Chapter 10
FINANCIAL MARKETS AND FINANCIAL CRISES

The previous two chapters study the behavior of households and firms in partial-equilibrium settings. In Chapter 8, households divide their income between consumption and saving, taking the set of available assets and the distribution of their rates of returns as given. In Chapter 9, firms decide how much investment to undertake, taking the way that future profits are valued as given.

Financial markets are where these saving and investment decisions meet. In the absence of asymmetric information, externalities, and other imperfections, they play a central role in getting the economy to its Arrow-Debreu outcome. The signals sent by asset prices and state-contingent returns are what frame the partial-equilibrium problems that households and firms face. They therefore determine how households allocate their resources among consumption and holdings of various risky assets and what investment projects are undertaken. And it is the interaction of the demand and supply of risky assets that determines their prices. General equilibrium occurs when households and firms are optimizing taking prices as given, and where prices cause asset markets to clear. Section 10.1 presents a model of perfectly functioning financial markets in general equilibrium that shows this interplay between saving and investment decisions.

The main reason that macroeconomists are so interested in financial markets, however, is that they do not appear to function in this idealized way. There are at least four distinct issues related to financial markets that are important to macroeconomics.

The first is whether there are important macroeconomic propagation mechanisms operating through financial markets. With perfect financial markets, asset prices passively summarize all available information. But if there are imperfections in financial markets that cause departures from first-best outcomes, those distortions may change endogenously in response to economic developments. As a result, they can magnify the macroeconomic effects of various types of shocks to the economy.

Sections 10.2 and 10.3 investigate this idea. Section 10.2 presents a microeconomic model of investment in the presence of asymmetric information between outside investors and entrepreneurs and examines the determinants and effects of the resulting distortions. It then shows that there are several channels that cause those distortions to be greater when the economy is weaker. Because the distortions reduce investment, their endogenous response to the state of the economy magnifies the macroeconomic effects of shocks—a mechanism that is known as the *financial accelerator*. Then in Section 10.3, we will examine some microeconomic evidence about the importance of financial-market imperfections to investment.

The second issue concerns whether departures of financial markets from the Arrow-Debreu baseline not only magnify the effects of other disturbances, but can also be an independent source of shocks to the economy. In particular, Section 10.4 is devoted to the issue of possible *excess volatility* of asset prices. In perfect financial markets, the price of any asset is the rational expectation given the available information of the present value of the asset's future payoffs using the stochastic discount factor that arises from agents' marginal utilities of consumption; the price of the asset changes only if there is new information about its payoffs or about the stochastic discount factor. Section 10.4 examines the possibility that this assumption might fail. It shows that the forces pushing asset prices toward fundamentals if they depart are not unlimited, and analyzes several factors that limit their strength. It also describes how movements in asset prices not driven by fundamentals can affect macroeconomic outcomes. Section 10.5 addresses the issue of whether the possibilities described in Section 10.4 are merely hypothetical or whether there is evidence that they are important in practice.

A third macroeconomic subject raised by financial markets and the possibility of financial-market imperfections is financial crises. One might expect that a large financial system in an economy with millions of participants would change smoothly in response to economic developments. In fact, however, financial markets are subject to convulsions not only at the level of individual assets and financial institutions, but at the level of broad swaths of the financial system. One notable episode occurred in the Great Depression, when the economic downturn and repeated waves of panics led to the failure of one-third of U.S. banks. For decades, the conventional wisdom was that such worldwide financial crises were a thing of the past. But in the fall of 2008, this view was proven wrong. Lehman Brothers, a major investment bank, declared bankruptcy in September. In the aftermath, equity prices fell by more than 25 percent in just four weeks; spreads between interest rates on conventional but slightly risky loans and those on the safest and most liquid assets skyrocketed; many borrowers were unable to borrow at any interest rate; and economies around the world went into severe recessions.

Financial crises are the subject of Sections 10.6 through 10.8. The first of these presents the classic Diamond–Dybvig model of the possibility of a self-fulfilling run on a financial institution that would otherwise be solvent. The second addresses the issue of how financial market disruptions and failures can spread among financial institutions. And the third discusses some microeconomic evidence that sheds light on the macroeconomic effects of crises.

The final issue concerns the social value of financial markets. Private marginal products in financial markets can be either bigger or smaller than social marginal products. For example, someone who channels funds from small savers to a poor entrepreneur with the potential to become the next Thomas Edison or Steve Jobs will probably capture only a tiny part of the social value of the resulting inventions. On the other hand, someone who makes an enormous profit by buying an undervalued asset whose price would otherwise have risen a few seconds later probably has almost no effect on any actual investment decisions, and so has negligible social product.

Although this issue is fascinating and important, we will not pursue it. McKinnon (1973) and others argue that the financial system has important effects on overall investment and on the quality of the investment projects that are undertaken, and thus on economies' growth over extended periods. But the development of the financial system may be a by-product rather than a cause of growth; and factors that lead to the development of the financial system may affect growth directly. As a result, this argument is difficult to test. Nonetheless, there is at least suggestive evidence that financial development is important to growth (for example, Jayaratne and Strahan, 1996; Levine and Zervos, 1998; Rajan and Zingales, 1998; Levine, 2005). Likewise, Banerjee and Newman (1993) and Buera, Kaboski, and Shin (2011) argue that financial-market imperfections can lead to large inefficiencies in both human- and physical-capital investment, and that this misallocation has large effects on development.

There is even less evidence about whether too many resources are devoted to the financial sector in advanced economies. Budish, Cramton, and Shim (2015) present compelling evidence that efforts to increase trading speed are close to pure rent-seeking with little social value. But whether this description fits with much of how other resources in the financial sector are used is not clear. Philippon and Reshef (2012) present evidence that compensation in finance in the United States in recent decades is puzzlingly high, and Philippon (2015) finds that the large improvements in information technology and other types of technological progress do not appear to have increased the efficiency of the U.S. financial sector over the past hundred years. But again, this does not come close to resolving the question of whether too many resources are devoted to the financial sector.

10.1 A Model of Perfect Financial Markets

This section presents a model of perfectly functioning financial markets to show how the interaction of consumer preferences and the set of possible investments determine what investment projects are undertaken and how claims on the projects' output are valued.

Assumptions and Equilibrium Conditions

The economy lasts for two periods. A representative household has an endowment E of the economy's sole good in period 1 and no endowment in period 2. It maximizes the expected value of its lifetime utility, which is given by

$$V = U(C_1) + \beta U(C_2), \qquad \beta > 0, \qquad U'(\bullet) > 0, \qquad U''(\bullet) < 0, \tag{10.1}$$

where C_t is the household's consumption in period t.

All period-2 output comes from investments undertaken in period 1. There are N possible investment projects. The output of each project is potentially uncertain. Specifically, there are S possible states of the world in period 2. If quantity K_i of period-1 output is devoted to project i, it produces $R_{is}K_i$ in period 2 in state s (where $R_{is} \geq 0$ for all i and s). We let π_s denote the probability of state s occurring; the π_s's satisfy $\pi_s \geq 0$ and $\sum_{s=1}^{S} \pi_s = 1$. The K_i's cannot be negative. It is convenient to think of each investment project as being undertaken by a distinct firm. Finally, the economy is perfectly competitive: households and firms are price-takers.

It is straightforward to write down the conditions that characterize the equilibrium of this economy. Because there are complete markets and no imperfections, we can describe the equilibrium in terms of Arrow-Debreu commodities—that is, claims on period-2 output in the various states of the world. Specifically, let q_s be the price, in units of period-1 output, of a claim on one unit of perod-2 output in state s. Then equilibrium is a set of prices, $\{q_s\}$, investment decisions, $\{K_i\}$, and consumption decisions, C_1 and $\{C_2^s\}$, with three properties.

First, households must be maximizing their utility subject to their budget constraint. The budget constraint is

$$C_1 + \sum_{s=1}^{S} q_s C_2^s = E. \tag{10.2}$$

Utility maximization requires that reducing C_1 by a small amount and using the savings to increase C_2^s does not affect lifetime utility. This yields the Euler equation.

$$U'(C_1) = \frac{1}{q_s} \pi_s \beta U'(C_2^s) \qquad \text{for all } s. \tag{10.3}$$

We can rearrange this as

$$q_s = \pi_s \beta \frac{U'(C_2^s)}{U'(C_1)} \qquad \text{for all } s. \tag{10.4}$$

That is, in equilibrium the price of a claim on output in state s equals the product of the probability that the state occurs and the marginal utility of consumption in that state relative to consumption today.

Second, there must be no unexploited profit opportunities. The cost of investing marginally more in project i, in terms of period-1 consumption, is just 1. The payoff is the revenues from selling the state-contingent output, which is $\sum_{s=1}^{S} q_s R_{is}$. If a strictly positive amount is invested in the project, the payoff to investing marginally more must equal the cost. And if nothing is invested in the project, the payoff from the first unit of investment must be less than or equal to the cost. Thus we have

$$\sum_{s=1}^{S} q_s R_{is} \begin{cases} = 1 & \text{if } K_i > 0 \\ \leq 1 & \text{if } K_i = 0 \qquad \text{for all } i. \end{cases} \tag{10.5}$$

Notice that since there is a full set of Arrow-Debreu commodities, there is no risk in undertaking the project: although the amount that the project produces depends on the state, claims on output in all states are sold in period 1.

Finally, markets must clear. The market-clearing condition in period 1 is

$$C_1 + \sum_{i=1}^{N} K_i = E. \tag{10.6}$$

And the market-clearing condition for claims on period-2 output in state s is

$$\sum_{i=1}^{N} K_i R_{is} = C_2^s \qquad \text{for all } s. \tag{10.7}$$

The number of equilibrium conditions is 1 (from [10.2]), plus S (from [10.3] or [10.4]), plus N (from [10.5]), plus 1 (from [10.6]), plus S (from [10.7]). The unknowns are the S q_s's, the S C_2's, the N K_i's, and C_1. The number of equations exceeds the number of unknowns by 1 because of Walras's law.

Discussion

From firms' perspective, this model is little different from the partial-equilibrium model of investment we studied in Chapter 9, particularly the model of investment under uncertainty in Section 9.7. And from households' perspective, the model is similar to the partial-equilibrium model of consumption in the presence of risky assets in Section 8.5. But because both investment and consumption are endogenous in the current model, the marginal utility of consumption in different states, and hence the payoff to investment projects in different states, is now endogenous.

The only assets with net supplies that are strictly positive are claims on the output of the investment projects that are undertaken. But although the net supplies of any other potential financial assets are zero, one can still think of markets where some agents can sell them and others buy them. The price of any asset (including ones without positive net supplies) depends on the pricing kernel of this economy, which is determined by the marginal utility of consumption in different states. That is, the price of an asset with payoff in state s of x_s is $\sum_{s=1}^{S} q_s x_s$. Two potentially interesting assets are debt and insurance. The price of riskless debt—that is, an asset that pays 1 unit regardless of the state—is $\sum_{s=1}^{S} q_s$. The price of insurance against state s occurring—that is, an asset that pays 1 unit in state s and 0 otherwise—is q_s. Of course, since all households are the same, in equilibrium we will not observe some agents selling these assets and others buying them. But with heterogeneity in preferences or income, we would.

Also, notice that trade in financial assets can get the economy to the Walrasian outcome without there literally being Arrow-Debreu commodities with all transactions taking place at the beginning of time. In the model, where agents are homogeneous, the Arrow-Debreu allocation can be achieved through equity markets where claims on firms' output are traded. With heterogeneous agents, additional assets, such as insurance contracts, would also be needed.

Crucially, because all markets are perfectly competitive, information is symmetric, and there are no externalities, the equilibrium is Pareto efficient. If we enriched the model to allow for the arrival of new information (for example, about the probabilities of the different states or the returns to investment projects in different states), there would be changes in asset prices, but they would be efficient. And the only reason for there to be sudden large changes in prices would be the sudden arrival of major news.

There are numerous possible extensions that would not affect these central features of the model. Examples include additional periods, heterogeneity among households, adjustment costs in investment, and a role for other inputs into production (most obviously, labor supplied by households). None of these would alter the messages that financial markets are where households and firms meet to efficiently share risk and determine the level and composition of investment. Rather than pursuing those extensions, the rest of the chapter turns to settings where financial markets have more significant consequences.

10.2 Agency Costs and the Financial Accelerator

Introduction

In the models of Chapter 9 and Sections 10.1, all parties are equally well informed, and so financial markets function efficiently. Potential investments are valued according to their state-contingent payoffs using the prevailing stochastic discount factor. As a result, they are undertaken if their value exceeds the cost of acquiring and installing the necessary capital.

In fact, however, firms are much better informed than potential outside investors about their investment projects. Outside financing must ultimately come from individuals. These individuals usually have little contact with the firm and little expertise concerning the firm's activities. In addition, their stakes in the firm are usually low enough that their incentive to acquire relevant information is small.

Because of these problems, institutions such as banks, mutual funds, and bond-rating agencies that specialize in acquiring and transmitting information play central roles in financial markets. But even they can be much less informed than the firms or individuals in whom they are investing their funds. The issuer of a credit card, for example, is usually much less informed than the holder of the card about the holder's financial circumstances and spending habits. In addition, the presence of intermediaries between the ultimate investors and firms means that there is a two-level problem of asymmetric information: there is asymmetric information not just between the intermediaries and the firms, but also between the individuals and the intermediaries (Diamond, 1984).

Asymmetric information creates *agency problems* between investors and firms. Some of the risk in the payoff to investment is usually borne by the investors rather than by the firm; this occurs, for example, in any situation where there is a possibility that the firm may go bankrupt. When this is the case, the firm can change its behavior to take advantage of its superior information. It can only borrow if it knows that its project is particularly risky, for example, or it can choose a high-risk strategy over a low-risk one even if this reduces overall expected returns. Thus asymmetric information can distort investment choices away from the most efficient projects. In addition, asymmetric information can lead the investors to expend resources monitoring the firm's activities, and the managers or entrepreneurs running the firm to devote less than the socially optimal amount of effort to the firm. Thus again, asymmetric information imposes costs.

This section presents a simple model of asymmetric information and the resulting agency problems and discusses some of their effects. We will find that when there is asymmetric information, investment depends on more than just interest rates and profitability; such factors as investors' ability to monitor firms and firms' ability to finance their investment using internal funds also matter. We will also see that asymmetric information changes how interest rates and profitability affect investment and that it magnifies the effects of shocks to the economy.

Assumptions

An entrepreneur has the opportunity to undertake a project that requires 1 unit of resources. The entrepreneur has wealth of W, which is less than 1.

Thus, he or she must obtain $1 - W$ units of outside financing to undertake the project. If the project is undertaken, it has an expected output of γ, which is positive. γ is heterogeneous across entrepreneurs and is publicly observable. Actual output can differ from expected output, however; specifically, the actual output of a project with an expected output of γ is distributed uniformly on $[0,2\gamma]$. Since the entrepreneur's wealth is all invested in the project, his or her payment to the outside investors cannot exceed the project's output. This limit on the amount that the entrepreneur can pay to outside investors means that the investors must bear some of the project's risk.

To keep things simple, we assume that the entrepreneur and outside investors are risk-neutral, and that there is a technology with no risk or asymmetric information that yields a rate of return of r for sure. We also assume that the outside investors are competitive. These assumptions have several implications. First, the project is socially desirable if and only if the expected rate of return is greater than r; that is, the requirement for a social planner to want the project to be undertaken is $\gamma > 1+r$. Second, because the entrepreneur can invest at the risk-free rate, he or she undertakes the project if the difference between γ and the expected payments to the outside investors is greater than $(1+r)W$. And third, competition implies that in equilibrium, outside investors' expected rate of return on any financing they provide to the entrepreneur is r.

The key assumption of the model is that entrepreneurs are better informed than outside investors about their projects' actual output. Specifically, an entrepreneur observes his or her output costlessly; an outside investor, however, must pay a cost c to observe output. c is assumed to be positive; for convenience, it is also assumed to be less than expected output, γ.

