In related work Shleifer (1986) has also shown how, if new products lead to temporary rents, firms may want to introduce them at times of high demand, leading to bunching in the introduction of new products and implementation cycles. These cycles emerge even though the rate at which new products are discovered is constant.

Like the Diamond model, the model presented above captures an interaction between firms that is surely part of reality. Whether such interactions are quantitatively sufficient to create multiple equilibria is open to question. However, even if they do not generate multiple equilibria, these models all suggest reasons why an economy in a recession or depression may not easily return to high levels of activity. In terms of the models of chapter 8 which emphasize the role of price and wage decisions in the adjustment process, they all suggest reasons why firms may see their costs increase as output decreases and may not want to cut prices given wages. One of the goals of current research is to assess the relative empirical importance of the many mechanisms presented in this section.

9.6 Financial Markets and Credit Rationing

Thus far in this chapter we have concentrated on behavior in the goods and labor markets, seeking to account for and to appraise the apparent rigidity of the real wage and other relative prices. A related set of issues arises in the capital markets, where monetary policy actions that have only small effects on interest rates appear to have significant impacts on the economy. We shift our attention now to financial markets and to credit rationing.

Despite the complexity and sophistication of the financial markets, they are typically represented in macroeconomic models by only two variables: the money stock and an interest rate. In this respect the financial markets are treated no differently than other complex markets such as the labor market. But there is a recurrent theme in the literature and among market participants that the interest rate alone does not adequately reflect the links between financial markets and the rest of the economy. Rather, it is argued, the availability of credit and the quality of balance sheets are important determinants of the rate of investment.

Further it is often argued that the money stock is not a key quantity in the determination of the price level and output, in part because it is endogenous and in part because the financial system is sufficiently flexible to generate as much inside money as might be needed to finance any given level of activity.

These views were emphasized by Gurley and Shaw (1960) and the Radcliffe Committee (1959). They have recently been revived by, among others, Bernanke and Gertler (1987), Blinder and Stiglitz (1983), and Greenwald and Stiglitz (1988), who emphasize the role of credit in the business cycle, and particularly in the transmission of monetary policy to the economy.[58]

The recent literature builds on the theory of imperfect information. The basic argument is that the capital markets not only intermediate in a mechanical way between savers and investors but in addition deal with a variety of problems that arise from asymmetric information about investment projects between borrower and lender. These informational problems both shape capital market institutions and debt instruments and affect the way in which policy actions are transmitted to the goods markets.

In this section we examine recent theories of the relationship between financial variables and economic activity, starting from the topic of *credit rationing*.[59] If credit is rationed, then it is possible that the interest rate is not a reliable indicator of the impact of financial variables on aggregate demand. It is quite likely in that case that quantity variables, such as the amount of credit, have to be looked at in appraising monetary and financial policy.

There are several definitions of credit rationing, all arising from the view and experience of capital market participants that borrowers cannot borrow as much as they would like to even when the markets appear to be operating well. *Type 1 credit rationing* occurs when an individual cannot borrow as much as he or she wants at the going interest rate. *Type 2 credit rationing* occurs when, among identical borrowers, some who wish to borrow are able to do so, while others cannot (Keeton 1979). Note that this notion is very close to definitions of involuntary unemployment.

Credit rationing is easy to understand when there are interest rate ceilings, for instance, usury laws. Although usury laws are not uncommon, we will not concentrate on them here.[60] Rather, we develop a theory of credit rationing that depends on asymmetric information between borrowers and lenders.

Two main reasons have been advanced for lenders to ration credit rather than raise interest rates to clear markets:

1. *Moral hazard.* When the contract between lender and borrower is a debt contract that allows for bankruptcy, the lender increases the incentive of the borrower to undertake risky investments by raising the interest rate. The increased risk of bankruptcy may actually reduce the lender's expected return when the interest rate rises. This would not be a problem if the lender could observe and control the type of project undertaken by the borrower.

