance effect — there is a consensus that a peanut theory of macroeconomic policy will not do; so too should a theory based on changes in the differential between money market accounts and short-term government bills be looked upon with suspicion, though there may indeed be links between this differential and the appropriate long-term real interest rates, and between those interest rates and the level of investment.

The fact that today money is interest bearing will play an important rôle in the analysis of the next section. But we should remember that even when banks were prohibited from paying interest directly, they may have been able to do so indirectly, e.g., in the context of extended creditor-debtor arrangements, by “charging” correspondingly lower interest rates during periods in
which the individual was a net debtor to the bank.

Monetary policy is supposed to affect the economy (in the traditional IS-LM framework) through changes in the interest rate, which affect the level of investment. Of course, the demand for money depends — in the traditional model — on the nominal rate of interest, while the demand for investment is supposed to depend on the real interest rate. There is a clear and direct link between nominal and real interest rates only if expectations concerning rates of inflation are unaffected by monetary policy itself — a dubious proposition.

In our view, there have been long periods in which it does not seem plausible that variations in long-term real interest rates have played an important rôle in fluctuations in investment. It is hard to know, of course, what the relevant real interest rates are. We have time series for nominal interest rates for loans of various maturities, but what is relevant for an investor, at the time he makes his decisions, is the real interest rate, and to calculate this, he must form a forecast of future prices. (This would not be true, of course, if all loans were appropriately indexed.) As a first pass at this problem, Jaffee and Stiglitz (1990) looked at the realised real rates of interest (Figure 1). We constructed a time series for what the real rate of interest would have been on five year government bonds (so we can ignore variations in rates of default), assuming the market acted as if it had perfect foresight and was risk neutral. Such a hypothetical rate would have been the rate that would — in neo-classical models without equity and credit rationing — be relevant for investment decisions in five year machines. Though there has been some secular changes in this real interest rate, rising in the 1980s to record heights, there is virtually no relationship between this, the relevant real interest rate, and the level of investment.\(^5\) (Similar results hold for the ten year real interest rate.) With a few exceptions — the Great Depression and the 1980s — one might as well treat the real interest rate as a constant. And constants do not provide a basis of a good theory of fluctuations.\(^6\)

5. The fact that, by and large, changes in the rate of inflation offset changes in nominal interest rates was noted long ago, and is often referred to as Fisher's Law. It has clearly not always been true, but it has been true over long enough periods to cast doubt on the rôle of the interest rate mechanism.

6. Hicks (1988) provides a figure of long-term interest rates (consols) over a period of more than a century (reproduced here as Figure 2b). Again, little variation in the rate of interest is observed. Since for much of the period, rates of inflation were low and stable, the low variation in long-term interest rates translates into low variation in real interest rates.
Figure 1a


Figure 1b

This is not surprising, given businessmen's accounts of how they make investment decisions. Though interest rates enter the calculations, they conventionally require real returns of 15, 20, or 25 per cent before they undertake a project; given the uncertainties associated with all aspects of long-term investment projects, a variation of the real interest rate from 3 to 4 per cent is absorbed in rounding error.

The scepticism that we have expressed here and elsewhere about the rôle of interest rates has, of course, been implicitly or explicitly shared by monetarists. They have noted the regularity between the money supply and the level of output, with interest rates seemingly playing a relatively minor rôle. Leaving to theorists the task of explaining this sometime regularity, they have been content to base policy prescriptions on the persistence of this relationship, to argue that if only we increase the supply of money, output will go up. The failure of the stable relationship between money and income (illustrated in Figure 2) to hold in the 1980s, and the failure of the prescriptions based on it should not detract us from the essential insight of the monetarists — the seeming unimportance over long periods of time of the interest rate. But the changes in essentially unpredictable ways of velocity during the decade do mean that theoretical models, based on the hypothesis that "somehow" the economy operates in such a way as to maintain a constant money income ratio, have lost their standing even as an intermediate level of theorising.

There are further objections to the transactions based theory of money: when Ireland faced a strike which closed down the clearing mechanisms for cheques, while the transactions based theories might have suggested that the economy would have come to a screeching halt, alternative arrangements were easily worked out, and the effects were indeed limited. Italy has periodically gone through periods of shortages of small currency, with little impediment in trade, the major effect perhaps being the increased number of cavities resulting from the slightly larger number of pieces of candy consumed, as candy became conventionally used for small change.