This type of asymmetric information is known as *costly state verification* (Townsend, 1979). In studying it, we will see not only how asymmetric information affects investment outcomes, but also how it shapes financial contracts. There are other types of information asymmetries, such as asymmetric information about the riskiness of projects or about entrepreneurs' actions, that may be more important than costly state verification in practice. However, they have broadly similar implications concerning distortions and the amplification of shocks, so it is instructive to study these issues through the lens of the simpler costly-state-verification model.

The Equilibrium under Symmetric Information

In the absence of the cost of observing the project's output, the equilibrium is straightforward. Entrepreneurs whose projects have an expected payoff that exceeds $1+r$ obtain financing and undertake their projects; entrepreneurs whose projects have an expected output less than $1+r$ do not. For the projects that are undertaken, the contract between the entrepreneur and the outside investors provides the investors with expected payments of $(1 - W)(1 + r)$. There are many contracts that do this. One example is a contract that gives to investors the fraction $(1 - W)(1 + r)/\gamma$ of whatever output turns out to be. Since expected output is γ, this yields an expected payment of $(1 - W)(1 + r)$. The entrepreneur's expected income is then $\gamma - (1 - W)(1 + r)$, which equals $W(1 + r) + \gamma - (1 + r)$. Since γ exceeds $1 + r$ by assumption, this is greater than $W(1 + r)$. Thus the entrepreneur is made better off by undertaking the project.

The Form of the Contract under Asymmetric Information

Now consider the case where it is costly for outside investors to observe a project's output. In addition, assume that each outsider's wealth is greater than $1 - W$. Thus we can focus on the case where, in equilibrium, each project has only a single outside investor. This allows us to avoid dealing with the complications that arise when there is more than one outside investor who may want to observe a project's output.

Since outside investors are risk-neutral and competitive, an entrepreneur's expected payment to the investor must equal $(1+r)(1-W)$ plus the investor's expected spending on verifying output. The entrepreneur's expected income equals the project's expected output, which is exogenous, minus the expected payment to the investor. Thus the optimal contract is the one that minimizes the fraction of the time that the investor verifies output while providing the outside investor with the required rate of return.

Given our assumptions, the contract that accomplishes this takes a simple form. If the payoff to the project exceeds some critical level D, then the entrepreneur pays the investor D and the investor does not verify output. But if the payoff is less than D, the investor pays the verification cost and takes all of output. Thus the contract is a debt contract. The entrepreneur borrows $1 - W$ and promises to pay back D if that is possible. If the entrepreneur's output exceeds the amount that is due, he or she pays off the loan and keeps the surplus. And if the entrepreneur cannot make the required payment, all of his or her resources go to the lender. This payment function is shown in Figure 10.1.

The argument that the optimal contract takes this form has several steps. First, when the investor does not verify output, the payment cannot depend on actual output. To see this, suppose that the payment is supposed to be Q_1 when output is Y_1 and Q_2 when output is Y_2, with $Q_2 > Q_1$, and that the investor does not verify output in either of these cases. Since the investor does not know output, when output is Y_2 the entrepreneur pretends that it is Y_1, and therefore pays Q_1. Thus the contract cannot make the payment when output is Y_2 exceed the payment when it is Y_1.

Second, and similarly, the payment with verification can never exceed the payment without verification, D; otherwise the entrepreneur always

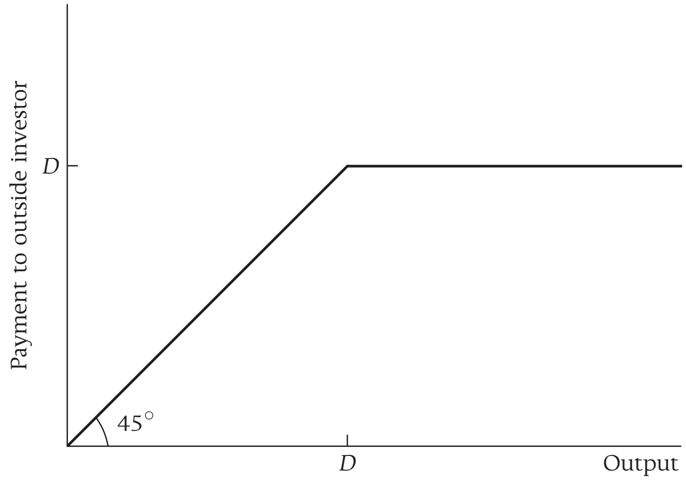

FIGURE 10.1 The form of the optimal payment function

pretends that output is not equal to the values of output that yield a payment greater than D. In addition, the payment with verification cannot equal D; otherwise it is possible to reduce expected expenditures on verification by not verifying whenever the entrepreneur pays D.

Third, the payment is D whenever output exceeds D. To see this, note that if the payment is ever less than D when output is greater than D, it is possible to increase the investor's expected receipts and reduce expected verification costs by changing the payment to D for these levels of output; as a result, it is possible to construct a more efficient contract.

Fourth, the entrepreneur cannot pay D if output is less than D. Thus in these cases the investor must verify output.

Finally, if the payment is less than all of output when output is less than D, increasing the payment in these situations raises the investor's expected receipts without changing expected verification costs. But this means that it is possible to reduce D, and thus to save on verification costs.

Together, these facts imply that the optimal contract is a debt contract.[1]

---

[1] For formal proofs, see Townsend (1979) and Gale and Hellwig (1985). This analysis neglects two subtleties. First, it assumes that verification must be a deterministic function of the state. One can show, however, that a contract that makes verification a random function of the entrepreneur's announcement of output can improve on the contract shown in Figure 10.1 (Bernanke and Gertler, 1989). Second, the analysis assumes that the investor can commit to verification if the entrepreneur announces that output is less than D. For any announced level of output less than D, the investor prefers to receive that amount without verifying than with verifying. But if the investor can decide ex post not to verify, the entrepreneur has an incentive to announce low output. Thus the contract is not *renegotiation-proof*. For simplicity, we neglect these complications.

The Equilibrium Value of *D*

The next step of the analysis is to determine what value of D is specified in the contract. Investors are risk-neutral and competitive, and the risk-free interest rate is r. Thus the expected payments to the investor, minus his or her expected spending on verification, must equal $1+r$ times the amount of the loan, $1-W$. To find the equilibrium value of D, we must therefore determine how the investor's expected receipts net of verification costs vary with D, and then find the value of D that provides the investor with the required expected net receipts.

To find the investor's expected net receipts, suppose first that D is less than the project's maximum possible output, 2γ. In this case, actual output can be either more or less than D. If output is more than D, the investor does not pay the verification cost and receives D. Since output is distributed uniformly on $[0,2\gamma]$, the probability of this occurring is $(2\gamma-D)/(2\gamma)$. If output is less than D, the investor pays the verification cost and receives all of output. The assumption that output is distributed uniformly implies that the probability of this occurring is $D/(2\gamma)$, and that average output conditional on this event is $D/2$.

If D exceeds 2γ, on the other hand, then output is always less than D. Thus in this case the investor always pays the verification cost and receives all of output. In this case the expected payment is γ.

Thus the investor's expected receipts minus verification costs are

$$R(D)=\begin{cases}\dfrac{2\gamma-D}{2\gamma}D+\dfrac{D}{2\gamma}\left(\dfrac{D}{2}-c\right) & \text{if } D\le 2\gamma\\[2ex] \gamma-c & \text{if } D>2\gamma.\end{cases} \tag{10.8}$$

Equation (10.8) also implies that when D is less than 2γ, $R'(D)$ is equal to $1-[c/(2\gamma)]-[D/(2\gamma)]$. Thus R increases until $D=2\gamma-c$ and then decreases. The reason that raising D above $2\gamma-c$ lowers the investor's expected net revenues is that when the investor verifies output, the net amount he or she receives is always less than $2\gamma-c$. Thus setting $D=2\gamma-c$ and accepting $2\gamma-c$ without verification when output exceeds $2\gamma-c$ makes the investor better off than setting $D>2\gamma-c$.

Equation (10.8) implies that when $D=2\gamma-c$, the investor's expected net revenues are $R(2\gamma-c)=[(2\gamma-c)/(2\gamma)]^2\gamma\equiv R^{\text{MAX}}$. Thus the maximum expected net revenues equal expected output when c is zero, but are less than this when c is greater than zero. Finally, R declines to $\gamma-c$ at $D=2\gamma$; thereafter further increases in D do not affect $R(D)$. The $R(D)$ function is plotted in Figure 10.2.

Figure 10.3 shows three possible values of the investor's required net revenues, $(1+r)(1-W)$. If the required net revenues equal V_1—more generally, if they are less than $\gamma-c$—there is a unique value of D that yields the investor the required net revenues. The contract therefore specifies this value

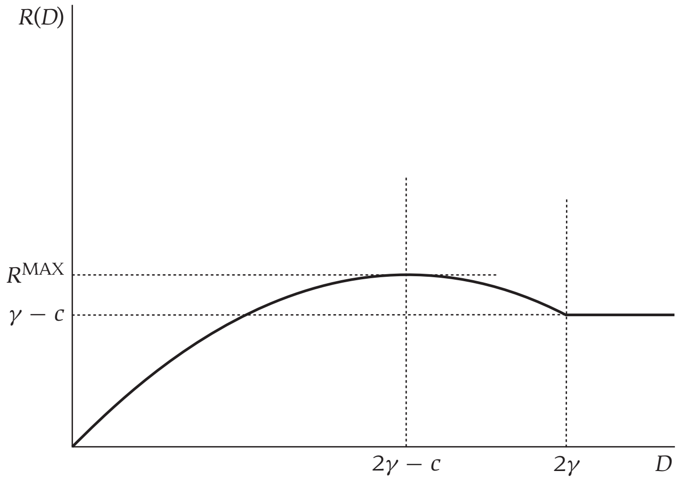

FIGURE 10.2 **The investor's expected revenues net of verification costs**

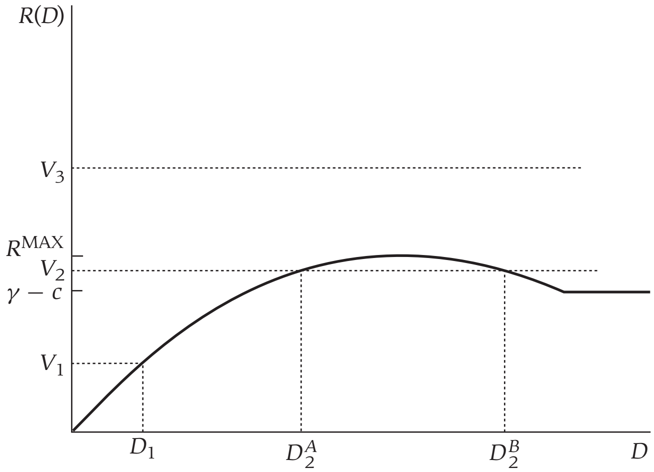

FIGURE 10.3 **The determination of the entrepreneur's required payment to the investor**

of D. For the case when the required payment equals V_1, the equilibrium value of D is given by D_1 in the figure.

If the required net revenues exceed R^{MAX}—if they equal V_3, for example—there is no value of D that yields the necessary revenues for the investor. Thus in this situation there is *credit rationing:* investors refuse to lend to the entrepreneur at any interest rate.

Finally, if the required net revenues are between $\gamma - c$ and R^{MAX}, there are two possible values of D. For example, the figure shows that a D of either D_2^A or D_2^B yields $R(D) = V_2$. The higher of these two D's (D_2^B in the figure) is not a competitive equilibrium, however: if an investor is making a loan to an entrepreneur with a required payment of D_2^B, other investors can profitably lend on more favorable terms. Thus competition drives D down to D_2^A. The equilibrium value of D is thus the smaller solution to $R(D) = (1 + r)(1 - W)$. Expression (10.9) implies that this solution is[2]

$$D^* = 2\gamma - c - \sqrt{(2\gamma - c)^2 - 4\gamma(1 + r)(1 - W)}$$
$$\text{for } (1 + r)(1 - W) \leq R^{\text{MAX}}. \tag{10.9}$$

Equilibrium Investment

The final step of the analysis is to determine when the entrepreneur undertakes the project. Clearly a necessary condition is that he or she can obtain financing at some interest rate. But this is not sufficient: some entrepreneurs who can obtain financing may be better off investing in the safe asset.

An entrepreneur who invests in the safe asset obtains $(1 + r)W$. If the entrepreneur instead undertakes the project, his or her expected receipts are expected output, γ, minus expected payments to the outside investor. If the entrepreneur can obtain financing, the expected payments to the investor are the opportunity cost of the investor's funds, $(1 + r)(1 - W)$, plus the investor's expected spending on verification costs. Thus to determine when a project is undertaken, we need to determine these expected verification costs.

These can be found from equation (10.9). The investor verifies when output is less than D^*, which occurs with probability $D^*/(2\gamma)$. Thus expected verification costs are

$$\begin{aligned} A &= \frac{D^*}{2\gamma}c \\ &= \left[\frac{2\gamma - c}{2\gamma} - \sqrt{\left(\frac{2\gamma - c}{2\gamma}\right)^2 - \frac{(1 + r)(1 - W)}{\gamma}}\,\right]c. \end{aligned} \tag{10.10}$$

Straightforward differentiation shows that A is increasing in c and r and decreasing in γ and W. We can therefore write

$$A = A(c, r, W, \gamma), \quad A_c > 0, \quad A_r > 0, \quad A_W < 0, \quad A_\gamma < 0. \tag{10.11}$$

[2] Note that the condition for the expression under the square root sign to be negative is that $[(2\gamma - c)/(2\gamma)]^2\gamma < (1 + r)(1 - W)$—that is, that R^{MAX} is less than required net revenues. Thus the case where the expression in (10.9) is not defined corresponds to the case where there is no value of D at which investors are willing to lend.

The entrepreneur's expected payments to the investor are $(1+r)(1-W)+A(c,r,W,\gamma)$. Thus the project is undertaken if $(1+r)(1-W) \leq R^{\mathrm{MAX}}$ and if

$$\gamma-(1+r)(1-W)-A(c,r,W,\gamma)>(1+r)W. \qquad (10.12)$$

In general, either constraint—investors' willingness to lend to the entrepreneur at some interest rate, or the entrepreneur's willingness to undertake the project if a loan is available—can be the relevant one. Suppose, for example, that W is very small. Then undertaking the project is attractive to the entrepreneur even when investors obtain the maximum possible revenues. Thus what determines whether the investment is undertaken is whether investors are willing to finance it. On the other hand, suppose W is substantial, but γ is only slightly above $1+r$. In this case, there is a level of D that gives investors their needed net revenues, $(1+r)(1-W)$, but the agency costs involved (that is, $A(c,r,W,\gamma)$) may exceed $\gamma-(1+r)$. In this case, whether the investment is undertaken is determined by whether the entrepreneur is willing to undertake it.

Discussion

Although we have derived these results from a particular model of asymmetric information, the basic ideas are general. Suppose, for example, there is asymmetric information about how much risk the entrepreneur is taking. In such a setting, if the investor bears some of the cost of poor outcomes, the entrepreneur has an incentive to increase the riskiness of his or her activities beyond the point that maximizes the expected return to the project. That is, asymmetric information about actions can create *moral hazard*. In that situation, asymmetric information again reduces the total expected returns to the entrepreneur and the investor, just as it does in our model of costly state verification. Under plausible assumptions, these agency costs are decreasing in the amount of financing that the entrepreneur can provide (W), increasing in the amount that the investor must be paid for a given amount of financing (r), decreasing in the expected payoff to the project (γ), and increasing in the magnitude of the asymmetric information (c when there is costly state verification, and the entrepreneur's ability to take high-risk actions when there is moral hazard).

Similarly, suppose that entrepreneurs are heterogeneous in terms of how risky their projects are, and that risk is not publicly observable—that is, suppose there is asymmetric information about types that creates *adverse selection*. Then again there are agency costs of outside finance, and again those costs are determined by the same types of considerations as in our model. Thus the qualitative results of this model apply to many other models of asymmetric information in financial markets.