2. *Adverse selection*. Similarly, again assuming that the contract between lender and borrower is a debt contract, lenders may prefer to ration credit rather than to raise the interest rate because more risk averse individuals drop out of the borrowing pool as the interest rate rises. The less risk averse the borrower, the more likely is the borrower to choose risky projects that increase the chance of bankruptcy. This problem would not occur if the lender had full information about the type of project to be undertaken by the borrower.

We now develop a simple model of credit rationing, due to Keeton (1979) and Stiglitz and Weiss (1981).[61] The model raises several major questions, among them whether the phenomenon of credit rationing is an artifact of the restricted form of contract studied, whether credit rationing implies some form of market failure, and whether the existence of credit rationing implies that interest rates are unreliable indicators of the effects of monetary policy actions on the economy. We turn to those and related questions after developing the model.

A Model of Credit Rationing

There is a continuum of entrepreneurs, each of whom has a project that requires an initial investment of K and is indivisible. Each entrepreneur has an endowment of $W < K$ and therefore has to borrow to invest.

All projects yield the same expected return, R, but they differ in risk. For simplicity suppose that projects either succeed, yielding R_i^s, where i is the index of the project, or fail, yielding the common value R^f, which could be zero. The probability of success is p_i. The relation between p_i and R_i^s implied by the assumption that the expected return is the same across projects is therefore

$$p_i R_i^s + (1 - p_i) R^f = R, \qquad \text{for all } i. \tag{46}$$

The distribution of p_i across entrepreneurs is characterized by a density function $g(p_i)$.

Financial institutions, banks for short, make loans to entrepreneurs. Entrepreneurs use their own wealth for self-finance to the maximum extent possible and need to borrow the amount $K - W = B$ in order to undertake a project. The loans are of a standard debt form, on which the borrower pays the specified amount $(1 + r)B$ if he is able to, but in the event of bankruptcy, which is assumed to occur if the project fails, he pays only the actual available return R^f. It is assumed that

$$R_i^s > (1 + r)B > R^f, \qquad \text{for all } i. \tag{47}$$

The key asymmetry of information is that though the entrepreneur knows his probability of success, the bank does not. Further, in the absence of mechanisms to sort individuals into probability classes, the bank potentially makes loans to all who are willing to borrow at the posted rate. If it should decide to ration credit, it cannot do so in a way that discriminates high-risk from low-risk borrowers among those willing to borrow.

Assume that both the bank and the entrepreneur are risk neutral. The expected return to the investor is

$$E(\pi_i) = p_i[R_i^s - (1 + r)B]. \tag{48}$$

The expected payoff to the bank that makes the loan is

$$E(\pi_b) = (1 + r)B \int_0^p p_i g(p_i)\, dp_i + R^f \int_0^p (1 - p_i) g(p_i)\, dp_i, \tag{49}$$

where p is the cutoff probability at which customers come to the bank for loans, to be determined below.

Now consider an entrepreneur deciding whether to borrow. A key feature of the payoff to the investor is that it is decreasing in the probability of success, p_i. (Remember that R is the same across projects so that a lower p_i implies a higher R_i^s.) To see this, substitute from (46) into (48) to obtain

$$E(\pi_i) = R - R^f - p_i[(1 + r)B - R^f],$$

which from (47) is decreasing in p_i. Thus high-risk investors are willing to pay more for a loan. This is the basic source of the credit-rationing result. It clearly depends on the fact that the contract between the borrower and the bank is a debt contract, and we return below to the reasons why the contract may take that form.

Assume that investors have the alternative of holding their wealth, W, in a safe asset that yields a rate of return ρ. They will therefore want to borrow so long as

$$E(\pi_i) \geqslant (1 + \rho)W. \tag{50}$$

Given (50) and the definition of $E(\pi_i)$ above, the higher the interest rate, r, the riskier is the marginal project, that project for which the entrepreneur is indifferent between undertaking the investment project and putting his wealth into the safe asset. This implies that $dp/dr < 0$; that is, the probability of success of the marginal project declines as the interest rate increases.