In light of the changing relationship between money and income, there has been a search for alternative definitions of money. But this simply serves to show the ad hoc nature of the enterprise. The failure of the money-income relationship should serve as a spur to a re-examination of the bases of monetary theory, no less than did the seeming failure of the inflation-unemployment trade-off in the late 1970s call for a re-examination of other aspects of the postulates of traditional Keynesian theory.
Velocity

Source: Citibase Data Bank

Figure 2
III THE BANKING SYSTEM OF THE FUTURE

To understand how our current banking system works, we need to think through how an idealised banking system might work — a banking system not too different from what may emerge in the not too distant future. The central features of this banking system are (a) government insured deposits; (b) government imposed reserve requirements, with reserves held at the Central Bank in interest free accounts; and (c) no transactions costs. The banking system must compete against money market mutual funds, which invest in government T-bills, and which provide checking service comparable to those provided by the bank. They have no reserve requirements, and they pay an interest rate equal to the T-bill rate. (This follows from assumptions concerning competition in the industry and no transactions costs.) Here, we want to highlight certain fairly obvious implications of the model.

The fact that deposits are government insured means that depositors should be indifferent between holding their funds in the money market accounts and in banks. This means that the interest rate paid by banks to their depositors must equal that on government bonds.

Each bank thus views itself as facing a horizontal supply curve for funds. If it pays ε more than its competitors, it can get as much funds as it wants.

While the assumption that deposit rates are set competitively seems at variance with "traditional" bank models, the discrepancy may be more apparent than real. Even when the bank was proscribed from paying interest, the bank may have provided other services, e.g., extending a credit line at below "the fair market value" of such a credit line; or lending funds at a lower interest rate. In a competitive market, one would expect banks to compete in the provision of these auxiliary services when they cannot compete directly in terms of interest rates paid.\footnote{An important difference between our model, and a more realistic version of the traditional theory, is that we assume that deposit rates adjust instantaneously. We suspect that these adjustment costs play an important role in understanding the short-run effects of monetary policy, but we cannot pursue the matter further here. See Greenwald and Stiglitz (1992).}

Each bank has a certain amount of equity. Banks — like other businesses — are averse to going bankrupt — a fact that might not be so evident from the behaviour of the S & Ls in the United States; we shall return to this puzzle later. For reasons that are by now well known,\footnote{There are three parts to the argument: (a) Equity markets are in fact a relatively unimportant source of funds (see, e.g., Mayer, 1990); (b) This is not surprising, given that when firms do issue equity, it has a marked negative effect on their net worth (see, e.g., Asquith and Mullins, 1986); and (c) There are strong theoretical reasons why this should be so. (See, e.g., Greenwald, Stiglitz and Weiss, 1984; and Meyers and Majluf, 1984 for a discussion of adverse selection models; or Stiglitz, 1974 or Jensen and Meckling, 1976 for an analysis of incentive effects. Other theories for the lack of use of equity include those focusing on costly state verification (see, e.g., Harris and Townsend, 1981).} we assume that banks

7. An important difference between our model, and a more realistic version of the traditional theory, is that we assume that deposit rates adjust instantaneously. We suspect that these adjustment costs play an important role in understanding the short-run effects of monetary policy, but we cannot pursue the matter further here. See Greenwald and Stiglitz (1992).

8. There are three parts to the argument: (a) Equity markets are in fact a relatively unimportant source of funds (see, e.g., Mayer, 1990); (b) This is not surprising, given that when firms do issue equity, it has a marked negative effect on their net worth (see, e.g., Asquith and Mullins, 1986); and (c) There are strong theoretical reasons why this should be so. (See, e.g., Greenwald, Stiglitz and Weiss, 1984; and Meyers and Majluf, 1984 for a discussion of adverse selection models; or Stiglitz, 1974 or Jensen and Meckling, 1976 for an analysis of incentive effects. Other theories for the lack of use of equity include those focusing on costly state verification (see, e.g., Harris and Townsend, 1981).
are constrained in the amount of equity that they can raise; for simplicity, we assume that their equity is fixed. There is a cost to going bankrupt,\(^9\) and thus banks maximise their expected profits taking into account the effects of their decisions on their bankruptcy probability.\(^{10}\)

The essential consequence, for our purposes, of the observation that banks are equity constrained is that they act in a risk averse manner, and that their level of net worth affects their behaviour. This is an important departure from the standard neo-classical paradigm, with perfect risk markets. In such markets, risks are effectively spread throughout the economy, and banks (at least with respect to risks which are uncorrelated with the business cycle) act in a risk neutral manner.\(^{11}\) And since banks can easily raise capital, the amount of a bank's net worth is of no relevance: if there are good lending opportunities, it can instantaneously raise the funds with which to make the loans; and it can raise the funds in a way which imposes no additional risk of bankruptcy.\(^{12}\)

For simplicity, in this paper, much of the time we shall just assume that banks are risk averse; and indeed, for diagrammatic purposes, we shall assume that banks' preferences can be related to the mean and standard deviation of their terminal wealth.