Likewise, although we have discussed the effects of asymmetric information in the context of an entrepreneur with a potential investment project, its effects in other settings are broadly similar. For example, they carry over to the case of a household with uncertain future income that is considering buying a house or an automobile, and to that of an established firm with uncertain future profits that is considering increasing its capital stock.

The Financial Accelerator

A central implication of this analysis is that financial-market imperfections act to magnify the effects of shocks to the economy. With perfect financial markets, whether an investment is undertaken depends on the expected value of its output computed using the stochastic discount factor; what resources the agent undertaking the investment has are irrelevant. But as our model shows, this is no longer true when financial-markets are imperfect. Equation (10.11) shows that agency costs are a decreasing function of the entrepreneur's wealth, W; thus a fall in W can cause the investment not to occur in the absence of any other changes. More broadly, the resources that firms can use to finance investment depend on the excess of their current revenues over current expenses, and the resources households can use to purchase houses and consumer durables depend on their current income. When some force reduces output, those resources fall. Thus the agency costs associated with a given level of investment rise. As a result, investment falls, magnifying the initial fall in output. Bernanke and Gertler (1989) show this formally in the context of technology shocks in a simple real-business-cycle model, but the logic is general. For example, it applies equally to monetary shocks in sticky-price models.

Financial-market imperfections can magnify the effects of shocks not only through the shocks' impact on current resources, but also through their impact on the value of collateral. If we extended the model of this section to include collateral that the outside investor could claim in the event of default, agency costs would be greater when the value of the collateral was lower. And there are two reasons that asset values tend to fall when economy-wide output falls. First, and most straightforwardly, since declines in output are generally at least somewhat persistent, a fall in output today is typically associated with a decrease in expectations of future output. As a result, the expected future marginal products or service flows from assets fall, which translates into lower values of those assets today. Second, and more subtly, the financial-market imperfections themselves act to amplify the falls in collateral values from declines in output. For example, Kiyotaki and Moore (1997) develop a model where a productive asset that can be used as collateral is held by two types of firms: credit-constrained and unconstrained firms. When a fall in economy-wide output raises agency costs, and so reduces the constrained firms' ability to borrow and purchase the asset, more of the asset must be held by the unconstrained firms. With diminishing marginal product, the marginal product of the asset at these firms falls. In the baseline case in which there is no change in the interest rate at which these firms (which can borrow and lend at the prevailing safe interest rate) discount future profits, the value of the asset falls. The result is further reductions in the constrained firms' purchases of the asset, further falls in its value, and so on.

The idea that financial-market imperfections magnify the effects of shocks to economy-wide output is known as the *financial accelerator*.[3]

Other Implications

The model of this section has many other implications. As with the financial accelerator, the most important arise from financial-market imperfections in general rather than from the specific model. Here we discuss four.

First, financial-market imperfections have a particularly large magnification effect for output movements stemming from changes in interest rates. Consider an increase in the safe interest rate, for example as a result of a tightening of monetary policy in an economy with sticky prices. Agency costs rise not just because of the fall in output, but also because interest rates affect agency costs directly. An increase in r raises the total amount the entrepreneur must pay the investor. This means that the probability that the entrepreneur is unable to make the required payment is higher, and thus that agency costs are higher. This acts to reduce investment, and so further magnifies the output effects of the increase in interest rates.

Second, in the case of monetary and other aggregate demand disturbances, there is yet another channel through which financial-market imperfections make the real effects of shocks larger: as described in Section 6.7, they increase the degree of real rigidity. Recall that firms' incentives to change prices are an important determinant of the real effects of nominal shocks. Greater real rigidity—that is, a smaller responsiveness of profit-maximizing real prices to changes in aggregate output—reduces incentives for price adjustment in response to nominal shocks, and so increases the shocks' real effects. Costs of obtaining financing are one component of firms' costs. Thus, the fact that agency costs are higher when aggregate output is lower mutes the decline in firms' costs when output is lower, and so reduces the fall in their profit-maximizing prices. If firms face costs of adjusting their nominal prices, the result is that they adjust their prices by less (or are less likely to adjust them at all) in response to a monetary shock, and so the effects of the shock on aggregate output are greater.

Fourth, many variables that do not affect investment when capital markets are perfect matter when they are imperfect. Entrepreneurs' wealth

[3] Since the financial accelerator has nothing to do with an increasing rate of change, a more logical name would be financial amplifier. Nonetheless, the term *financial accelerator* is standard. See Bernanke, Gertler, and Gilchrist (1999) for a model of the financial accelerator in a much richer model of business cycles than the highly stylized models of Bernanke and Gertler (1989) and Kiyotaki and Moore (1997).

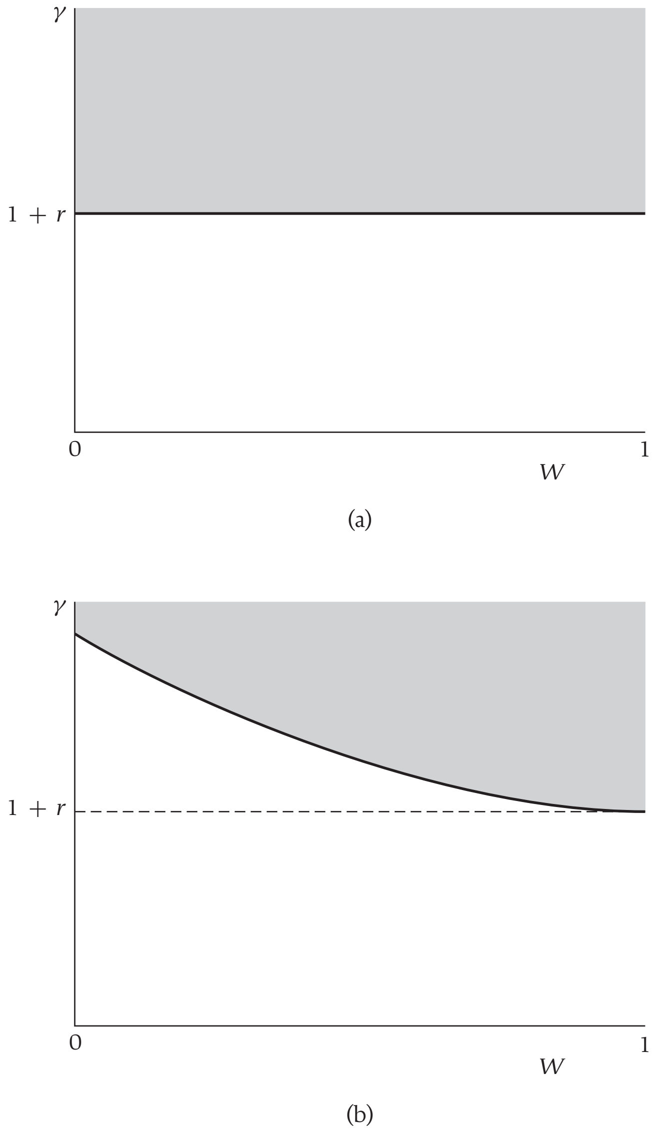

FIGURE 10.4 **The determination of the projects that are undertaken under symmetric and asymmetric information**

provides a simple example. Suppose that γ and W are heterogeneous across entrepreneurs. With perfect financial markets, whether a project is funded depends only on γ. Thus the projects that are undertaken are the most productive ones. This is shown in panel (a) of Figure 10.4. With asymmetric information, in contrast, since W affects the agency costs, whether a project is funded depends on both γ and W. Thus a project with a lower expected payoff than another can be funded if the entrepreneur with the less productive project is wealthier. This is shown in panel (b) of the figure.

Two other examples of variables that affect investment only when capital markets are imperfect are average tax rates and idiosyncratic risk. If taxes are added to the model, the average rate (rather than just the marginal rate) affects investment through its impact on firms' ability to use internal finance. And risk, even if it is uncorrelated with consumption, affects investment through its impact on agency costs. Outside finance of a project whose payoff is certain, for example, involves no agency costs, since there is no possibility that the entrepreneur will be unable to repay the investor. But, as our model shows, outside finance of a risky project involves agency costs.

Finally, and critically, our analysis implies that the financial system itself can be important to investment. The model implies that increases in c, the cost of verification, reduce investment. More generally, the existence of agency costs suggests that the efficiency of the financial system in processing information and monitoring borrowers is a potentially important determinant of investment.

This observation has implications for both long-run growth and short-run fluctuations. With regard to long-run growth, as discussed in the introduction to this chapter, there is evidence that the financial system is important to overall investment and growth. With regard to short-run fluctuations, our analysis implies that disruptions to the financial system can affect investment, and thus aggregate output. Recall that the transformation of saving into investment is often done via financial intermediaries, creating a two-level asymmetric information problem. This creates a potentially large propagation mechanism for shocks. Suppose some development—for example, the crash of the stock market in 1929 and the contraction of the economy in 1930, or the fall in house prices in 2007 and 2008—lowers borrowers' wealth. This not only reduces their ability to borrow and invest; it also weakens the position of financial intermediaries, and so reduces their ability to obtain funds from ultimate wealthholders. This reduces their lending, further depressing investment and output. This amplification can be compounded by links among intermediaries. In the extreme, some intermediaries fail. The end result can be catastrophic. Precisely these types of financial amplification mechanisms were at work in the Great Depression (Bernanke, 1983b) and in the crisis that began in 2007. We will return to these issues in the final two sections of the chapter.

10.3 Empirical Application: Cash Flow and Investment

Theories of financial-market imperfections imply that internal finance is less costly than external finance. They therefore imply that all else equal, firms with higher profits invest more.

A naive way to test this prediction is to regress investment on measures of the cost of capital and on *cash flow*—loosely speaking, current revenues minus expenses and taxes. Such regressions can use either firm-level data at a point in time or aggregate data over time. In either form, they typically find a strong link between cash flow and investment.

There is a problem with such tests, however: cash flow is likely to be correlated with the future profitability of capital. We saw in Section 9.5, for example, that our model of investment without financial-market imperfections predicts that a rise in profitability that is not immediately reversed raises investment. The reason is not that higher current profitability reduces firms' need to rely on outside finance, but that higher future profitability means that capital is more valuable. A similar relationship is likely to hold across firms at a point in time: firms with high cash flow probably have successful products or low costs, and thus have incentives to expand output. Because of this potential correlation between cash flow and future profitability, the regression may show a relationship between cash flow and investment even if financial markets are perfect.

A large literature, begun by Fazzari, Hubbard, and Petersen (1988), addresses this problem by comparing the investment behavior of different types of firms. Fazzari, Hubbard, and Petersen's idea is to identify groups of firms that may differ in their costs of obtaining outside funds. There is likely to be an association between cash flow and investment among both types of firms even if financial-market imperfections are not important. But the theory that financial-market imperfections have large effects on investment predicts that the association is stronger among the firms that face greater barriers to external finance. And unless the association between current cash flow and future profitability is stronger for the firms with less access to financial markets, the view that financial-market imperfections are not important predicts no difference in relationship between cash flow and investment for the two groups. Thus, Fazzari, Hubbard, and Petersen argue, the *difference* in the relationship between cash flow and investment between the two groups can be used to test for the importance of financial-market imperfections to investment.

The specific way that Fazzari, Hubbard, and Petersen group firms is by their dividend payments. Firms that pay high dividends can finance additional investment by reducing their dividends. Firms that pay low dividends, in contrast, must rely on external finance.[4]

[4] One complication is that it may be costly for high-dividend firms to reduce their dividends: there is evidence that reductions in dividends are interpreted as a signal of lower future profitability, and that the reductions therefore lower the value of firms' shares. Thus it is possible that the test could fail to find differences between the two groups of firms not because financial-market imperfections are unimportant, but because they are important to both groups.

Fazzari, Hubbard, and Petersen's test is thus an example of *difference in differences*. To see this, suppose for simplicity that cash flow has only two possible values, low and high. First consider the regression,

$$I_i = a + bD_i^{\text{HighCashFlow}} + e_i, \tag{10.13}$$

where I_i is investment by firm i and $D^{\text{HighCashFlow}}$ is a dummy for firms with high cash flow. The estimate of b from this regression shows the difference in investment between firms with high and low cash flow. As we just discussed, because there are reasons other than effects through the relative cost of internal and external finance that firms with higher cash flow might invest more, this regression is not a persuasive way of estimating the effect of cash flow on investment. So consider instead,

$$I_i = a + b_1 D_i^{\text{HighCashFlow}} + b_2 D_i^{\text{LowDividends}} + b_3 D_i^{\text{HighCashFlow}} D_i^{\text{LowDividends}} + e_i, \tag{10.14}$$

where $D^{\text{LowDividends}}$ is a dummy for firms with low dividend payments. With this specification, the estimate of b_1 shows the difference in investment among high-dividend firms between ones with high and low cash flow. The estimate of $b_1 + b_3$ shows the analogous difference among low-dividend firms. Thus the estimate of b_3 shows the difference in differences—specifically, the difference between low-dividend and high-dividend firms in the difference in investment between firms with high cash flow and firms with low cash flow. Fazzari, Hubbard, and Petersen's argument is that any correlation between cash flow and investment that is not causal is captured by b_1, and so finding a positive value of b_3 would indicate a causal impact of cash flow on investment.[5]

Although this discussion captures the essence of Fazzari, Hubbard, and Petersen's approach, their actual specification is more complicated than (10.14). They measure cash flow as a continuous variable rather than as a dummy. To account for differences in firm size, they measure both investment and cash flow as fractions of the firm's capital stock. In addition, their regression includes an estimate of q and dummy variables for each year (all entered both directly and interacted with the dummy for low-dividend firms), as well as dummy variables for each firm. The sample consists of 422 relatively large U.S. firms over the period 1970 to 1984. Low-dividend firms are defined as those with ratios of dividends to income consistently under 10 percent, and high-dividend firms are defined as those with

[5] Using differences in differences is even more appealing if there is a compelling argument that any correlation in the control group is not causal. In Fazzari, Hubbard, and Petersen's setting, if there were strong reasons to believe that the high-dividend firms faced no difference in the costs of external and internal finance (and that there was no other reason for cash flow to affect investment at those firms), the estimate of b_3 would be an estimate of the impact of cash flow on investment for low-dividend firms. In Fazzari, Hubbard, and Petersen's setting, however, such an interpretation is not persuasive.

dividend–income ratios consistently over 20 percent. (Fazzari, Hubbard, and Petersen also consider an intermediate-dividend group.)

Fazzari, Hubbard, and Petersen find that for the high-dividend firms, the coefficient on cash flow is 0.230 (with a standard error of 0.010); for the low-dividend firms, it is 0.461 (0.027). The t-statistic for the hypothesis that the two coefficients are equal is 12.1; thus equality is overwhelmingly rejected. The point estimates imply that low-dividend firms invest 23 cents more of each extra dollar of cash flow than the high-dividend firms do. Thus even if we interpret the estimate for the high-dividend firms as reflecting only the correlation between cash flow and future profitability, the results still suggest that financial-market imperfections have a large effect on investment by low-dividend firms.

Fazzari, Hubbard, and Petersen's identification is not airtight. Low-dividend firms differ systematically from high-dividend firms. For example, they are generally younger and less established. As a result, it is not out of the question that the correlation between current cash flow and future profitability differs importantly between low-dividend and high-dividend firms, in which case the differing relations between cash flow and investment could arise for reasons other than differential access to external finance.

Because of this concern, many authors have explored variations on Fazzari, Hubbard, and Petersen's approach. For example, Lamont (1997) compares the investment behavior of the nonoil subsidiaries of oil companies after the collapse in oil prices in 1986 with the investment behavior of comparable companies not connected with oil companies. The view that internal finance is cheaper than external finance predicts that a decline in oil prices, by reducing the availability of internal funds, should reduce the subsidiaries' investment; the view that financial-market imperfections are unimportant predicts that it should have no effect. Thus Lamont is employing differences in differences, looking at the difference between oil-affiliated subsidiaries and nonoil companies in the difference in investment between times of high and low oil prices. He finds a statistically significant and quantitatively large difference in the behavior of the two groups. Thus his results suggest that the barriers to outside finance are considerably larger than the barriers to finance between different parts of a company.