Now consider the impact of an increase in the loan interest rate on the expected return of any bank that is making loans. Differentiating (49) with respect to r, we obtain

$$\frac{dE(\pi_b)}{dr} = B\int_0^p p_i g(p_i)\,dp_i + \left(\frac{dp}{dr}\right)[(1+r)Bpg(p) + R^f(1-p)g(p)]. \quad (51)$$

The first term on the right-hand side reflects the higher repayments by those who repay. The second term reflects the deterioration in the quality of the pool of applicants. This second term is negative, and it is accordingly possible that an increase in the interest rate charged by the bank reduces its expected profits. Whether this happens depends on the properties of the density function. The bank's profits are maximized at the interest rate at which $dE(\pi_b)/dr = 0$.

In the fourth quadrant of figure 9.10, we show one possible relationship between the rate of interest charged to lenders, r, and the expected return to an individual bank, denoted ρ_b, equal to $E(\pi_b)/B$. Examples of density functions that produce such a relationship between the expected return of the bank and the interest rate to lenders are given in Stiglitz and Weiss (1981) and English (1986); the key to producing a negative relationship is that a small increase in the interest rate drives a large number of relatively safe borrowers out of the market.

The remainder of the diagram shows how credit rationing may arise.[62] The demand for credit is shown by the downward-sloping L_d curve in the first quadrant. The demand for credit is simply

$$B\int_0^p g(p_i)\,dp_i.$$

Its negative slope results from the fact that $dp/dr < 0$. In the third quadrant we show the supply of funds to the bank as a function of ρ_b, assumed here to be increasing. We also assume a direct relationship between the bank's rate of return and the rate it pays on deposits. In the absence of reserve holdings and operating costs, and with a competitive banking system, ρ_b would be equal to the rate of return offered on deposits. If banks hold reserves and have other operating expenses, the rate of return on deposits would be below ρ_b but, in general, an increasing function of ρ_b. The analysis would not change if the elasticity of supply of funds to the bank were infinite.[63]

We can draw the implied "supply" curve in the first quadrant. In the fourth quadrant a given value of r implies a given expected rate of return to the bank, ρ_b. The third quadrant gives the supply of deposits at rate ρ_b. The