**The Bank's Portfolio**

Any bank faces a potential loan portfolio which, for simplicity, can be described by its mean, \(\mu\), and standard deviation, \(\sigma\), per dollar lent. Each of these is a function of:

---

9. See, e.g., Greenwald and Stiglitz (1988a) for a discussion of the nature of these costs; these include both the direct costs of reorganisation; the cost in terms of lost opportunities which arise as a result of the restrictions that are effectively imposed during the period of reorganisation (e.g., the decreased willingness of others to engage in certain types of trades); and the costs to managers — not only the (often partial) loss of control, but the adverse signal concerning their competency which results from bankruptcy.

10. Actually, for a regulated industry like banking, we should substitute the condition that firms are averse to going bankrupt with the condition that firms are averse to regulatory intervention, the strong form of which is a regulator mandated change in management, as occurred in the case of Continental Illinois.

11. In fact, since many of the important risks upon which we shall focus on business cycle risks, some of our results would obtain even if banks were not equity constrained.

12. When a bank's net worth is too low, in the formulation in which banks maximise expected terminal wealth minus expected bankruptcy costs, they will actually act as risk preferers: bankruptcy itself, as is well known, introduces a non-concavity into the pay-off function, which, at low levels of net worth, more than offsets the concavity introduced by the expected bankruptcy costs. This non-concavity is important in understanding the risk taking behaviour of the S & Ls in the United States during the 1980s, leading to the eventual demise of so many of them. See below and Stiglitz (1991a).
(i) The total number of dollars lent. Normally, banks have specialised information about firms with whom they have regular relationships; they have less information about others. As they lend more, their ability to screen accurately is reduced; they go "down the list" of borrowing prospects. At the same time, lending more to the same borrowers increases the risk of default, lowering the mean and increasing the standard deviation.

By the same token, increasing the interest increases the standard deviation of the loan portfolio. We assume that the returns to the different loans (the bankruptcy probabilities) are not independent. There is a strong cyclical component to bankruptcy. Accordingly, the more the bank lends, the greater the risk.

(ii) The interest rate charged. Normally, we would have expected an increase in the interest rate charged to increase the bank's expected return. Stiglitz and Weiss (1981) showed that, among a group of seemingly identical borrowers, increasing the rate of interest had an adverse effect on the mix of loan applicants. If this effect is strong enough, the expected return may actually be lowered as the interest rate charged increases. (See Figure 3.) At higher interest rates, safer borrowers decide it is not worth borrowing; those who have high probabilities of defaulting remain interested.

![Figure 3](image-url)
They also showed that there were adverse incentive effects associated with charging higher interest rates (see also Keeton, 1980). At higher interest rates, borrowers undertake riskier projects, lowering the bank's expected return and increasing its risk.13

These interest rate effects arise because the bank has only imperfect information, both concerning the nature of the loan applicants and the actions that loan applicants undertake. Of course, the bank has some information, and it has some control over what borrowers do with its funds. But so long as its information is imperfect, and its control is limited, the effects that we have just described may occur, and they may be important.

(iii) The bank's expenditure on screening loan applicants and monitoring the uses of its funds. The screening costs, in particular, represent up front expenditures, the return to which accrues only subsequently, when the loans are repaid.

There are other determinants of the mean and standard deviation of the bank's loan portfolio, e.g., its collateral requirements. For simplicity, however, we shall ignore these other determinants.

We thus write for the representative bank

\[ \mu = \mu(N, r, e) \] (1)

\[ \sigma = \sigma(N, r, e) \] (2)

where

\[ N = \text{number of dollars lent} \]
\[ r = \text{interest rate charged} \]
\[ e = \text{expenditures on screening and monitoring.} \]

More generally, we can write the bank's gross returns from its loan portfolio as

\[ Y = Y(N, r, e, \Theta) \] (3)

where

\[ \Theta = \text{state of the business cycle, representing the undiversifiable risk of the banks' portfolio.} \]

Without loss of generality, we assume that \( Y_\Theta > 0 \). Also,

\[ \mu N = EY, (\sigma N)^2 = E(Y - \bar{Y})^2 \] (4)