Another example of a variation on Fazzari, Hubbard, and Petersen's approach is the study by Rauh (2006), who focuses on how the idiosyncrasies in rules about required funding of pensions affect the internal funds firms have available for investment. Like Fazzari, Hubbard, and Petersen and Lamont, he finds large effects of internal funds. In addition, as we will discuss in Section 10.8, there is extensive evidence that credit availability was important to firms' behavior in the 2008 financial crisis, again supporting the view that financial-market imperfections are important.

Another line of work looks at the predictions of theories of financial-market imperfections concerning the importance not of firms' cash flow, but of the value of the assets they can use as collateral. The theories imply that a positive shock to the value of collateral makes it easier for a firm to obtain finance, and so increase its investment; if financial-market imperfections are unimportant, on the other hand, changes in the value of collateral should not matter. Gan (2007) and Chaney, Sraer, and Thesmar (2012) find that shocks to the value of firms' collateral that appear to be unrelated to their investment opportunities have large effects on investment.

More broadly, the literature on financial-market imperfections is one of unusual empirical consensus: there is strong and consistent evidence that cash flow and other determinants of access to internal resources affect investment, and that they do so in ways that suggest the relationship is the result of financial-market imperfections.[6]

10.4 Mispricing and Excess Volatility

In the models of financial markets we have considered so far—the partial-equilibrium models of consumers' asset allocation and the determinants of investment in Sections 8.5 and 9.7, and the general-equilibrium model of Section 10.1—asset prices equal their fundamental values. That is, they are the rational expectations, given agents' information and their stochastic discount factors, of the present value of assets' payoffs.

But a quick look at the actual behavior of asset prices raises doubts about whether this view is correct. There are many asset-price movements—including ones that are macroeconomically important—that, at least at first glance, cannot be easily explained by fundamentals. Examples include the 20 percent fall in the stock market in a single day with no evident major news in October 1987 and the run-ups and sharp declines in the overall stock market in the late 1920s, in internet stocks in the 1990s and early 2000s, and in house prices in the early 2000s. But of course, simply pointing to some puzzling asset-price movements hardly proves that there are large mispricings of assets.

The question of whether asset prices equal their fundamental values or whether they move much more than is warranted by fundamentals is important to two of the key issues about financial markets and the macroeconomy that we discussed at the start of the chapter. First, it is central to the question of whether financial markets can be an independent source of disturbances to the economy—a point we will return to at the end of this section. Second, if departures of asset prices from fundamentals take the form of sharp rises relative to fundamentals followed by sharp reversals, the result can be defaults by households and firms and failures of financial

[6] Kaplan and Zingales (1997) challenge this consensus both theoretically and empirically, focusing especially on Fazzari, Hubbard, and Petersen's evidence. But see Fazzari, Hubbard, and Petersen (2000) for a rebuttal.

intermediaries. As a result, as we will discuss in Section 10.7, such movements are a potential source of financial crises.

The assumption that asset prices equal their fundamental values is the natural benchmark. It is generally a good idea to start by assuming rationality and no market imperfections. And it is natural to think that even if only a small fraction of participants in financial markets are rational, that is enough to prevent mispricing: mispricing would create profit opportunities, and the profit opportunities would create incentives to trade against the mispricing (Friedman, 1953).

In fact, however, there is not an open-and-shut theoretical case that asset prices cannot differ from fundamentals. For a small number of rational agents to be enough to make mispricing impossible, one of two conditions must hold: either there must be no risk in trading to take advantage of mispricing (that is, there must be riskless arbitrage opportunities), or the agents must not only be rational, but risk-neutral with unlimited resources. Neither condition holds in practice.

This section is therefore devoted to examining the possibility of departures of asset prices from the values warranted by fundamentals. We start by considering a model, due to DeLong, Shleifer, Summers, and Waldmann (1990), that demonstrates what is perhaps the most economically interesting force making such departures possible. DeLong et al. show that departures of prices from fundamentals can be self-reinforcing: the very fact that there can be departures creates a source of risk, and so limits the willingness of rational investors to trade to correct the mispricing. Indeed, they develop a model in which the presence of some traders who act on the basis of incorrect beliefs has the effect that the price of an asset is not equal to its fundamental value even though its payoffs are certain and some investors are fully rational. After examining DeLong et al.'s model in detail, we will turn to other factors that limit the forces correcting mispricing and discuss the macroeconomic implications of our analysis.

Assumptions

DeLong et al. consider an economy with two seemingly identical assets. One unit of either asset pays a constant, known amount $r > 0$ of the economy's single good each period. Where the two assets differ is in their supply. The first asset, which we refer to as safe, can be converted into one unit of the economy's good at any time, and one unit of the good can be converted into the asset. This ensures that its price (in units of the good) is always 1. If not, agents could earn immediate riskless profits by selling the asset for goods and converting the proceeds into the asset (if its price exceeded 1), or buying the asset with goods and converting it into goods (if its price was less than 1). In contrast, the other asset, which we refer to as risky, cannot be created or destroyed. Thus its supply is equal to a constant, which for simplicity we normalize to 1. We denote its price in period t (in terms of the good in period t) by P_t.

The economy is an overlapping-generations economy with two types of agents. They are the same in several ways. They live for two periods, are price-takers, and value consumption only in the second period of life. For tractability, each agent's utility takes the constant-absolute-risk-aversion form, $U(C) = -e^{-2\gamma C}, \gamma > 0$, where C is the agent's consumption in his or her second period. Finally, each agent has the same amount of first-period income, which is constant over time. Since there is nothing special about whether consumption is positive or negative when agents have constant-absolute-risk-aversion utility, we normalize that income to zero.

Where the types differ is in their beliefs about the returns to the risky asset. The first type has rational expectations. That is, this type correctly perceives the distribution of returns from holding the asset. We refer to these agents as arbitrageurs. The second type misestimates the mean return on the asset. In particular, in period t, the entire distribution of the price of the asset in period $t+1$, p_{t+1}, perceived by each agent of this type, is shifted relative to the true distribution by an amount η_t. The η_t's are independent over time and normally distributed with mean zero and variance V_η. We refer to these agents as *noise traders*, which is the conventional term for agents who trade in financial markets based on considerations unrelated to fundamentals. The fraction of noise traders in each generation is f, where $0 \le f \le 1$. For simplicity, population growth is assumed to be zero, and the size of each generation is set to 1.

Analyzing the Model

The assumptions of the model are chosen so there is a stationary equilibrium where the price of the asset is linear in the shocks and where the distribution of agents' second-period consumption, conditional on their first-period information, is normal. Two assumptions are central to this result. The most obvious is that the only shock in the model (the shift in the noise traders' beliefs) is normally distributed with the same distribution each period. The other is that utility takes the constant-absolute-risk-aversion form. To see why this assumption leads to tractable results, note that if C is normally distributed with mean μ and variance W, then $-2\gamma C$ is normal with mean $-2\gamma\mu$ and variance $4\gamma^2 W$. Thus (by the properties of lognormal distributions), expected utility, $E[-e^{-2\gamma C}]$, is $2\gamma\mu - 2\gamma^2 W$, which is proportional to $\mu - \gamma W$. That is, the combination of normally distributed consumption and constant-absolute-risk-aversion utility implies that agents act as if they have linear preferences over the mean and variance of consumption. This causes outcomes to be linear in the shock. And since the shock is normally distributed, this linearity causes outcomes to be normally distributed as well.

Our strategy will therefore be to look for a stationary equilibrium where the price of the asset (P_t) is linear in the shock (η_t). To find such an equilibrium, consider first the arbitrageurs. Let $E_t[p_{t+1}]$ denote the rational expectation of p_{t+1} (the price of the risky asset in period $t+1$) given the information available at t, and let V denote the variance of $p_{t+1} - E_t[p_{t+1}]$. If an arbitrageur in period t buys X_t^a of the risky asset, he or she must hold $-p_t X_t^a$ of the safe asset (recall that first-period income is assumed to be zero). Both assets pay r per unit in period $t+1$. The risky asset is sold at a price of p_{t+1}, while the safe asset is sold at a price of 1. Thus the agent's second-period consumption is

$$\begin{aligned} C_{2t}^a &= r(X_t^a - p_t X_t^a) + p_{t+1} X_t^a - p_t X_t^a \\ &= [r + p_{t+1} - (1+r)p_t]X_t^a. \end{aligned} \tag{10.15}$$

Equation (10.15) implies that given the information available at time t, C_{2t}^a has mean $[r + E_t[p_{t+1}] - (1+r)p_t]X_t^a$ and variance $(X_t^a)^2 V$. Note also that the only variable in the expression for C_{2t}^a that is uncertain as of period t is p_{t+1}. Thus if p_{t+1} is linear in η_{t+1}, the distribution of C_{2t}^a (given the information available at t) is normal.

We saw above that an agent's expected utility is proportional to the mean of his or her consumption minus γ times its variance. Thus an arbitrageur's expected utility is proportional to $[r + E_t[p_{t+1}] - (1+r)p_t]X_t^a - \gamma(X_t^a)^2 V$. The first-order condition for the level of X_t^a that maximizes the agent's expected utility is therefore

$$[r + E_t[p_{t+1}] - (1+r)p_t] - 2\gamma X_t^a V = 0. \tag{10.16}$$

Solving for X_t^a yields

$$X_t^a = \frac{r + E_t[p_{t+1}] - (1+r)p_t}{2\gamma V}. \tag{10.17}$$

Since all period-t arbitrageurs are the same, each purchases this quantity of the risky asset.

The analysis of the representative noise trader's behavior is identical, except that $E_t[p_{t+1}]$ is replaced by the agent's incorrect belief about the mean of p_{t+1}, which is $E_t[p_{t+1}] + \eta_t$. His or her demand is therefore

$$X_t^n = \frac{r + E_t[p_{t+1}] + \eta_t - (1+r)p_t}{2\gamma V}. \tag{10.18}$$

When the economy enters period t, the fixed supply of the risky asset is held by old agents, who sell their holdings regardless of the price. Thus equilibrium requires that the sum of the demands of the $1-f$ arbitrageurs and the f noise traders equals the fixed supply, which we have set to 1:

$$(1-f)X_t^a + fX_t^n = 1. \tag{10.19}$$

Substituting expressions (10.17) and (10.18) for X_t^a and X_t^n and then solving for p_t gives

$$p_t = \frac{r + E_t[p_{t+1}] + f\eta_t - 2\gamma V}{1+r}. \tag{10.20}$$

Applying (10.20) to future periods (and taking expectations of both sides as of period t) implies that $E_t[p_{t+1}] = (r + E_t[p_{t+2}] - 2\gamma V)/(1+r)$, $E_t[p_{t+2}] = (r + E_t[p_{t+3}] - 2\gamma V)/(1+r)$, and so on.[7] Repeated substitution into (10.20) therefore gives us:

$$\begin{aligned} p_t = {} & \left(\frac{1}{1+r} + \frac{1}{(1+r)^2} + \frac{1}{(1+r)^3} + \cdots\right)(r - 2\gamma V) \\ & + \lim_{n\to\infty} \frac{E_t[p_{t+n}]}{(1+r)^n} + \frac{f\eta_t}{1+r}. \end{aligned} \tag{10.21}$$

The fact that we are focusing on stationary equilibria implies that the mean of p is constant over time, and thus that $\lim_{n\to\infty}[E_t[p_{t+n}]/(1+r)^n]$ is zero. In addition, the infinite sum in (10.21) simplifies to $1/r$. Thus (10.21) implies:

$$p_t = 1 - \frac{2\gamma V}{r} + \frac{f\eta_t}{1+r}. \tag{10.22}$$

The last step in solving the model is to rewrite V, the variance of $p_{t+1} - E_t[p_{t+1}]$, in terms of primitive parameters. The only stochastic term in (10.22) is $f\eta_t/(1+r)$. It follows that $V = [f^2/(1+r)^2]V_\eta$. Substituting this expression into (10.22) gives us our final equation for the equilibrium price in a given period:

$$p_t = 1 - \frac{2\gamma}{r}\frac{f^2}{(1+r)^2}V_\eta + \frac{f\eta_t}{1+r}. \tag{10.23}$$

Note that the price is linear in η_t. It follows that the distributions of the consumption of agents born at t, conditional on the information available at t, are normal (see [10.15]). Equation (10.23) also implies that the distribution of p is the same each period. Thus we have found an equilibrium of the form we were looking for.

[7] Recall that we defined V as the variance of $p_{t+1} - E_t[p_{t+1}]$. Thus this step implicitly assumes that the variances of $p_{t+2} - E_{t+1}[p_{t+2}]$, $p_{t+3} - E_{t+2}[p_{t+3}]$, and so on are all equal to V. Equation (10.22) shows that this assumption is correct in the equilibrium that we find. Another approach would be to appeal to the assumption of stationarity and the fact that no past variables appear in (10.20) to conjecture that $E_{t-1}[p_t] = E_{t-1}[p_{t+1}] = E_t[p_{t+1}]$. Taking expectations of both sides of (10.20) as of $t-1$ and using the conjecture then gives $E_t[p_{t+1}] = 1 - (2\gamma V/r)$. Substituting this expression into (10.20) then yields (10.22), and (10.22) shows that the conjecture is correct.

Discussion

The model's key implication is that the price of the asset is risky despite the fact that there is no uncertainty about its payoffs—everyone knows that it will pay r each period with certainty. The reason is that the fluctuations in the beliefs of the noise traders are themselves a source of risk. Rational traders considering buying the asset must be concerned not only about its dividend payments, but also about the price at which they will sell the asset. Even if the sentiments of the noise traders are pushing down the asset's current price, a rational trader who buys the asset at the low price faces the possibility that sentiments could deteriorate further by the time he or she needs to sell the asset, and so the price could fall even more. Thus, *noise-trader risk* limits the willingness of rational but risk-averse investors to trade to offset departures of prices from fundamentals. As DeLong et al. put it, noise traders "create their own space." The result is excess volatility of the price of the risky asset: the variance of the risky asset's price (which is positive) exceeds the variance of its fundamental value (which is zero).[8]

The short horizons of the rational traders are critical to this finding. If they had infinite horizons, they could hold the risky asset indefinitely, and so the selling price would be irrelevant. As a result, departures of the asset's price from its fundamental value would not be possible. The assumption of limited horizons is reasonable, however. The ultimate holders of assets are individuals. They are likely to need to sell their assets at some point, for example to pay for consumption in retirement or to smooth their consumption in the face of fluctuations in labor income. In addition, as we will discuss below, there are forces that may make the horizons of portfolio managers shorter than those of the underlying asset-holders.

The model implies not only that the price of the risky asset fluctuates, but also that on average it is less than its fundamental value of 1 (see [10.23]). This result is a direct consequence of the noise-trader risk: the asset is riskier than is warranted by fundamentals, which reduces the average amount that the agents in the model (who are risk averse) are willing to pay for it.

[8] One might expect that even if there is an equilibrium where asset prices fluctuate in response to noise traders' sentiments, there would be another where they do not. After all, if rational traders know that the price of the "risky" asset always equals 1, the asset is riskless, and so their demand for the asset is perfectly elastic. What rules out this potential equilibrium in DeLong et al.'s model is their assumption that what is exogenous and constant is the variance of the error in noise traders' beliefs about the mean of next period's price, rather than the variance of the shock to the quantity of the risky asset they demand. As a result, in the proposed equilibrium where the asset's price is constant and equal to 1, noise traders' demand each period would be infinite (either positive or negative). This discussion implies that the result that there is not a second equilibrium where the noise traders do not cause prices to depart from fundamentals is not general. Rather, DeLong et al. make a particular assumption that eliminates this equilibrium, which allows them to focus on the interesting case where there are departures from fundamentals. See Problem 10.10 for more on these issues.

Finally, the model has implications about the effects of changes in the parameters. Most are unsurprising. Equation (10.23) shows that the mean departure of prices from fundamentals is larger when agents are more risk averse, when there are more noise traders, and when the variance of sentiment shocks is greater. And it is smaller when r is larger, which corresponds to a larger fraction of the present value of the payoff to the asset being paid in the next period. Similarly, the effect of a given shift in sentiment is larger when there are more noise traders and smaller when r is greater.