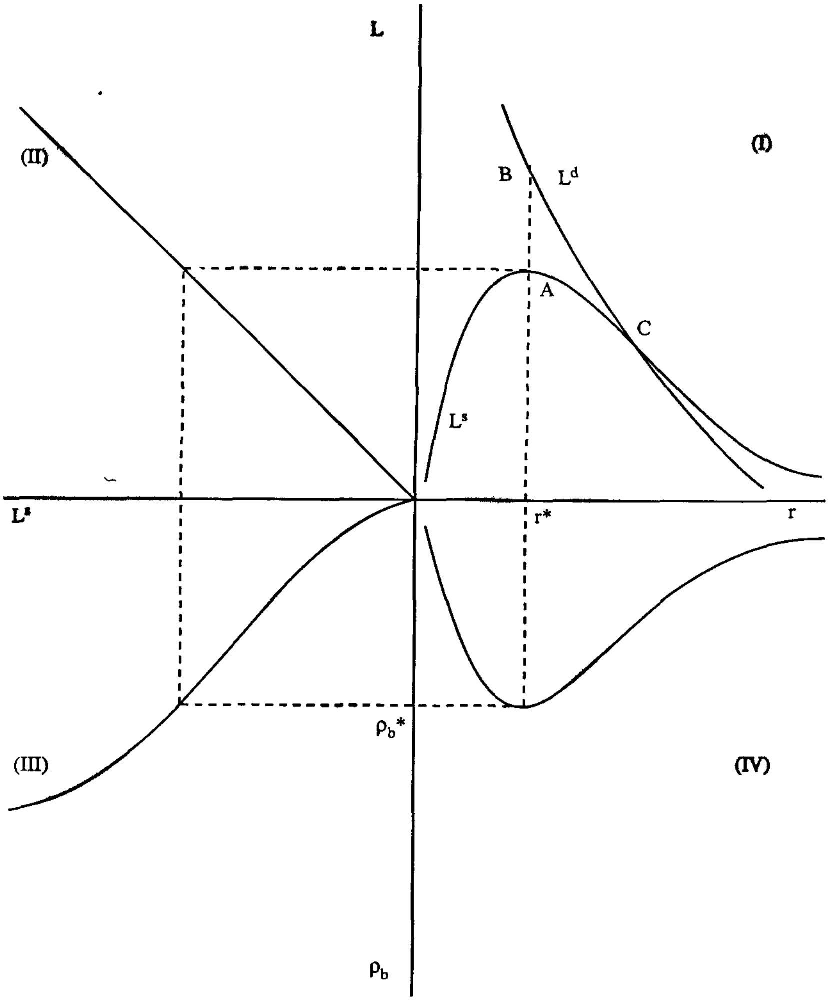

Figure 9.10
Equilibrium credit rationing

45-degree line in the second quadrant allows us to derive the supply of loans in the first quadrant. In the case shown, there is a credit-rationing equilibrium at point A, at the interest rate r^* that maximizes the bank's profits. The demand for loans at that interest rate, shown by point B, exceeds the amount supplied. Credit is therefore rationed. This is type 2 rationing, whereby, among identical projects, some receive financing and others do not—a result of projects having a given minimum size.

There appears to be a standard nonrationed equilibrium at point C. Any bank operating at that point, however, could earn higher profits by cutting its interest rate and returning to point A. It might seem that credit rationing could not persist because banks are earning profits at point A. But nothing in the analysis is inconsistent with the assumption that at point A all expected profits have been competed away; thus point A could be a point of industry equilibrium, implying the possible existence of credit rationing in equilibrium in the setup shown in figure 9.10.

It is also possible to generate credit rationing in a setup in which each investor has several different types of projects. Then, by an analysis similar to that above, we can show the possible existence of a credit-rationing equilibrium in which each investor obtains less financing than he wants at the prevailing interest rate.

Note that in these cases it is not easy to distinguish whether credit rationing results from moral hazard or adverse selection. In both examples there is adverse selection of projects, the selection of riskier projects being responsible for credit rationing. Likewise, in each case credit rationing may be described as arising from moral hazard because it is only the possibility of bankruptcy—willingly entertained by the borrower because of limited liability—that causes rationing.

The Form of Contract

The use of a very restricted form of debt contract in the above model raises the issue of the robustness of the credit-rationing result. One obvious possibility is that the borrower puts up collateral that he forfeits in the event of bankruptcy. If each individual has collateral equal to B, then there is no risk to the lender and no credit rationing. But in this case there would be no need to borrow either.

Could the use of a collateral–interest rate schedule induce borrowers to reveal their type? Stiglitz and Weiss (1981) show that this is not generally the case in a model in which banks set the collateral level. Their proof turns on the assumption that at some point individuals come up against the constraint of their existing wealth, and therefore the above analysis applies again.

Hart (1986) raises the question of why it is possible for the lender to ascertain the amount earned by the borrower in one state of nature (when the payoff is R^f) but not another (R^s). Some progress has been made in answering this question by Gale and Hellwig (1985). They assume that it is costly to monitor the state of nature[64] and then show that in such a case optimal contracts between borrower and lender take the form of debt. By having the borrower make a fixed payment independent of the state of nature for good outcomes, the contract saves on expected monitoring costs. By monitoring when the borrower declares an inability to meet the fixed payment, the contract ensures that the borrower cannot declare an inability to meet the stated payments at will.[65]

Optimality of Credit Rationing

The term rationing automatically creates the impression of nonoptimality. De Meza and Webb (1987) show that if, in the Stiglitz-Weiss model, the supply of funds to the bank is nondecreasing in the rate of return, there is too little investment at the credit-rationed equilibrium. The argument is that since both banks and entrepreneurs are risk neutral, all projects with expected return greater than or equal to the safe interest rate ρ should be undertaken. Thus at the social optimum all projects for which

$$p_i R_i^s + (1 - p_i)R^f \geqslant (1 + \rho)K \tag{52}$$

should be undertaken.