13. Risk will only be affected to the extent that the riskier projects are more highly correlated with, say, the business cycle. Generalisations of the arguments presented in Stiglitz and Weiss (1981) can be used to show that this will in fact be the case. See also Greenwald and Stiglitz (1992).
To simplify matters, we formulate the bank's problem within a two period model. The bank has an initial net worth, $a_t$. It can obtain funds at an interest cost of $\rho$. The end-of-the-period wealth of the bank is (taking account of the possibility of bankruptcy)

$$a_{t+1} = \max \{Y + M(1 + \rho) - \rho(1 + \tau) (N + M + e - a_t), 0\}$$

(5)

where $\rho = $ the interest rate paid depositors (the T-bill rate), $M$ is bank's investment in T-bills, and $\tau$ is the effective tax rate imposed on deposits (as a result of the reserve requirements).\(^{14}\) Thus, if $\phi$ is the reserve requirement, if a bank wishes to make $\$B$ of investments, financed by borrowing, it must borrow $\$D$, where

$$(1 - \phi) D = B$$

(6)

and the cost of those funds is

$$D(1 + \rho) = B(1 + \rho)/(1 - \phi)$$

(7)

Thus

$$1 + \tau = 1/(1 - \phi).$$

(8)

The bank goes bankrupt if

$$Y + M(1 + \rho) - (1 + \tau)(N + M + e - a_t) \leq 0,$$

i.e., there exists a critical value of $\Theta$, $\hat{\Theta}$, such that the bank goes bankrupt if $\Theta \leq \hat{\Theta}$, and not otherwise.

Let $F(\hat{\Theta})$ be the probability that $\Theta \leq \hat{\Theta}$, i.e., the probability that the bank goes bankrupt, and let $c$ be the cost of bankruptcy ($c$ will normally be a function of the scale of the bank; bankruptcy costs will be larger the larger the bank, e.g., measured by the value of assets or liabilities).

The Bank’s Objectives

There are two alternative ways of modelling bank behaviour. They both yield similar results, namely, that the bank acts in a risk averse manner.

The first is based on the assumption that the bank is risk neutral, provided it does not go bankrupt, but there is a high cost to bankruptcy, so it naturally

\(^{14}\) It will become clear later the sense in which, and the conditions under which, reserve requirements act as a tax.
wishes to avert bankruptcy.\textsuperscript{15} Formally, we assume the bank
\begin{equation}
\max_{Ea_{t+1}} - cF_{(N, M, r)} \tag{10}
\end{equation}

s.t.
\[ N \leq N^d (r, e), \]

where $N^d$ is the demand for loans at the interest rate charged, $r$. This formulation recognises that banks might want to choose an interest rate below that at which demand equalled supply; but that it cannot "force" borrowers to borrow. When the constraint is not binding, it means that there is credit rationing: those willing to borrow at the interest rate (and other credit terms) offered by the bank exceed the loans that the bank is willing to make. We focus on that regime, because it greatly simplifies the analysis. Most of our results, however, remain valid even in the absence of credit rationing.

Equation (10) says that the bank chooses its loan portfolio, its loan policy (here reflected in the interest rate charged) and its investments in government bonds to maximise the expected value of its terminal wealth, minus the expected value of the bankruptcy penalty.

The second alternative formulation, and the one upon which we shall focus in much of this paper, is simply to assume that banks act in a risk averse manner, that is they maximise the expected utility of terminal wealth, where utility is a (strictly) concave function of terminal wealth,

\[
\text{Max } EU_{t+1}
\]

where $U'' < 0$. A special case of this, which will be particularly useful for diagrammatic analysis, is that where\textsuperscript{16} the bank can be viewed as maximising

\[ U = V(\mu N, \sigma N), \tag{12} \]

a utility function which can be represented as a function only of the mean and standard deviation of terminal wealth.\textsuperscript{17}

\textsuperscript{15} See Greenwald and Stiglitz (1988a, 1990c) for alternative interpretations of this assumption, e.g., it may represent the consequences of managerial behaviour.

\textsuperscript{16} We assume that the pattern of returns and the utility functions satisfy one of the conditions under which mean variance analysis is valid. See Cass and Stiglitz (1972), Stiglitz (1972).

\textsuperscript{17} For simplicity, when we use the mean standard deviation model, we shall ignore the implications of bankruptcy for the mean and standard deviation of return. The mean standard deviation diagram is presented as only a heuristic for understanding what is going on, and the corrections required to incorporate bankruptcy are sufficiently complicated that they obfuscate the essential points which we wish to emphasise.
For the remainder of the paper, we focus on banks' lending and borrowing behaviour, keeping $e$ fixed (for simplicity we set it equal to zero).