The only implications about the effects of the parameters that may be surprising concern two parameters that do not appear in the last term of (10.23): neither risk aversion (γ) nor the variance of sentiment shocks (V_η) influences how a given shift in sentiment affects the price of the asset. The reason is that the sentiment shocks correspond to a given change in expectations of next period's price rather than in the quantity demanded. As a result, higher risk aversion and a greater variance of sentiment shocks mute not only the arbitrageurs' willingness to trade to correct mispricing, but also noise traders' willingness to trade on the basis of their sentiments (see equation [10.18]). These two forces offset one another, with the result that neither γ nor V_η enters into the last term of (10.23).

Other Factors Limiting Arbitrageurs' Willingness to Trade to Correct Mispricing

Researchers have identified two factors in addition to noise-trader risk that mute the extent to which sophisticated investors are willing to trade to move asset prices back toward fundamentals if they depart from them. Like noise-trader risk, these forces make it easier for asset prices to differ from fundamentals.

The most obvious additional factor is *fundamental risk*. Suppose an asset is undervalued given currently available information. An investor who buys the asset faces the risk that its price will fall because of the arrival of new information about its future payoffs. That is, the price of an undervalued asset can fall not just because its price falls further below fundamentals (as in our model of noise-trader risk), but also because the fundamental value of the asset declines. By reducing sophisticated investors' willingness to trade against mispricings, fundamental risk magnifies the effects of departures from fundamentals coming from other sources.

The second is *agency risk* (Shleifer and Vishny, 1997). Many of the investors who are most likely to trade to exploit mispricing rely mainly on funds obtained from others. If the funders base their assessment of the investors' abilities partly on their short-run performance, they may withdraw their funds—and so force investors to sell undervalued assets—precisely in situations where the mispricing has worsened in the short run, and so the prices of the undervalued assets have fallen. That is, the fact that sophisticated traders are often acting as agents of less knowledgeable individuals can force the sophisticated investors to behave as if they have short horizons. Notice that in contrast to the baseline model of noise-trader risk, in this case the sophisticated investors' short horizons arise endogenously, and only occur when the departure of an asset's price from fundamentals becomes larger.[9]

Problems 10.6 and 10.7 develop a model where noise-trader risk, fundamental risk, and agency risk are present together.[10]

Macroeconomic Implications

Our analysis shows that the forces working against possible mispricing are not infinitely strong. Risk-neutral agents with unbounded funds at their disposal do not exist. As a result, rational investors' willingness to trade to correct a possible mispricing is inherently limited. We have discussed three factors that weaken the forces working to mute mispricing: noise-trader risk, fundamental risk, and agency risk.

But the fact that departures of prices from fundamentals are possible does not prove that they are important for the economy. A first question is whether they can be quantitatively large. A simple calculation suggests that they can if two key conditions are met: they involve assets whose prices are relatively volatile, and they are on average fairly persistent. Consider a class of assets that rational investors estimate is overvalued relative to fundamentals by 30 percent. Suppose first that the asset class is all equities and that on average a tenth of a mispricing disappears over a year. Recall from Section 8.5 that the average annual excess return of equities over a safe interest rate is about 6 percentage points, and that the standard deviation of the excess return is about 17 percentage points. Then rational investors believe that the expected excess return over the next year is 3 percentage points, and (if returns are normal) that there is a 43 percent chance rather than the usual 50 percent chance that the excess return will be greater than its long-term average. The optimal response to this information is simply to hold fewer equities than usual, not to make large trades against them. Thus the forces acting to prevent such a mispricing are not particularly strong. On the other hand, if the mispricing concerns a class of assets whose

[9] The essence of agency risk is nicely summarized by a statement that is sometimes attributed (erroneously) to Keynes: "The market can remain irrational longer than you can remain solvent."

[10] A fourth factor potentially limiting sophisticated investors' willingness to trade to correct mispricing is *model-based risk*: arbitrageurs cannot be certain that their estimates of fundamental values are in fact the best estimates given the available information. This risk can be thought of as just a subtle form of fundamental risk: the fundamental value of the asset may turn out to be less than currently expected not just because of the arrival of conventional types of news, but also because of new information showing that the model that sophisticated investors were using to estimate fundamentals was incorrect.

excess return has a standard deviation of 5 percentage points and on average one-third of the mispricing is corrected in a year, then rational investors believe that the expected excess return over the coming year is −4 percent, and that the chance of the excess return being negative is about 80 percent. Thus the incentives to trade against the mispricing are vastly larger, and so the forces preventing such a mispricing are vastly stronger.[11]

A second question about whether departures of asset prices from fundamentals can be important to the economy is whether substantial mispricings can have large macroeconomic effects. Again the answer appears to be yes: large movements in asset prices appear to have large effects on the composition and level of economic activity. Theoretically, we know that asset prices affect both investment and consumption. In the q theory model of investment, the market prices of various capital goods are critical inputs into firms' investment decisions. And as described in Section 10.2, increases in asset prices reduce distortions in financial markets, and so raise investment. Likewise, the permanent-income hypothesis implies that increases in wealth resulting from higher asset prices raise consumption. And if households are liquidity constrained, increases in wealth relax the constraints, and so potentially raise consumption by more than implied by the permanent-income hypothesis. In the extreme, a household that is at a corner solution for its choice of consumption may raise its consumption one-for-one with increases in wealth (see equation [8.64] in Section 8.6).

Empirically, the huge run-up in the prices of technology stocks in the late 1990s appears to have led to a large increase in investment in fiber-optic cable, capital goods of various internet start-ups, and so on, and to have been an important driver of the boom in overall economic activity in that period. Likewise, the enormous rise in house prices in the early 2000s appears to have caused very large increases in housing investment. More interestingly, there is evidence that it also led to large increases in consumption, particularly among low-income homeowners, as homeowners tapped their new wealth (Mian and Sufi, 2014). As with the 1990s dot-com boom, these forces again had a notable impact on the overall economy.

Thus, there are strong reasons to believe that substantial departures of asset prices from fundamentals are possible, and that if they take place, they are likely to have significant macroeconomic consequences. But that does not tell us whether they do occur. The next section therefore turns to empirical evidence on that issue.

[11] Endogenous information acquisition is likely to make these effects self-reinforcing. Information about mispricings of assets with low volatility where departures of prices from fundamentals are corrected quickly is much more valuable than information about mispricings of assets with high volatility where departures from fundamentals are very persistent. Individuals interested in profiting from trading against mispricings therefore concentrate on acquiring the first type of information. As a result, the pool of agents who know enough to identify large, fairly persistent mispricings and who trade against them is endogenously reduced.

10.5 Empirical Application: Evidence on Excess Volatility

The analysis in the previous section shows that the prices of large categories of assets could depart substantially from fundamentals, and that this could have important effects on the economy. But it does not tell us whether those things in fact occur.

Unfortunately, a direct attack on this question is unlikely to be insightful. As we just discussed, there were large run-ups in the prices of technology stocks in the 1990s and of houses in the early 2000s. It is tempting to interpret the fact that those prices then fall sharply (as well as other episodes where prices of assets first rose greatly and then plummeted) as showing that there are sometimes large departures of prices from fundamentals. This judgment is implicit in such phrases as "the dot-com bubble" and "the housing bubble." But there could be run-ups and collapses of asset prices not because of mispricing, but because the best available information first suggests large upward revisions to the rational expectations of the present value of payoffs and then suggests that those revisions were unwarranted. Moreover, the real-time assessments of experts in periods of rapidly rising asset prices show that determining fundamental values is difficult, and thus that large ex post errors are at best weak evidence of mispricing. For example, after Federal Reserve chair Alan Greenspan famously warned of "irrational exuberance" in the stock market in December 1996, stock prices almost doubled over the next three years—a performance that does not fit easily with Greenspan's implication that stocks were overvalued. More tellingly, Gerardi, Foote, and Willen (2011) document that expert housing economists held a wide range of views about the reasonableness of house prices during their run-up in the early 2000s. And Cheng, Raina, and Xiong (2014) show that the personal investment decisions of real-estate professionals in the run-up suggest they did not believe that houses were overvalued.[12]

More broadly, the same forces that mute agents' willingness to trade to correct mispricings can make the mispricing of an asset (or of a category of assets) almost impossible to detect. Summers (1986) constructs an example in which the standard deviation of the departure of an asset's price from fundamentals is 30 percent, yet because of the persistence of the departures and the volatility of the asset's price, it would typically take thousands of years of data to confidently reject the null hypothesis that the price is a random walk. And even establishing that an asset price is not a random walk would not show that there are departures from fundamentals. In well-functioning

[12] One corollary of this discussion is that the term "bubble" does not have a clear-cut meaning, unless it is being used to describe a rational bubble (see Problem 8.9). Thus any use of the term should be accompanied by a clear statement of what it is intended to mean.

financial markets, assets' payoffs are valued using the appropriate stochastic discount factor, not a constant discount rate. As a result, expected returns can vary over time in response to variation in the expected change in the marginal utility of consumption and its comovement with assets' payoffs.

Because of these difficulties, more compelling efforts to assess whether there are macroeconomically important departures of asset prices from fundamentals generally involve working up from microeconomic evidence. At the level of individual stocks, there are cases where there is clear-cut evidence of mispricing. In particular, although it is almost never possible to confidently determine the fundamental value of a stock, there are settings where two assets are so closely related that one can be confident of their *relative* fundamental values. One type of case along these lines arises when two assets represent claims on identical payoffs. For example, for historical reasons, shares of Shell Oil and shares of Royal Dutch Petroleum entitle the holder to the same dividend streams. Thus the fundamental values of the two assets are the same. But as Rosenthal and Young (1990) and Froot and Dabora (1999) document, the prices of the two stocks differ, often by substantial amounts. They also show that such complications as voting rights and taxes cannot come close to explaining the differences.

Another type of case along these lines arises when one asset includes another as one component. Barring highly unusual circumstances, the fundamental value of the broader asset must be higher than that of the narrower one. But Lamont and Thaler (2003) document several cases of asset pairs where the price of the narrower asset exceeded the price of the broader one. One striking example is that of 3Com and Palm in 2000, documented in Figure 10.5. During this period, 3Com owned much of Palm, yet shares of Palm also traded separately. Thus investors had two ways of obtaining a claim on Palm's payoffs: by buying shares in 3Com or by buying shares in Palm directly. Since the first approach entitled the investor to something else as well, namely a claim on 3Com's non-Palm payoffs (which, as Lamont and Thaler document, could not plausibly have had negative value), the fundamental value of the first asset was necessarily larger than that of the second. But as Figure 10.5 shows, the prices of the two assets often strongly violated that relationship: from early March 2000 until the middle of May, the implied value of 3Com's non-Palm payoffs was negative—often substantially so. Thus at least one of the two assets must have been mispriced. The relationship between the two prices only became sensible when it was known that 3Com was about to divest itself of Palm, so that holders of 3Com shares would soon receive shares of Palm.

The fact that there are mispricings at the microeconomic level makes it more plausible that there are macroeconomically important mispricings, but it does no more than that. A step that brings us closer to the macroeconomic level is a comparison of closed-end mutual funds and their underlying assets. A closed-end fund holds shares in other publicly traded stocks. Since a share in a closed-end mutual fund is a claim on a portfolio, its fundamental value is

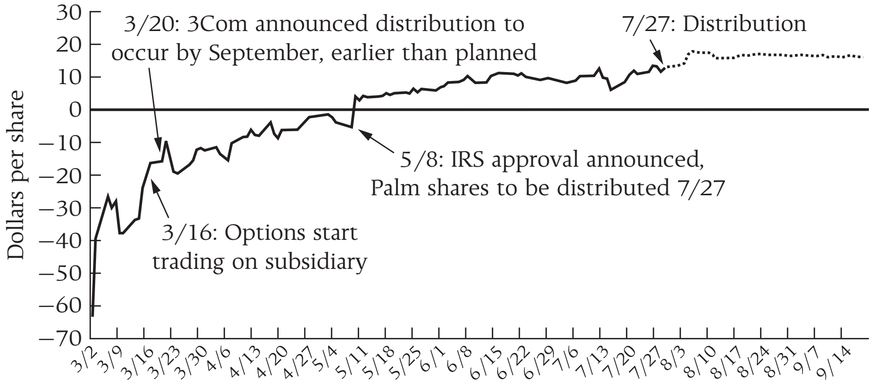

FIGURE 10.5 The implied value of the non-Palm portion of 3Com, March 2–September 18, 2000 (from Lamont and Thaler, 2003)[13]

the same as the sum of the fundamental values of the assets in the portfolio. But shares in a closed-end fund cannot be redeemed with the fund; instead, they can only be sold to other investors.[14] Thus a departure of the price of a closed-end fund from the price of the underlying portfolio does not create an opportunity for immediate riskless profit.

Lee, Shleifer, and Thaler (1991) document that the prices of closed-end funds often differ from those of their underlying portfolios. More importantly, they show that the differences are in line with what one would expect if noise traders are especially likely to trade closed-end funds and if those funds are subject to noise-trader risk. Shares in closed-end funds on average trade at a discount relative to the underlying portfolios (as one would expect if they were subject to an additional source of risk—see equation [10.23]); they occasionally trade above (which is generally when new closed-end funds are created); and their returns are more variable than those on the underlying portfolios.

Taking the analysis a step further, Lee, Shleifer, and Thaler (1991) and DeLong and Shleifer (1991) document a relationship between the discount on closed-end funds and the behavior of overall stock prices. If noise traders are particularly likely to trade closed-end funds, then the behavior of the closed-end discount may reflect "sentiment" about stocks. That is, it may be that optimism about stocks that is not warranted by fundamentals drives the closed-end discount down, and pessimism drives the discount up. These authors therefore ask whether the value of the overall stock market relative

[13] Reprinted by permission from the *Journal of Political Economy*. Copyright 2003 by the University of Chicago.

[14] This is the defining feature of a closed-end fund; the shares of an open-end fund, in contrast, can be redeemed with the fund.

to current dividends or earnings is higher when the closed-end discount is low or negative, and low when the closed-end discount is high. They find that it is. For example, the closed-end discount was sharply negative (that is, there was a substantial closed-end premium) in the huge run-up in stock prices before the 1929 crash. In the same spirit, Froot and Dabora (1999) show that the price of Royal Dutch Petroleum (which is traded mainly on U.S. stock exchanges) relative to the price of Shell Oil (which is traded mainly on U.K. exchanges) moves with the overall value of U.S. stocks relative to U.K. stocks. Of course, these results do not prove that the rise in stock prices in the late 1920s (and, more generally, the component of overall stock-price movements that is correlated with movements in the closed-end discount) was not warranted by fundamentals, or that sentiment plays a role in driving the relative value of the U.S. and U.K. stock markets. But the results are consistent with what one would expect if these things were true.

A very different way of obtaining evidence that bears on mispricing is by asking market participants about their thinking. For instance, Case, Shiller, and Thompson (2012) find that homebuyers often report views that seem very difficult to justify on rational grounds, such as expectations of double-digit annual price appreciation for a decade. Moreover, such expectations are particularly common in periods of rapidly rising house prices. Again, this does not prove that these arguably irrational expectations are the source of the price movements, but the patterns are certainly consistent with that view.

The idea that asset prices equal the rational expectation of the appropriately discounted value of assets' payoffs given the available information is extremely appealing. But the analysis of the previous section and this one leads to the conclusion that that hypothesis may nonetheless be very far off as a description of reality. Theoretically, the forces that would correct mispricing if it arose are limited, and so large departures of asset prices from fundamentals are possible. Empirically, one cannot reject large mispricing at the level of large classes of assets; there is clear evidence of mispricing at the microeconomic level; and some of the microeconomic patterns fit with what one would expect if there were important aggregate mispricing.