Investors finance through the bank if

$$p_i[R_i^s - (1 + r)B] \geqslant (1 + r)W. \tag{53}$$

Consider now the marginal project for which (53) holds with equality. On any project that it finances, the bank's expected return is

$$E(\pi) = p_i(1 + r)B + (1 - p_i)R^f - (1 + \rho)B.$$

On the marginal project the bank's expected return is

$$E(\pi) = p_i R_i^s + (1 - p_i)R_f - (1 + \rho)K.$$

Suppose now that this is also the socially marginal project, for which (52) holds with equality. Then the bank expects to break even on this project but nevertheless will lose money on all other projects as a result, because they are riskier. Since in an equilibrium the bank expects to make zero or positive profits, it cannot at the margin be financing the marginal socially optimal investment. Therefore investments that should be undertaken are not undertaken.

De Meza and Webb (1987) show that in this model an interest subsidy can restore the first-best allocation. They also show that with a backward-bending supply of funds to the banking system, there will be overinvestment at the credit-rationed equilibrium.

However, the nonoptimality of equilibria in the presence of credit rationing is not general: there are examples in which the allocation with credit rationing is that which would be produced by a central planner with the same information as is available to market participants. Thus the allocation *in a credit-rationed equilibrium is optimal in the model setup by Williamson* (1986); similarly, Keeton (1979) shows in a model with type 1 credit rationing that the allocation is efficient. The point in all of these cases is that credit rationing is an efficient method of preventing overinvestment in risky projects that would otherwise take place because of the lack of information by the lender.

An interesting model in which the allocation of credit is nonoptimal in the absence of credit rationing is presented by Mankiw (1986). The nonoptimality arises from the use of debt contracts, which imply that at equilibrium (where supply equals demand) some safe borrowers who should invest do not do so and some borrowers who should not invest are doing so. Mankiw shows that an increase in the interest rate may destroy the market equilibrium (this is the financial collapse). Because this leads to a nonoptimal allocation of investment, government intervention to prevent market collapse may be socially justified.

The Transmission of Monetary Policy

Interest in credit rationing in the 1950s was motivated by the question whether monetary policy could have powerful effects on the economy if interest rates did not move substantially. At that time monetary policy was constrained by the fear that large interest rate increases would significantly increase the interest burden on the budget. It was accordingly hoped that monetary policy could affect aggregate demand even without having a large impact on interest rates.

The "availability doctrine" argued that because of interest rate ceilings, changes in the quantity of financial assets would affect economic activity even without changes in the interest rate. The question that therefore arises is whether the existence of equilibrium credit rationing implies that monetary policy can have significant impacts on aggregate demand without changing interest rates.

Using figure 9.10, assume that an easing of monetary policy increases the supply of loans to the banks at each deposit interest rate, shifting the locus in the third quadrant to the left. If it does not move the locus in the fourth quadrant, the increase in the supply of funds will shift the L_s curve in the first quadrant up, increasing the quantity of loans offered at each interest rate. The cost of bank loans would be unchanged, and the amount of loans and of investment increased.

If we measured the interest rate by the cost of bank loans, this would show monetary policy affecting the economy without changing the interest rate. The question remains, however, to find the mechanism through which the loan supply curve in the third quadrant moved. English (1986) shows, in a model with a safe asset and credit rationing, that although credit rationing removes a close link between the quantity of bank loans and interest rates charged by banks, the rate of return on the safe asset remains a good indicator of the impact of monetary policy on borrowing.

Stiglitz and Weiss (1981) expand the model developed above to include a variable level of collateral: loan contracts then specify both the level of collateral and an interest rate. The loan interest rate may move either pro- or countercyclically. By including two types of projects, relatively safe and relatively risky, and by allowing productivity shocks to affect the probabilities of success of these projects differentially, they are able to produce a countercyclical pattern in the interest rate paid to bank depositors. The model, however, does not fully specify the links between monetary policy and the availability of funds (in real terms) to the banks.

Bernanke and Gertler (1986) develop a general equilibrium real business cycle model in which collateral plays an important role, thereby opening up a channel through which the quality of balance sheets affects investment and output. In their model investment projects are too large for an individual saver to finance alone. There are nevertheless agency costs associated with external finance so that the more collateral (equivalently one may think of firms' internal financing) that is needed, the greater the deadweight loss associated with the recourse to external finance.[66] Under a general assumption of risk neutrality, the expected return on investment net of agency costs has to be equal to the safe rate of interest, which is itself endogenous. Bernanke and Gertler show, in an overlapping generations model (of the Diamond type studied in chapter 3), that positive productivity shocks that increase output and saving lead to more efficient investment (because collateral is greater), thereby accentuating the effects of the productivity shock relative to a full information economy.