**Basic Propositions**

We summarise the implications of this model for bank behaviour in five propositions:

**Proposition 1.** The bank never borrows to buy government bonds, i.e.,

$$\text{If } p > 0, \text{ so long as there is a strictly positive reserve requirement, and } B = N + M - a_t = \text{bank deposits} > 0, \text{ then } M = 0.$$  

This follows from the fact that there is a tax on deposits; the bank must pay (before tax) an interest rate which is exactly equal to the interest rate it receives on government bonds. Thus, in every state of nature, $a_{t+1}$ is lower, and hence $Ea_{t+1}$ is lower, and the probability of bankruptcy is higher. (Moreover, if the costs of bankruptcy increase with the size of the bank's assets or liabilities, the cost of bankruptcy also increases.)

The Diagrammatics of the Mean Variance Model

To proceed to the remaining propositions, it will be useful if we recast our model into the mean-standard deviation framework. Assume the bank took all of its net worth and invested it in loans. By choosing its policies efficiently, it can generate a mean standard deviation frontier, illustrated in Figure 4, denoted by RR. We call this the loan opportunity set. It can obtain higher returns by undertaking greater risk, e.g., by charging higher interest rates.

The bank has two additional alternatives. It can borrow and lend (i.e., buy government bonds) at the safe rate of interest. These opportunities are denoted by the point S along the vertical axis. And it can, by allocating different proportions of its wealth between T-bills and loans, obtain any point on the straight line through S which is tangent to the loan opportunity set, RR, at P.

If there were no reserve requirements, the bank could also borrow at the same cost that it can lend to the government, so that its whole opportunity set is the extension of the straight line through S that is tangent to RR. But so long as there are reserve requirements, and reserves pay a lower interest rate than government T-bills, the bank's total opportunity set appears as in Figure 4. There are four possibilities.

---

18. Actually, in terms of real consumption, government bonds are not safe. Mean-variance analysis can be extended in a straightforward way to situations which the "safe" asset is still risky, so long as it represents the minimum variance asset. See Cass and Stiglitz (1970).
In the first three cases, it is accepting no deposits. It is acting as a closed investment bank. First, it could simply buy government bonds (point S): the bank is, in those circumstances, hardly operating as a bank. It can take some of its funds and invest them in T-bills, and it can lend others out. (These are the points along the locus SP.) It can take its net worth and simply make loans, neither investing in T-bills, nor accepting deposits. These are the points along the loan opportunity locus between P and P'.

The fourth case, where it accepts deposits, and lends, it is acting as a conventional bank; it chooses one of the points along the line PL.

Which it chooses depends on its attitudes towards risk, as well as the trade-offs it perceives. Figure 4 illustrates the four possibilities. If it perceives the return on T-bills as close to the maximum it can obtain on its loan portfolio, it is more likely to be close to the corner solution S; alternatively, if it is very risky averse, so that it requires a large increase in mean to compen-

19. If T-bills were perfectly safe, they would not be at the corner solution unless T-bills yielded the same expected return as the loan portfolio. But because of uncertainties associated with inflation, T-bills are, in real terms, risky, and so it is conceivable that there be a corner solution, even when T-bills yield a lower expected return.
sate it for undertaking additional risk, it is more likely to be near S. If it perceives the expected return to loans as much higher than what it must pay to obtain funds, and it is not very risk averse, then it will borrow a great deal (accept many deposits), and lend a great deal.

Implicitly in our analysis so far we have assumed that the mean and standard deviation of returns per dollar lent are independent of the scale of lending. If, however, there is diminishing returns, then the bank's opportunity set appears as in Figure 5, where we assume that there is a fixed supply of good loan opportunities, and a perfectly elastic supply of second rate lending opportunities. Under these assumptions, there is a kink in the total opportunity set facing the bank.

All of this analysis presumes the bank can choose what interest rate it charges, and get as many loan applicants as it wishes, that is, that there is credit rationing. Figure 6 illustrates what happens in the absence of credit rationing. Then, the point on the mean standard deviation curve that the bank faces is determined by competition. There is a single point, denoted by C. Assume all loan applicants are identical. Then, if it offers competitive terms, the bank can get as many borrowers as it wishes. Its opportunity locus is thus the kinked line SCL.
We can now use this framework to establish our remaining propositions.