10.6 The Diamond–Dybvig Model

We have discussed how asymmetric information can have important effects on investment and how it can magnify the macroeconomic impact of shocks. And we have seen that asset prices may depart from the values warranted by fundamentals and that such departures can have important consequences for the economy. But none of the analysis we have done so far gets at a central feature of financial markets: they are subject to sudden, convulsive changes. These changes happen at both the microeconomic and macroeconomic levels. At the microeconomic level, a bank that appears to be operating normally one day can face a run the next; or an investment bank that is funding itself by rolling over short-term loans at low interest rates can suddenly find that it cannot borrow at any interest rate. And at the macroeconomic level, sudden changes like these can affect many institutions almost simultaneously.

This section focuses on convulsions at the level of individual institutions. In particular, it presents the classic Diamond–Dybvig model of the possibility of a bank run (Diamond and Dybvig, 1983; see also Bryant, 1980). What drives the possibility of a run in the model is demand for liquidity—that is, a desire on the part of savers to be able to retrieve their funds at any time. If the underlying investment projects that financial intermediaries are funding are long-term, this creates *maturity mismatch*: the intermediaries' assets (their claims on a portion of the returns to the investment projects) are long-term, but their liabilities (the savers' claims on the intermediary) can be redeemed at any time, and so are short-term. Diamond and Dybvig develop a model that shows why banks may create maturity mismatch to meet the demand for liquidity and how this mismatch creates the possibility of a run.

Assumptions

Diamond and Dybvig develop a model where there is a demand for assets that resemble traditional demand deposits. That is, the assets have a preset value and can be redeemed at any time. They then show that if investment projects are long-term, a financial institution that issues demand deposits is vulnerable to runs.

Specifically, there are three periods, denoted 0, 1, and 2. The economy consists of a continuum of agents, each of whom is endowed with 1 unit of the economy's single good in period 0. If the good is invested, it yields $R > 1$ units of the good if it is held until period 2, but only 1 unit if the project is stopped in period 1. The fact that the two-period return exceeds the one-period return gives investment a long-term character.

Ex ante, all individuals are the same. But in period 1, fraction θ learn that they only value consumption in period 1. The remainder are willing to consume in either period 1 or period 2. We refer to the individuals who want to consume only in period 1 as "type-a" individuals and the others as "type-b" individuals, and we assume $0 < \theta < 1$. Importantly, an individual's type is not observable by others.

Let c_t^i be the consumption of a type-i individual in period t. The utilities of the two types are given by

$$U^a = \ln c_1^a, \tag{10.24}$$

$$U^b = \rho \ln\left(c_1^b + c_2^b\right), \tag{10.25}$$

where $0 < \rho < 1$ and $\rho R > 0$.

These assumptions, which are obviously not general, are chosen so that the model leads to economically interesting possibilities. For example, the assumption that $\rho < 1$ has the effect that the type a's particularly value consumption.

Two Baseline Cases

Before introducing the possibility of a financial intermediary, it is useful to consider two simple variants of the model.

The first case is autarchy. That is, we rule out any type of trade or insurance among individuals. In this case, individuals' optimization problem is trivial. Since they do not value period-0 consumption, each individual invests his or her 1 unit of endowment. Those who learn in period 1 that they are type a's liquidate their investment projects early and have period-1 consumption of 1 and period-2 consumption of 0. Individuals who turn out to be type b's hold their projects to period 2 and have period-1 consumption of 0 and period-2 consumption of R. Thus expected utility under autarchy is

$$\begin{aligned} U^{AUTARCHY} &= \theta \ln 1 + (1-\theta)\rho \ln R \\ &= (1-\theta)\rho \ln R. \end{aligned} \tag{10.26}$$

The second special case is a social planner who can observe individuals' realized types. Because type a's get no utility from period-2 consumption, the planner will clearly choose $c_2^a = 0$. And because type b's are indifferent between consumption in the two periods and projects yield more if they are held to maturity, the planner will also clearly choose $c_1^b = 0$. Thus the interesting choice variables are c_1^a and c_2^b.

A project that is liquidated early yields 1 unit. It follows that the fraction of projects liquidated early must equal θ (the fraction of individuals who consume in period 1) times c_1^a (period-1 consumption per individual who consumes in period 1). This leaves fraction $1-\theta c_1^a$ that are held to maturity. Each yields R, and the output is divided among the type b's, who are fraction $1-\theta$ of the population. Thus the planner's resource constraint (when $c_2^a = c_1^b = 0$) is

$$c_2^b = \frac{(1-\theta c_1^a)R}{1-\theta}. \tag{10.27}$$

A representative individual's expected utility is $\theta \ln c_1^a + (1-\theta)\rho \ln c_2^b$. Using the budget constraint, (10.27), to substitute for c_2^b, we can write this as

$$E[U] = \theta \ln c_1^a + (1-\theta)\rho[\ln(1-\theta c_1^a) + \ln R - \ln(1-\theta)]. \tag{10.28}$$

Before solving for the utility-maximizing value of c_1^a, it is helpful to ask what happens if the social planner changes c_1^a marginally from the autarchy outcome. As shown earlier, the autarchy outcome is $c_1^a = 1$. Thus,

$$\begin{aligned}\left.\frac{\partial E[U]}{\partial c_1^a}\right|_{c_1^a = c_1^{a,AUTARCHY}} &= \left.\frac{\partial E[U]}{\partial c_1^a}\right|_{c_1^a = 1}\\ &= \theta + \frac{(1-\theta)\rho}{1-\theta}(-\theta)\\ &= (1-\rho)\theta\\ &> 0.\end{aligned} \tag{10.29}$$

It is easy to check that $\partial^2 E[U]/\partial c_1^{a2} < 0$. Hence a planner who wants to maximize the representative individual's expected utility and who can observe types will transfer some resources from the type b's to the type a's. The intuition is simply that the type a's particularly value consumption.

Equation (10.28) implies that the first-order condition for the optimal level of c_1^a under full information is

$$\frac{\theta}{c_1^{a*}} + \frac{(1-\theta)\rho}{1-\theta c_1^{a*}}(-\theta) = 0, \tag{10.30}$$

which implies

$$\begin{aligned}c_1^{a*} &= \frac{1}{\theta + (1-\theta)\rho}\\ &> 1.\end{aligned} \tag{10.31}$$

Substituting this expression into (10.27) gives

$$\begin{aligned}c_1^{b*} &= \frac{\rho R}{\theta + (1-\theta)\rho}\\ &< R.\end{aligned} \tag{10.32}$$

Notice that although c_1^{b*} is less than its level under autarchy, the assumption that $\rho R > 1$ implies it is greater than c_1^{a*}.

A Bank

One of Diamond and Dybvig's key results is that we do not need either observability of types or a social planner to achieve the first best. Consider what happens if one individual sets up a bank. The bank offers to take deposits on the following terms. Any individual—regardless of type—who deposits 1 unit can withdraw c_1^{a*} in period 1 if the bank has funds available. Whatever funds the bank has in period 2 are divided equally among the depositors who do not withdraw in period 1. The bank pays the depositors by investing its deposits in the projects and liquidating projects as needed to meet the demand for early withdrawals. Note that these assumptions imply that the owner of the bank breaks even. He or she does not put in any resources of his or her own, and the output obtained with the depositors' resources is all paid out to the depositors.[15]

Since each unit invested yields only 1 unit if it is liquidated in period 1 and c_1^{a*} is greater than 1, it is necessary to specify what happens if a large fraction of depositors (specifically, more than $1/c_1^{a*}$) asks to withdraw early. Diamond and Dybvig assume that in this situation, the bank provides c_1^{a*} to as many of the early withdrawers as possible and nothing to the remainder. Because the bank has no way of distinguishing among individuals, the ones who receive c_1^{a*} are assumed to be chosen at random. The assumption that the bank pays the promised amount to as many early withdrawers as possible and nothing to the remainder is intended as a shortcut way of modeling the idea that instead of there being a single moment when some individuals discover that they need liquidity and make early withdrawals, liquidity needs arise at different times for different individuals, and so there is some heterogeneity in the timing of early withdrawals. In the context of banking, this first-come, first-served assumption is known as a *sequential service constraint*. Notice that in the case where more than $1/c_1^{a*}$ of depositors withdraw early, the bank liquidates all its projects in period 1, and so depositors who wait until period 2 get nothing.

Under these assumptions, the social optimum—type a's getting c_1^{a*} and type b's getting c_2^{b*}—is a Nash equilibrium. To see this, suppose everyone believes that the type a's, and only the type a's, will withdraw in period 1. Since the bank's period-2 resources are divided equally among period-2 withdrawers, in the proposed equilibrium the amount that each period-2 withdrawer receives is

$$\begin{aligned} c_2 &= \frac{(1-\theta c_1^{a*})R}{1-\theta} \\ &= c_2^{b*}, \end{aligned} \tag{10.33}$$

where the second line uses the economy's resource constraint, (10.27). A representative type-a individual will clearly choose to withdraw in period 1, since he or she only values period-1 consumption. And since $c_2^{b*} > c_1^{a*}$ and type b's are indifferent about the timing of their consumption, the type b's will wait until period 2. That is, there is a Nash equilibrium where the economy attains the first best even though individuals' types are unobserved and there is no government intervention.

Thus, one of Diamond and Dybvig's central results is that a bank can provide liquidity: it makes long-term investments but allows depositors to access funds before the investments mature. Depositors value liquidity

---

[15] Thus we are implicitly assuming free entry into banking, so that profits are driven to zero.

because of their uncertainty about when they will most want their funds. And the bank is able to provide liquidity because depositors end up wanting their funds at different times and it can pool risks about the timing of desired withdrawals across depositors.

Also, notice that early withdrawers obtain an above-market return: c_1^a is greater than 1, the realized value of an investment that is liquidated in period 1. This result is realistic. A bank that pays interest on deposits makes some investments (in the form of loans, for example), but also holds some cash and highly liquid, low-interest-rate assets (Treasury bills, for example) to allow for early withdrawals. Despite the lower interest rate the bank earns on those assets, depositors who withdraw early are paid the same interest rate as ones who do not.

The Possibility of a Run

Unfortunately, although the social optimum is a Nash equilibrium, there is a second equilibrium: a bank run. Consider what happens if each type b believes that all agents, not just the type a's, will try to withdraw their deposits in period 1. As described above, the fact that $c_1^{a*} > 1$ means that if every agent tries to withdraw early, the bank is not able to satisfy them all. It has to liquidate all its investments, and there is nothing left in period 2. Thus, if a type b believes all other type b's will try to withdraw in period 1, he or she is better off trying to withdraw in period 1 (and having a positive probability of getting c_1^{a*}) than waiting until period 2 (and getting zero for sure). That is, a bank run—all agents trying to withdraw early—is a Nash equilibrium.[16]

[16] This analysis ignores one complication. Because utility is logarithmic (see [10.24] and [10.25]), the fact that individuals face some chance of having zero consumption when there is a run means that their expected utility in the run equilibrium is infinitely negative. As a result, any positive probability of a bank run would cause individuals to be unwilling to deposit their endowment in the bank. That is, it appears that rather than providing a candidate explanation of bank runs, the model shows that the possibility that there might be a bank run could prevent banks from attracting any deposits, and so prevent the economy from reaping the benefits of pooling liquidity risk across individuals.

There are at least two ways to address this complication. First, as Diamond and Dybvig point out, if there is a positive probability of a run, individuals could still want to deposit part of their endowments in the bank. By depositing some of their endowment but retaining some to invest themselves, they would guarantee themselves positive consumption for sure even in the event of a run. Thus there can be an equilibrium with a bank that attracts deposits and where there is a positive probability of a run, and so the model provides a candidate explanation of runs. The second approach is to make a minor change to the utility functions. Specifically, suppose we replace (10.24) and (10.25) with functions that are logarithmic over the range from 1 to R, but not infinitely negative when consumption is zero. With this change, the autarchy and first-best outcomes (and expected utilities in those cases) are unchanged. But expected utility in the run equilibrium is now well defined. As a result, as long as the probability of a run is not too large, individuals are willing to deposit their entire endowment in the bank even in the face of a strictly positive probability of a run. Thus again the model provides a candidate explanation of runs.

Discussion

The possibility of a run is inherent when a bank has illiquid assets and liquid liabilities. Liquid liabilities give depositors the option of withdrawing early. But since the bank's assets are illiquid, if all depositors try to withdraw early, the bank will not be able to satisfy them. As a result, if each agent believes that all others are trying to withdraw early, they believe the bank cannot meet its obligations, and so they too will try to withdraw early.

One implication of this discussion is that the precise source of agents' desire for liquidity is not critical to the possibility of a run. In Diamond and Dybvig's model, the desire for liquidity arises from agents' uncertainty about the timing of their consumption needs. But the results would be similar if it arose instead from a desire on the part of entrepreneurs for flexibility to pursue unexpected investment opportunities (Holmström and Tirole, 1998; Diamond and Rajan, 2001). More intriguingly, Dang, Gorton, and Holmström (2015) argue that a desire for liquidity can arise from problems of asymmetric information. If funders know they can withdraw their funds at the first indication of trouble, their need to carefully assess the quality of the underlying assets they are investing in is greatly reduced. In Dang, Gorton, and Holmstrom's terminology, asymmetric information gives rise to a desire for *informationally insensitive* assets—of which assets that can be liquidated for a predetermined price at any time are a prime example.

Likewise, a run can take various forms. At a traditional bank with demand deposits, it can involve depositors physically rushing to the bank to try to withdraw their funds. But modern bank runs rarely resemble this. Think of a bank that is financing itself both by attracting deposits from retail investors (that is, individual households) and by rolling over very short-term loans from wholesale investors (that is, institutions such as money market mutual funds); or think of an investment bank financing itself almost entirely by such short-term loans. Then a run may take the form of many wholesale investors simultaneously refusing to roll over their loans, or simultaneously making the terms of the loans much more onerous. As a result, the bank may be forced to liquidate its investments early, and fail as a result (Gorton and Metrick, 2012). Because the bank is harmed when many lenders do not roll over their loans, each lender's belief about whether others are rolling over their loans is important to its decision about whether to roll over its own loan. Thus, as in a bank run in the Diamond–Dybvig model, widespread refusal to roll over loans can be a self-fulfilling equilibrium. Or consider a financial institution that is funding itself by rolling over medium-term loans that come due at different times. Then a run can take the form of each lender refusing to roll over its loan when it comes due out of a belief that later lenders will do the same. In that case, the "run" unfolds over time rather than occurring all at once (He and Xiong, 2012).

In the Diamond–Dybvig model, a run is a pure *liquidity crisis* for the bank. All agents know that if the type b's did not try to withdraw in period 1, the bank would have enough funds to make its promised payments in period 2. In that sense, the bank is completely solvent. It is only the fact that all depositors want their funds immediately that makes it unable to meet its obligations.

The alternative to a liquidity crisis is a *solvency crisis*. Suppose, for example, there is a possibility of the bank manager absconding with some of the funds, or of a shock in period 1 that causes a substantial fraction of the investment projects to fail. With such extensions of the model, the bank is sometimes unable to repay all the type b's even if they wait until period 2; that is, it is sometimes insolvent. In such situations, all agents have an incentive to withdraw their funds early—but now that is true regardless of whether they believe other agents are also trying to withdraw early.

In cases like these, the distinction between a liquidity run and a solvency run is clear. But in other cases, it is not. For example, suppose there is a small (but strictly positive) chance the bank would be unable to meet all its obligations to the type b's if they waited, and suppose there is some heterogeneity among the type b's (perhaps in terms of their risk aversion or their degree of impatience). The small probability of insolvency may make it a dominant strategy for some type b's (such as the most risk averse or the least patient) to withdraw early, but this may lead others to withdraw early, which may lead yet others to do so, and so on. The end result may be that a small probability of insolvency leads to a run that causes the bank to fail for sure. Such a run cannot be fruitfully described as either a pure liquidity run or a pure solvency run.

Policies to Prevent Runs

Diamond and Dybvig consider three policies that can prevent a liquidity run. The first can be implemented by the bank, while the other two require government action.