A favorable productivity shock reduces the gap between the safe interest rate and the gross of agency cost expected return on investment; in recessions the safe interest rate may fall while the gross expected return on investment rises. This is a result that is consistent with earlier work by Bernanke (1983), claiming that one of the major propagating mechanisms in the Great Depression was the collapse of financial intermediation that opened the gap between safe and risky rates of interest.

Bernanke and Gertler use their model in addition to discuss the famous Irving Fisher notion of a "debt deflation." Fisher's argument was that deflation increases debt burdens and bankruptcies and therefore reduces investment. If the debt deflation is interpreted as reducing the amount of collateral, the Bernanke-Gertler model can be used to demonstrate the Fisher hypothesis.

Summary

There are now several models in which investment can change without significant movements in interest rates and in which financial variables other than the interest rate affect the rate of investment. In this sense the interest rate may be an inadequate indicator of the thrust of monetary policy.

But these models have not yet reached the point where monetary policy effects themselves are endogenized; at the present state of development they all rely implicitly on some form of nominal rigidity, either on price rigidity so that changes in nominal money affect real money balances and credit or on the existence of nominal bonds, whose value is affected by unexpected inflation or deflation. Once these real effects allow a monetary change to provide an impulse to the system, it is clear that the details of the financial system affect the propagation or transmission mechanism.

9.7 Conclusions

The contrast between the unwieldy and inconclusive presentation of the material in this chapter and the clarity and sense of direction in chapters 2 and 3 is quite obvious. Such is the difference between material which is in large part about 20 years old, thoroughly absorbed and understood, and current research, whose implications and ultimate payoffs cannot yet be fully grasped.

In this and the previous two chapters we have been attempting to provide satisfactory theories to account for the macroeconomic facts outlined in chapter 1—particularly the joint behavior of output and prices, employment and wages, money and output, and aggregate demand and output. Although we have presented many different approaches, they have in common the desire to account for macroeconomic phenomena in a framework in which the motivation and environments of the economic agents and institutions in the model are fully specified. In other words, the new models go beyond the notion common in the 1970s and 1960s of providing a "microeconomic foundation for macroeconomics" and seek rather to provide complete macroeconomic models that are not only analytically coherent but also empirically relevant.

At this stage we cannot discern which model or combination of models will 20 years hence be regarded as absolutely essential to serious macroeconomic theory, as the Ramsey-Sidrauski model and the overlapping generations model are now. However, we can give educated guesses. In the labor markets, notions of efficiency wages have a definite ring of truth. So does monopolistic competition, as a stand-in for imperfect competition, in the goods markets. We believe that the recent work attempting to account for certain features of the financial markets from the viewpoint of asymmetric information is extremely important and that it will be increasingly integrated in complete macroeconomic models. Finally, we are quite sure that nominal rigidities are an important part of any account of macroeconomic fluctuations and that staggering of price and wage decisions is an important element of any complete story.

How those theories must be combined, and whether some unifying principle can be found in the dizzying diversity of explanations, remains to be seen. No doubt some theories that now look promising will turn out to be dead ends, and some that now look moribund or already dead will turn out to be important. How does the process work? To some extent the theories that win are those that are more appealing to our professional standards and prejudices. To a greater extent, they will be the theories that succeed in accounting for the macroeconomic facts as well as the microeconomic evidence.

Problems

1. *Labor supply and efficient contracts.*

Consider a labor market with one firm and one worker. The firm is risk neutral. It operates under constant returns to labor and is subject to multiplicative technological shocks s_t. Thus, if it uses l_t hours of labor, its output is $s_t l_t$. The worker has utility $E[\ln(l_t w_t) \quad l_t^b]$, where l is hours of labor, w is the hourly wage, and b is larger than one. l_t can take any nonnegative value.