**Proposition 2.** A decrease in bank's net worth leads to a decrease in bank lending.

Banks pursue lending to the point where the expected return equals the costs of the funds, including the increased expected bankruptcy costs:

\[ EY_N - \rho = \phi \]

where \( \phi \) is the marginal bankruptcy cost

\[ \phi = \partial cF/\partial N > 0. \]

Under mild restrictions, it can be shown that

\[ \phi_a < 0, \]

that is, if the bank initially has less net worth, then it must borrow more to maintain a given level of loans; but the return on the loans is risky. With the
higher amount "borrowed" by the bank, there is a higher probability that

\[ Y \leq B(1 + \rho), \]

that is, that it will not be able to pay back its depositors. Thus, reduced \( a_t \) increases (at a fixed \( N \)) the default probability; and under suitable restrictions, this also increases the marginal default probability. The effect of this on lending is illustrated in Figure 7a.

![Figure 7a](image)

With a fixed cost of funds, \( \rho \), the sole effect of a reduced \( a_t \) is to increase, at any level of lending activity, the marginal bankruptcy cost, as illustrated in the Figure, leading to a lower level of lending.

Figure 7b illustrates the same results in a mean variance diagram. The point \( X \) on the new mean standard deviation opportunity locus corresponds to the original equilibrium \( E \): at the lower wealth, the standard deviation of terminal wealth is the same if the bank lends out the same amount, but the expected value of terminal wealth is lower. Moreover, the trade-off is the same. With "normal" utility functions, at lower levels of wealth, the bank is willing to undertake less risk, i.e., it moves to the left of the point \( X \).

20. The analysis is similar to Greenwald and Stiglitz (1988a), except now the interest rate paid does not depend on the amount borrowed.
Proposition 3. An increase in reserve requirements leads to reduced lending, under “normal” conditions.

An increase in reserve requirements, at a fixed set of “policies” by the bank (including a given interest rate charged to borrowers) increases the cost of funds to the bank (remember our characterisation of reserve requirements as a tax on the bank). This increases the average bankruptcy probability at any level of borrowing, and under normal conditions, this increases the marginal bankruptcy probability.\(^{21}\) The increase in the marginal probability, at each level of lending, implies that the amount of lending will be reduced. The bank lends up to the point where the marginal (expected) return per dollar lent, \(\mu - \rho\), equals the marginal bankruptcy cost. The former is unchanged, the latter increased. (See Figure 8.)

Again, under “normal” conditions, the adjustment of the various terms of the loan contract will not alter this basic conclusion.

\(^{21}\) See Greenwald and Stiglitz (1988a) for a discussion of the conditions under which this occurs.
Figure 9 illustrates the same results in a mean standard diagram, for a bank which accepts deposits. The higher reserve requirements has both a wealth effect (banks are worse off) and a substitution effect (the net cost of funds are higher) and both of these lead to less lending. Moreover, in Figure 9 we show the bank as changing its interest rate to a higher level, leading to higher mean returns, but also higher risk. The shift in policy induced by the higher reserve requirements leads to a higher marginal (expected) return and a higher marginal bankruptcy probability, at each level of lending. Under normal conditions, these secondary adjustments will not undo the first effect.  

**Proposition 4.** Under normal conditions, an increase in the rate of interest on T-bills leads to less lending for banks which accept deposits.

For such banks, the analysis is exactly the same as an increase in the reserve requirement. Note that for banks which do not accept deposits, an increase in the interest rate paid on T-bills has a positive wealth effect and a negative substitution effect, so the net effect on lending activity is ambiguous.

Finally, we need to analyse the consequences of a change in risk (or more generally, a change in the loan portfolio confronting the bank).

---

22. Indeed, they normally reinforce the contraction in loan supply. Note in Figure 9 that $E_1$ could be close to $p'_1$, in which case the bank would be making almost no loans other than out of its own portfolio (i.e., it would accept no deposits).
**Proposition 5.** A mean preserving increase in risk means that, at any given level of loan activity, the risk of default is increased, and hence the marginal probability of the bank's going into bankruptcy, at any given level of lending activity, is increased, as illustrated in Figure 10. Lending activity is accordingly reduced.

Figure 11 shows the same result in terms of a mean standard deviation diagram. The loan opportunity locus shifts to the right, and this has both a wealth effect and a substitution effect, both of which serve to reduce lending activity.

Before leaving this section, we note one further set of possibilities: what happens in the case where there are a limited number of lending opportunities, as illustrated earlier in Figure 5? Then, if the bank is lending to all good loan opportunities, and only to good loan opportunities, changes in reserve requirements or interest rates may have no effect on lending. As Figure 12 illustrates, the bank operates at the kink, both before and after the change. Of course, large enough changes in interest rates, reserve requirements, or bank net worth move the bank away from the kink.
Figure 10

Figure 11
IV MONETARY POLICY AND CYCLICAL FLUCTUATIONS

The model presented in the previous section explains both changes over the cycle in the level of lending activity and why monetary policy may work, even in the absence of seignorage (which, because of our assumptions concerning competition, is totally absent).