The policy that can be implemented by the bank is a *suspension of payments*. Specifically, suppose it offers a slight variant of the contract we have been considering: it will pay out c_1^{a*} in period 1 to at most fraction θ of depositors. With this contract, a decision by some type-b agents try to withdraw early has no impact on the amount the bank pays out in period 1, and so has no effect on the resources the bank has available in period 2. Each type-b agent is therefore better off waiting until period 2 regardless of what he or she thinks others will do. Thus the policy eliminates the run equilibrium.

Such a policy is similar to what banks actually did before the advent of government deposit insurance. A bank facing a run would announce that depositors could withdraw their funds only at a discount. In the model, setting the discount such that any depositor who waits until period 2 is sure to get more than what he or she can get in period 1 eliminates the run.

In practice, a discount of a few percent typically helped to stabilize a bank facing a run, but did not move it immediately to a no-run equilibrium.

In the model, the policy of redeeming no more than fraction θ of deposits in period 1 restores the first-best outcome. Diamond and Dybvig show, however, that in an extension of the model where θ is uncertain, it does not. They therefore consider two possible government policies.

The first is deposit insurance. If the government can guarantee that anyone who waits until period 2 receives c_2^{b*}, this eliminates the run equilibrium. In this situation, the type b's always wait, and so the bank can always pay them c_2^{b*} in period 2. Thus the government never needs to pay out funds.

This simple analysis leaves out an important issue, however: for the guarantee to be credible, the government must have a way of obtaining the resources needed to pay depositors who wait until period 2 if some type b's run in period 1. Without this ability, the run is still an equilibrium: a guarantee that is not credible provides no reason for a depositor to not run if he or she believes others are running.

What makes the guarantee credible, Diamond and Dybvig argue, is the government's power to tax. Concretely, suppose the government's policy is that if more than fraction θ of agents withdraw in period 1, so that the bank's period-2 resources will be less than c_2^{b*} per remaining depositor, it will levy a tax on each agent in period 1 sufficient to increase the consumption of depositors who did not withdraw in period 1 to c_2^{b*}. Then each agent has no incentive to run.

The other government policy that Diamond and Dybvig consider is for it to act as a *lender of last resort*. Concretely, suppose the government—in practice, the central bank—announces that it stands ready to lend to the bank at a gross interest rate of c_2^{b*}/c_1^{a*}. Consider what happens if fraction $\phi > \theta$ of depositors withdraw their funds in period 1 when such a policy is in place. The bank can pay θ of them by liquidating projects and the remaining $\phi - \theta$ by borrowing from the central bank. Since it has liquidated only fraction θ of its projects, it has $(1-\theta)c_2^{b*}$ of resources in period 2, just as it would if only fraction θ of depositors had withdrawn in period 1. It can use $(\phi-\theta)c_2^{b*}$ to repay the central bank and the remaining $(1-\phi)c_2^{b*}$ to pay the depositors who withdraw in period 2. Thus each depositor knows that he or she can obtain c_2^{b*} in period 2 regardless of how many depositors withdraw in period 1. As a result, a type-b depositor will not want to withdraw early even if others do; that is, the central bank's policy eliminates the run equilibrium.[17]

This discussion raises two issues. The first, paralleling our discussion of deposit insurance, is how the central bank's offer to lend resources to the

[17] See Richardson and Troost (2009) for compelling evidence from a fascinating natural experiment that showed lender-of-last-resort policies can have large effects on banks' chances of survival in the face of runs.

bank in the event of excess withdrawals is credible. After all, the presence of a central bank does not increase the quantity of goods in the economy. One possibility is that the central bank is backed by a fiscal authority with the power to tax. In this case, a lender-of-last-resort policy is similar to deposit insurance.

Another (more interesting) possibility is to introduce money and the possibility of inflation into the model. Suppose that deposits are a claim not on amounts of goods in period 1 or period 2, but on amounts of money. In such a setting, the central bank can respond to a large number of period-1 withdrawals by lending money to the bank. The resulting increase in the money supply raises the price of the good in period 1. This in turn reduces the consumption of the early withdrawers below c_1^{a*} (since in this scenario their withdrawals are denominated in dollars, not goods), which means that the economy continues to satisfy its resource constraint regardless of the number of early withdrawals. With the central bank's promise rendered credible by the possibility of inflation, type *b*'s have no incentive to withdraw early, and so the central bank never has to act on its promise. Although these ideas are intriguing, explicitly making the model a monetary one is complicated, and so we will not pursue them formally.

The second issue raised by the lender-of-last-resort policy concerns the terms under which the central bank lends to the bank. The technological tradeoff between goods in the two periods is 1 to R. But if the central bank merely stood ready to lend to the bank at a gross interest rate of R, the bank could not borrow enough to prevent a run. With this interest rate, if all depositors withdrew in period 1, the bank would need to borrow $(1-\theta)c_1^{a*}$ to meet period-1 demand, and so it would need to repay $(1-\theta)c_1^{a*}R$ in period 2. But it would have only $(1-\theta)c_2^{b*}$ available. Thus the ratio of what it would have available to the amount it would need to repay the loan is

$$\begin{aligned}\frac{(1-\theta)c_2^{b*}}{(1-\theta)c_1^{a*}R} &= \frac{c_2^{b*}}{c_1^{a*}R} \\ &< 1,\end{aligned} \tag{10.34}$$

where the second line uses the fact that c_2^{b*}/c_1^{a*} is less than R. That is, if the bank borrowed $(1-\theta)c_1^{a*}$ at a gross interest rate of R, it could not repay the loan—which means that it is not feasible for the central bank to make a loan of $(1-\theta)c_1^{a*}$ on those terms. It follows that if the central bank charged a gross interest rate of R on its loans, the run equilibrium would remain.

This analysis shows that to eliminate the run equilibrium, the central bank needs to stand ready to lend to the bank at a below-market interest rate. In particular, as described previously, being willing to lend at a gross interest rate of c_2^{b*}/c_1^{a*}, which is less than R, solves the problem. But the fact that the needed interest rate is less than the market rate means that the central bank cannot make an unconditional offer to lend to the bank. Suppose it does. Then the bank is better off satisfying all demands for period-1 withdrawals by borrowing from the central bank rather than by liquidating projects. If only the type a's withdraw in period 1, for example, the bank liquidates no projects, borrows θc_1^{a*} in period 1, and has resources R in period 2. It repays θc_2^{b*} to the central bank and pays $(1-\theta)c_2^{b*}$ to the period-2 withdrawers, leaving it with a profit of $R - c_2^{b*}$ at the expense of the central bank.[18]

A solution to this danger is for the central bank to require the bank to meet the first θ of period-1 withdrawals by liquidating assets, and to lend only for withdrawals beyond that level. One can think of such a rule as a reserve requirement: the bank is required to be able to meet demands for immediate withdrawal of some fraction (but not all) of its demand deposits from its own assets.

The most famous prescription for how policymakers should respond to a banking panic is Bagehot's dictum that they should lend freely against good collateral at a penalty rate (Bagehot, 1873, Chapter 7). Bagehot's prescription is often invoked today. The Diamond–Dybvig model supports the part of the prescription about good collateral—in the model, the central bank's loans are certain to be repaid. But it leads to conclusions that are the opposite of the other parts of the prescription. Under the lender-of-last-resort policy in the model, the central bank lends not freely but subject to restrictions, and not at a penalty rate but at a discount. Interestingly, the behavior of modern central banks in panics seems to follow the implications of the Diamond–Dybvig model rather than Bagehot's rule.

Finally, note that in the event of an economy-wide run (an issue we will consider in the next section), it is inherent that the private sector cannot provide deposit insurance or serve as a lender of last resort. In the context of the model, the economy's resource constraint makes it impossible for all individuals to have c_1^{a*} in period 1. More broadly, if the financial system as a whole is providing liquidity by issuing short-term liabilities and holding long-term assets, it necessarily does not have enough short-term assets to meet the demand if all its creditors try to redeem their assets. Thus there is no private entity that can provide insurance or make loans to the entire financial system. A government with the power to tax or to create inflation is needed.

10.7 Contagion and Financial Crises

We have encountered two models in this chapter where there can be sudden changes in a borrower's ability to obtain funds. Most obviously, this is a central feature of the Diamond–Dybvig model: banks in that model are

[18] Another way of writing the bank's profits is as the amount it borrows, θc_1^{a*}, times the difference between the market interest rate and the rate charged by the central bank, $R - (c_2^{b*}/c_1^{a*})$. This product is $\theta R c_1^{a*} - \theta c_2^{b*}$. The economy's resource constraint, $\theta c_1^{a*} + [(1-\theta)c_2^{b*}/R] = 1$, implies $\theta R c_1^{a*} = R - (1-\theta)c_2^{b*}$. Thus we can write the bank's profits as $[R - (1-\theta)c_2^{b*}] - \theta c_2^{b*}$, or $R - c_2^{b*}$. This approach therefore leads to the same conclusion.

subject to runs, and so they may find themselves suddenly unable to attract or keep deposits. But it is also a property of Section 10.2's model of investment in the presence of financial-market imperfections. As we discussed, in that model there is a region of the parameter space where an entrepreneur cannot get funding under any terms. It follows that a small change in parameter values can cause the entrepreneur's borrowing to drop discontinuously from a strictly positive amount to zero. Moreover, this possibility of credit rationing is not an artifact of the specific assumptions of the model of Section 10.2, but a natural outcome when there is financing under asymmetric information. As we saw, higher interest rates increase agency costs; one consequence is that a higher interest rate does not necessarily increase expected payments to a lender. As a result, there can be situations where there is no interest rate at which investors are willing to supply funds.

Unfortunately, we cannot jump from the finding that there can be sudden changes in financial relationships at the microeconomic level to the conclusion that we have found an explanation of financial crises at the macroeconomic level. In a Diamond–Dybvig model extended to include multiple banks, there is no particular reason for there to be runs on many banks at the same time. Likewise, consider the model of Section 10.2 with heterogeneous entrepreneurs. Any change in fundamentals is likely to change the number of entrepreneurs who cannot borrow. But for a small change to lead to a large change in that number, there would have to be many entrepreneurs right on the boundary of being rationed, and there is no clear reason for that to be the case. In addition, there are forces that would tend to stabilize the overall financial system in the face of microeconomic disruptions. If one entrepreneur becomes unable to borrow at any interest rate, prevailing interest rates are likely to fall, making it easier for others to borrow. Similarly, if borrowers run on one Diamond–Dybvig bank, this might increase the supply of funds to other banks.

Contagion

A crucial fact about the macroeconomy, however, is that there are financial crises—that is, simultaneous difficulties at many financial institutions and disruptions of many agents' ability to borrow. Thus a critical question about financial disruptions is how they spread. Researchers have identified four possible sources of such *contagion*.

The first and most intuitive source of contagion is *counterparty contagion*. Financial institutions often hold various types of claims on one another. When one institution faces a run, and hence a risk of failure, the value of other institutions' claims on it generally fall. Thus the financial health of the institution's counterparties—that is, the institutions on the other side of its financial transactions—is likely to suffer. This can push them into insolvency, or at least cause sufficient doubts about their solvency to trigger runs.

Similarly, if one of the entrepreneurs in the model of Section 10.2 has obligations to other entrepreneurs, something that makes the first entrepreneur no longer able to borrow may worsen the financial health of the other entrepreneurs, potentially making them unable to borrow, and so on.[19]

The second possibility is *confidence contagion.* A development that makes one institution or agent unable to obtain funds can cause suppliers of funds to reassess the conditions of other institutions or agents, and so potentially become unwilling to lend to them. For example, recall the variant of the Diamond–Dybvig model in which the distinction between a liquidity run and a solvency run is blurred and in which information suggesting a small probability of insolvency triggers a run. Suppose that in such a situation, there are many banks, depositors have heterogeneous information about banks' health, and the values of the assets of different banks are positively correlated. Then if depositors at one bank observe a run at another, they infer that depositors there have adverse information about the health of that bank, and so they lower their estimates of the value of their own bank's assets. The result may be that they run on their own bank—leading depositors at yet other banks to lower their estimates of their banks' asset values, potentially triggering additional runs, and so on. Notice that with confidence contagion, contagion occurs even though a run on one bank has no direct impact on the value of other banks' assets.

The third type of contagion is *fire-sale contagion.* An institution facing a run is likely to sell assets to meet its depositors' demand. If financial markets were frictionless, the price of any asset would be determined by its state-contingent payoffs and the marginal utility of consumption in each state, and a decision by one institution to sell an asset would have no effect on its price. But when financial markets are less than perfect, the sale is likely to reduce the price of the asset. If other institutions also hold some of the asset (or similar assets), the value of their assets will fall—potentially leading to insolvency or runs.

The final type of contagion is *macroeconomic contagion.* Difficulties or failures at some institutions or borrowers are likely to reduce overall economic activity (for example, by making it harder for firms to borrow to pay suppliers or workers or by causing some firms to shut down). Although this could in principle help other firms (for example, by shifting demand toward them), it could also hurt them by reducing overall demand. If the negative effect dominates, it could weaken some firms' situations to the point where financial institutions are no longer willing to lend to them, or increase

[19] This discussion focuses on counterparty contagion from borrowers to lenders: a borrower who cannot repay a loan harms the lender, which may make the lender less able to repay its own lenders, and so on. But counterparty contagion can also operate in the opposite direction: if a lender finds that it has fewer funds and so cuts back on lending to one of its usual borrowers, the borrower may adjust by reducing its lending to its own borrowers, and so on (Glasserman and Young, 2016).

the likelihood of loan defaults by enough to make some financial institutions insolvent or to trigger runs on them. This could reduce economic activity further, and the process could continue. Moreover, the knowledge that such effects are likely to occur will cause some of the effects on institutions' health and ability to borrow to be immediate, rather than occurring only as economic activity declines.

All four types of contagion almost surely played a role in the massive financial crisis that followed Lehman Brothers' declaration of bankruptcy in September 2008. A large money market mutual fund, the Reserve Primary Fund, had made loans to Lehman Brothers. As a result, Lehman's failure affected Reserve Primary through counterparty contagion, leading to a run. That led to runs on other money market mutual funds; since they had no direct link with Reserve Primary, the effect appears to have been through confidence contagion. Meanwhile, many institutions had to sell assets to meet demand from their creditors or to make up for the fact that some of their normal sources of funds had dried up. Although we cannot be sure whether the prices of those assets fell below what was warranted by fundamentals, both the magnitude of the sales and the fact that many normal relationships among asset prices broke down suggest that the sales led to fire-sale contagion. Finally, it was clear that the disruptions would exacerbate the downturn that was already underway, and the prices of stocks and of all but the safest bonds plummeted. This macroeconomic contagion surely played a role in propagating the financial troubles.

Some Issues in Understanding Financial Crises

This discussion raises a host of questions concerning financial crises. Because this is an active area of research with few clear answers, we will only discuss some broad issues without delving into details.

One obvious issue concerns the relative importance of the various types of contagion. Is there one that plays a dominant role in most financial crises? Are different ones dominant in different cases? Are all four usually important? The honest answer to these questions is that at this point we know very little.

A second issue raised by contagion concerns the efficiency of decentralized outcomes. If the disruption of one financial relationship affects agents not involved in that relationship, this suggests the possibility of externalities, and thus the possibility of inefficiency. Such external effects appear to be present with confidence, fire-sale, and macroeconomic contagion; whether they arise with counterparty contagion is not as clear. In settings where contagion with negative externalities is possible, the equilibrium probability of a financial crisis is likely to be inefficiently high. If we take the forms of financial contracts as given, the inefficiency could manifest itself as excessively high levels of actions that generate risk of contagion, such as taking on more debt or issuing more demand deposits. But taking a step back, the inefficiency could also involve contract forms that are not socially optimal. For example, the social optimum might involve less use of debt contracts and more use of contracts with contingencies triggered by macroeconomic outcomes. And there could be other types of inefficiency, such as excessive interconnections among financial institutions (if the possibility of counterparty contagion leads to inefficiency), or excessive correlation in the types of assets held by different institutions (if fire-sale contagion is important).

The issues of the relative importance of different types of contagion and of potential inefficiency naturally raise questions for policy: if there is inefficiency, there may be scope for welfare-improving government intervention. But determining what interventions would be most beneficial requires knowing what externalities are most important.