Reduced Lending Activity in a Recession

Lending activity can be adversely affected both by changes in expectations (concerning the mean return or risk of lending activity) as well as changes in firms' net worth. In a recession, both tend to move in the same direction. Thus, the 1991-1992 recession in the United States has been marked by significant decreases in bank net worth, as a result of high rates of default on real estate loans, as well as a concomitant view of increased risk associated with further lending activity. (Of course, at a macroeconomic level, there is a feedback — the depressed lending activity leads to a lower level of investment, lower levels of GNP, and higher default rates.)

While there is little doubt that as the economy goes into a recession, expected returns from loans decrease and risk increases, there are several distinct reasons for this deterioration in the quality of the loan portfolio,
which it may be useful to isolate. The first is simply the increased uncertainty in the environment, the consequences of which were the subject of Proposition 5. Loans represent a particular form of risk sharing, where the lender does not participate in the upscale risk, but bears fully the risk of loss. And accordingly, when risk becomes greater (as reflected in a mean preserving spread), the expected return on loans decreases. This "financial fragility" has previously been noted in the case of risk neutral investors. Figure 13 repeats Figure 3, showing the expected return as a function of the interest rate charged. Beyond a certain point, higher interest rates lead to lower expected returns. We then show that increased uncertainty shifts the curve down, placing it below the opportunity cost of capital, the expected return that could be obtained say on T-bills. In that case, lending would dry up.

The changed net worth of borrowers has mixed affects on the expected return to the bank. On the one hand, risk averse borrowers will take greater care; on the other hand, they have less cushion to protect themselves; normally, we might expect that the net effect is a slight increase in bankruptcy probability. Perhaps more important, however, is the fact that when net worth is sufficiently low, with limited liability, firms become risk lovers. (This is the phenomenon made all too familiar by the S & L debacle in the United
States. In a well regulated banking system, banks whose low net worth would lead them to act in a risk loving manner would be shut down, and hence we ignore this possibility here.) In that case, the actions taken by borrowers will adversely affect both the mean return and risk faced by lenders. Moreover, not only are there more of these risk loving borrowers, but while the risk averse borrowers contract their lending activity, risk loving borrowers expand it, so there is a further adverse change in the mix of loan applicants.

**Monetary Policy**

We have seen how monetary policy works both through reserve requirements and open market operations. Reserve requirements are like a tax on deposits: changes in reserve requirements make obtaining deposits, and thus lending, less attractive. Open market operations change the supply of reserves; equilibrating the demand and supply of reserves necessitates changes in the interest rate, but have direct effects on the magnitude of lending activity.

The model also explains the circumstances in which monetary policy is relatively (or totally) ineffective. If, for instance, there are sharply diminishing returns in the loan portfolio — further lending yields lower expected returns and/or higher expected risks — then lowering reserve requirements (lowering the tax on deposits) may result in little extra lending activity. If the rate of interest on deposits has been driven to zero, then there is no “tax” on deposits, and banks would be willing to hold an indefinite amount of them — without engaging in any lending activity.\(^{23}\)

The problem is that as the economy goes into a recession, bank lending would normally decrease, and significantly so. Thus, monetary policy has to push hard simply to “stay still”. The fact that the money supply decreases in a recession does not necessarily reflect misguided monetary policy. The money supply, it will be recalled, is nothing more than the other side of the banks’ balance sheet. Reductions in lending activity will be associated with reductions in money supply. Monetary authorities may be doing a great deal — preventing lending activity (and money supply) from falling further than it would otherwise have. Doing more may not be possible.

In the perspective of this paper, then, monetary authorities affect macroeconomic activity (when they do so) not so much (or at least only) through the interest rates (or other credit terms) at which they make loans available as through the availability of credit.

---

\(^{23}\) In the “traditional” banking model, in which interest rates on deposits are fixed at zero, open market operations may become ineffective long before interest rates fall to zero. In that model, banks divide their total investible assets (deposits plus net worth) between loans and Treasury bills. Open market operations may have little effect on lending activity, if loans are viewed to have lower risk adjusted yields than Treasury bills.
Resuscitating the IS-LM Framework

In Greenwald-Stiglitz (1990b), we have shown how these insights can be incorporated into a modified IS-LM framework, consisting, like the traditional model, of two equations. Earlier, we explained why the traditional LM money market equilibrium condition is unpersuasive. We replace it with a "financial market equilibrium condition", which, in reduced form, gives the equilibrium level of the (real) interest rate on Treasury bills (in our idealised model, this equals the rate of interest on bank deposits) to the level of income. Open market operations and changes in reserve requirements shift the LM curve, in a way analogous to that of the standard theory. As in standard monetarist doctrines, the magnitude of lending activity, given the supply of outstanding reserves, is fixed, regardless of interest rates charged as long as deposit rates are non-negative.

The other equation is the goods market equilibrium curve, the IS curve; and it is this curve which is most altered. As in traditional Keynesian theory, neither investment nor savings is highly interest elastic, so the curve is relatively flat. But more germane to the current discussion of monetary policy is the result that some monetary policies, such as changes in reserve requirements, affect banks' willingness to lend at any market rate of interest; accordingly, they shift the IS curve directly. Investment is affected, and not just through the effect on interest rates.

Portfolio Theory

In many ways, the theory we have developed has some similarities to the portfolio approach of Tobin, but there are some important differences. Tobin stressed focusing on financial markets as a whole, not just on the money rate of interest; and he particularly focused on the price of equity (Tobin's q). When the price of equity exceeded the cost of producing capital goods (that is, market values exceeded the value of the underlying assets, or more precisely, marginal market values, the increase in market values from investment exceeded the cost of the investment) then firms would invest.

While the theory has some superficial plausibility — surely firms should invest more if they can increase the value of the shares of the original shareholders by doing so — it takes little more cognisance of the nature of financial markets than does the neo-classical model. The fact of the matter, as we noted earlier, is that firms raise little of the capital that they use for financing new investment through equity issues, and for good reason: new equity issues have a negative effect on market value. Our colleague Bob Hall once

24. As described here, the model is incomplete; we need to add a wage determination equation. In Greenwald-Stiglitz (1990b) we show how this can be done.
put the matter graphically by pointing out that a year or so ago, the *Wall Street Journal* had finally got "it" right: they had separated the financial section from the business section. To be sure, there are links, but the links are different, and far weaker, than the Tobin theory envisaged.

The basic problem is that managers of firms do not look to the stock market for judgements about the value of investment opportunities. They do not believe that the marginal investor, the dentist in Peoria Illinois or the retired doctor living in Miami, has more insight into the future of the steel or electrical industry than they do. And, in any case, the information provided by the stock market is too gross. The steel company needs to know what kind of steel mill to build, and the stock market cannot provide that kind of information.

In empirical tests, the q-model has, for the most part, performed only mediocely. Supporters are quick to explain that the empirical data do not easily enable us to see "marginal" q's, but there have been serious attempts to get around this problem, and these too have not fared well.

But even had they done better, the observed success may be little more than a spurious correlation, providing little insight into causal structures: events which make the market more optimistic increase q and, at the same time, are likely to make managers more optimistic about investing. It is not that the high q induced the higher level of investment.

The distinction is important, because there may be changes in economic policy which change q, at least in the short run, without at the same time changing long-run investment prospects, and we might not expect such changes to have much effect on investment.25

V CONCLUDING REMARKS

This paper is part of our more extended research programme exploring how imperfections of the capital market — imperfections which can be understood in terms of theories of imperfect information — lead to fluctuations in economic activity. If our findings seem so "commonsensical" and "unsurprising" we find it a virtue, not a vice.

Our findings are of considerable similarity to those of Keynes. And yet there are fundamental differences. Keynes focused on household demand for money; a horizontal demand curve for money (the liquidity trap) was what

25. There are further problems with the theory. In the standard portfolio model, open market operations work because of an increase in the supply of money offset by a decrease in the supply of T-bills. But with money yielding a return equal to that of T-bills, and with money being equally safe (government guarantees), from households' perspectives, these are perfect substitutes.
prevented interest rates from being driven still lower.

We have shifted the focus to the other side of the banks’ balance sheet, to credit rather than demand deposits.26 We have argued that monetary policy may be ineffective, not because households are willing to hold whatever money the banking system creates, but because banks are unwilling to lend. Monetary authorities have only limited powers to induce them to do so. If they cannot induce banks to lend, they have only limited powers to stimulate the economy. And we have argued that the effects of monetary policy are, at most, only partially mediated through changes in the interest rate.

REFERENCES


26. For earlier discussions of this perspective, see Blinder and Stiglitz (1983).


STIGLITZ, J.E., 1982. "Information and Capital Markets", in William F. Sharpe and