In the aftermath of the global financial crisis that erupted in 2008, policymakers have pursued a wide range of interventions intended to reduce the chances of future crises—suggesting that either they are quite uncertain about what externalities are most important or they believe many to be important. The policies include higher capital requirements for financial institutions (making them less likely to become insolvent if their assets fall in value); higher liquidity requirements (making them less likely to fail in the face of a run); special capital and liquidity requirements for "systemically important" financial institutions, or even putting a ceiling on the size of financial institutions; requiring some types of contracts to be aggregated and cleared centrally rather than contract-by-contract (a policy that appears to be aimed at reducing counterparty contagion); making capital requirements cyclical, so that institutions must build up buffers in good times that they can run down in bad times; "stress tests" intended to show how institutions would fare in an adverse scenario (implying that policymakers believe that the equilibrium level of information provision is too low); blanket prohibitions on some types of activities by financial institutions that issue government-insured deposits or that could be candidates for lender-of-last-resort loans; changes in how employees or executives of financial institutions can be compensated; changes in how firms that assess the riskiness of bonds can be compensated; and more.

This list can be seen not just as a recitation of policies and policy proposals, but as a research agenda. For each proposal, we can ask: what is the externality that warrants the intervention? Can we write down a complete model where the externality is present? What is the evidence about the quantitative importance of the externality? Is the proposed policy the optimal way of dealing with the externality, or might a different approach, such as a Pigovian tax, be preferable? Would the proposed policy have other effects?

A more subtle issue raised by our discussion of contagion is why financial crises end. If problems at a handful of institutions can trigger contagion that disrupts a large portion of the financial system, then one might expect the additional contagion unleashed by those further disruptions to be vastly more powerful, and so lead to vastly more problems—with the contagion continually magnified until it produced a conflagration that destroyed the entire financial system. Yet even in the absence of government intervention, we do not see this. In the Great Depression, for example, although one-third of banks failed, the remaining two-thirds (all of which were surely substantially affected by various types of contagion) survived. This suggests that a simple emphasis on contagion is missing something. One story that fits the broad contours of what we observe is that there are not only important destabilizing forces from contagion, but also important stabilizing forces; and that initially the destabilizing forces dominate, but as financial failures and disruptions accumulate, the stabilizing forces eventually win out. But this is just a sketch of an idea that does not even begin to address some basic questions—most notably, what the most important stabilizing forces are, and why they are dominated by contagion over some ranges but not others.

The final issue raised by this discussion that we will touch on is the link between mispricings and financial crises. A large fall in the prices of an important category of assets, such as stocks or houses, can trigger a financial crisis through its impacts on the wealth and collateral values of borrowers, the values of items on the asset side of financial institutions' balance sheets, and the aggregate level of output. And a sharp fall in asset prices can occur either because of news implying that the rational estimate of fundamental values is much lower than before, or because of the end of a large overpricing of assets. Thus, large mispricings may make financial crises more likely.

The possibility of a link between mispricings and financial crises again raises issues for both research and policy. On the research side, it increases the importance of determining whether large mispricings in fact ever occur, and if so, what their causes are. On the policy side, it increases the importance of measures to prevent large mispricings (which requires understanding what causes them) and of detecting them in real time.

A Whirlwind Tour of Some of Current Research

The issues we have been discussing are all active areas of research. Fire sales have been the subject of a particularly large amount of recent work. The classic analysis is that by Shleifer and Vishny (1992). On the theoretical side, Gromb and Vayanos (2002), Lorenzoni (2008), and Dávila and Korinek (2017) are leading examples of papers that identify externalities that arise from fire-sale effects and show conditions under which the result is excessive aggregate risk-taking, and so an inefficiently high probability of a financial crisis. Recent empirical investigations of fire sales include Mitchell, Pedersen, and Pulvino (2007); Ellul, Jotikasthira, and Lundblad

(2011); Merrill, Nadauld, Stulz, and Sherlund (2014); and Koijen and Yogo (2015). All find strong evidence of fire-sale effects.

Two strands of recent work investigate factors other than possible fire-sale externalities that can lead to excessive aggregate risk-taking, and so again increase the risk of a crisis. First, Farhi and Werning (2016), Korinek and Simsek (2016), and Schmitt-Grohé and Uribe (2016) show that because the zero lower bound on nominal interest rates (which we will discuss in Section 12.7) or membership in a currency union may prevent monetary policy from being able to offset aggregate demand effects and macroeconomic contagion in a downturn, there can be negative externalities from high levels of credit creation in a boom. Second, Farhi and Tirole (2012) and Gertler, Kiyotaki, and Queralto (2012) study how the government's incentives to intervene ex post to support the financial system in the event of a crisis distorts banks' ex ante decisions about risk-taking.

Another area that—not surprisingly—has been a major area of work is runs on financial institutions. Because of deposit insurance, old-fashioned runs of retail depositors on banks were not a central part of the recent crisis. But the crisis did involve important "run-like" behavior. Gorton and Metrick (2012); Iyer and Puri (2012); Covitz, Liang, and Suarez (2013); and Schmidt, Timmermann, and Wermers (2016) all study the microeconomics of runs, in many cases looking at highly disaggregated and very-high-frequency data. An example of recent theoretical work on runs is the extension of the Diamond–Dybvig model to a general-equilibrium setting with many banks by Gertler and Kiyotaki (2015).

Many interesting lines of work involve narrower issues related to the subjects we have been discussing. An example of a narrower area where there has been insightful recent research is credit ratings agencies. It is appealing to argue that the fact that ratings agencies were paid by the issuers of the bonds being rated led to ratings that were systematically too generous, and that this caused buyers to systemically underestimate the riskiness of the bonds they were purchasing in the run-up to the crisis. But this argument faces the theoretical problem that purchasers should be able to account for the agencies' incentives, and the empirical problem that reputational considerations have allowed the "rater pays" model to succeed in other contexts—most famously in the case of Underwriters Laboratories, which rates products' safety. Thus it is necessary to go beyond the simple argument. Coval, Jurek, and Stafford (2009) present evidence that the missing ingredient is that many purchasers of bonds did not account for the distinction between systematic and idiosyncratic risk in interpreting ratings. In contrast, Skreta and Veldkamp (2009) argue that the missing ingredient is that bond purchasers did not fully understand how the interaction of competition among rating agencies and the increased complexity of the assets the agencies were rating affected equilibrium ratings. Other examples of recent work on ratings agencies include Becker and Milbourn (2011), Griffin and Tang (2011), and Bolton and Freixas (2012).

This brief tour omits important areas of modern research on issues related to the contagion, crises, and policy. For example, Allen and Gale (2000), Elliott, Golub, and Jackson (2014), and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015) discuss contagion through networks of interconnections, and so shed light on counterparty contagion. To give another example, Hachem (2014) investigates the conditions under which agency problems and compensation policies within financial institutions can lead to socially inefficient levels of risk-taking. And in a very different vein, Admati and Pfleiderer (2013) and Mian and Sufi (2015, Chapter 12) propose major shifts in policies involving the financial system.

Two messages should come through from this quick and very incomplete survey. First, the crisis that began in 2007 and worsened dramatically in the fall of 2008 is having a profound effect on macroeconomics. Issues that were at the periphery of the field—or that were not viewed as part of macroeconomics at all and that were getting little attention in any part of economics—are now the subjects of intensive research. And second, that research is still in its early stages: many new questions are being asked and many new ideas are being proposed, but we are far from having answers. One implication is that today is a particularly exciting, and particularly important, time for macroeconomics.

10.8 Empirical Application: Microeconomic Evidence on the Macroeconomic Effects of Financial Crises

Determining the effects of financial crises on the overall economy is challenging. The main problem is one we have encountered repeatedly in tackling empirical questions: disentangling correlation and causation. Financial crises in large modern economies do not occur in a vacuum. They often happen after large buildups of household and corporate debt, and around the same time as large falls in asset prices; and they sometimes occur together with losses of confidence in a country's currency or in its government's debt. As a result, there are generally factors correlated with financial crises that are likely to affect the economy directly. Furthermore, because financial markets are forward-looking, news that provides information about the likely path of the economy (for example, news suggesting that homebuilding is likely to fall) can raise doubts about the solvency of financial institutions, and so trigger an immediate financial crisis. Thus even if financial crises precede other developments that are likely to reduce output directly, we cannot be confident that the crises are the cause of the output declines.

Because of these inherent difficulties in trying to use aggregate evidence to determine the macroeconomic effects of financial crises, a great deal of research uses microeconomic evidence to shed light on this issue. Not surprisingly, much of this research focuses on the recent global financial crisis. This section discusses two leading examples of this work.

Ivashina and Scharfstein's Evidence on Lending

In a very timely paper, Ivashina and Scharfstein (2010) investigate the impact of the crisis on lending by major financial institutions. One virtue of their paper is that it shows the wisdom of the common statement that in studying any issue, it is important to understand the relevant institutional details.

The starting point of Ivashina and Scharfstein's analysis is the observation by Chari, Christiano, and Kehoe (2008) that the stock of commercial and industrial loans on bank balance sheets *rose* sharply immediately after the collapse of Lehman Brothers in September 2008. It is tempting to interpret this as showing that there was not any shift back in credit supply. Ivashina and Scharfstein point out, however, that this proposed interpretation neglects an important feature of loan contracts. Many contracts give borrowers lines of credit—that is, access to amounts of credit they can draw on as they wish. Both borrowers' fears about the health of their lenders and the deteriorating condition of the economy could have made firms want to draw on their credit lines after the failure of Lehman. Thus perhaps the sharp rise in lending reflects only firms' use of existing credit lines and is not informative about the availability of new loans.

Unfortunately, there are no comprehensive data on drawdowns of credit lines. Ivashina and Scharfstein therefore take several indirect approaches to getting information on the importance of drawdowns in this period. One is to search for media reports of firms drawing on their credit lines. They find no such reports in the three months before mid-August 2008, but many between mid-August and December. The sum of the reported drawdowns is roughly one-quarter of the overall rise in commercial and industrial lending over this period. Because there are surely many drawdowns that are not reported, this evidence, although clearly not definitive, is consistent with the view that the large rise in lending largely reflected drawdowns. And it is fatal to the argument that the large rise in lending shows that there could not have been a fall in credit supply.

Ivashina and Scharfstein's main interest, however, is in the question of whether the financial disruptions in this period affected banks' lending. Again, they rely on institutional knowledge, this time to identify two variables likely to be correlated with differences among banks in how the crisis affected their ability to obtain funds. The first variable is based on the fact that banks have two main sources of funds: retail deposits and wholesale short-term debt. Deposits are usually stable, while the ability to issue short-term debt can change quickly. Thus, Ivashina and Scharfstein argue, banks where short-term debt was a larger share of total liabilities were likely to have had more trouble continuing to obtain funds after Lehman's collapse.

The second variable is based on the fact that many large loans and lines of credit are made jointly by multiple banks. Ivashina and Scharfstein argue that banks that had more such *cosyndicated* credit lines with Lehman were likely to see more credit-line drawdowns, both because Lehman's bankruptcy meant they were now responsible for a larger fraction of lending under the credit lines, and because the fact that one of the lenders had failed might make borrowers more nervous about the other lenders. With more credit-line drawdowns, these banks would have fewer funds available to make other loans.

Ivashina and Scharfstein therefore examine whether financial institutions that obtained a larger fraction of their funds from wholesale sources and that had a larger fraction of their lending in the form of loans cosyndicated with Lehman reduced their new lending by more when the crisis hit. They find that they did. Looking at the change in lending from the pre-crisis period August 2006–July 2007 to the crisis period August 2008–November 2008, they find that both variables are associated with statistically significant and economically large reductions in lending. The point estimates imply that a bank whose fraction of liabilities coming from wholesale deposits was one standard deviation larger than average reduced its lending by about 30 percentage points more than an average bank. The estimated difference between a bank with an exposure to Lehman one standard deviation greater than average and an average bank is similar.

The most obvious concern with Ivashina and Scharfstein's evidence is the possibility that there were systematic differences in the types of firms that banks lent to that were correlated with their reliance on wholesale deposits and their exposure to Lehman. For example, perhaps banks that relied more on wholesale funding or that cosyndicated more loans with Lehman made more real-estate loans, or loans to riskier borrowers. If so, they might have reduced their lending by more during the crisis not because it was harder for them to supply credit, but because they had fewer attractive lending opportunities. Ivashina and Scharfstein present a range of evidence against this possibility. Most notably, they show that the borrowers of the different types of banks look similar on a range of dimensions. Of course, they cannot rule out the possibility that there are important differences on dimensions they cannot observe. But the most plausible interpretation of their results is that the crisis disrupted many banks' ability to make loans.

As Ivashina and Scharfstein note, a deeper limitation of their findings is that even incontrovertible evidence that the crisis reduced credit supply by some banks would not establish that those reductions had important consequences. The crisis could have merely rearranged lending as borrowers switched from banks more affected by the crisis to less affected ones. And even if overall credit supply was affected, perhaps credit supply was not important to firms' employment and production decisions. Thus Ivashina and Scharfstein's analysis only partly addresses the issue of the economic consequences of financial crises.

Chodorow-Reich's Evidence on Employment Effects

Chodorow-Reich (2014) builds on Ivashina and Scharfstein's insights to take their analysis further. His work illustrates two methods that have contributed enormously to a great deal of microeconomically based empirical work in macroeconomics (and in other fields): the merging of disparate data sets, and the use of administrative rather than survey data.

Chodorow-Reich's goal is to determine whether firms connected to banks that reduced their credit supply by more in the crisis cut their employment by more. Answering this question is not easy. In fact, Chodorow-Reich's analysis involves five distinct steps.

The first step is to determine whether it even makes sense to talk about firms being connected to specific banks: if firms can switch easily among lenders, then any ties between specific firms and banks are unlikely to be important. The fact that asymmetric information appears central to many financial relationships suggests that there could be important ongoing lender-borrower relationships, but fundamentally the question is empirical. To address it, Chodorow-Reich asks whether there is important inertia in lender-borrower ties. Specifically, he asks whether a bank is more likely to be the lead lender on a syndicated loan to a firm if it was the lead lender on the firm's previous syndicated loan; and he asks the analogous question about a bank that is a participant in a syndicated loan but not the lead lender. The results of this step are clear-cut: Chodorow-Reich finds overwhelming evidence of very large inertia in lender-borrower relationships.

The second step is to address a concern analogous to one we just discussed in the context of Ivashina and Scharfstein's analysis. If some banks got into trouble *because* of the types of firms they were lending to, observing worse outcomes at firms connected with banks that got into trouble would not tell us about the impact of bank health on firms. Chodorow-Reich argues, however, that this concern is unlikely to be important in the 2008 financial crisis. There is a great deal of evidence that the causes of banks' difficulties in this period were largely unrelated to their lending to firms; most of the variation in bank health came from variation in their real-estate lending, for example through their holdings of mortgage-backed securities. Chodorow-Reich's starting point is therefore to view all the variation across banks in the change in their lending to firms during the crisis as reflecting variations in loan supply. But he also considers specifications in which he controls for numerous firm characteristics, and ones in which he uses measures of banks' real-estate exposure as instruments for their corporate lending. The results are similar across the various specifications, which is consistent with his argument that systematic differences in the types of firms that banks lent to were not important.

Step three—which may be the most important—is to construct the data needed for relating employment outcomes to lending relationships. The challenge is that Chodorow-Reich has separate data on syndicated loans, which are gathered by a private company, and on employment by firm, which are collected by the Bureau of Labor Statistics (BLS). The BLS data are at the level of establishments rather than firms, and so they have to be aggregated to the firm level. In addition, firms often have somewhat different names in the two data sets. Chodorow-Reich links the data sets using several different methods, including manually checking hundreds of cases where the match is not clear-cut.[20] He is ultimately able to match about half the firms in the loans data to the employment data.

The fourth step is to perform the analysis. We already know from Ivashina and Scharfstein that banks in worse health reduced their lending by more during the crisis. But as we discussed, their analysis does not rule out the possibility that their borrowers simply switched to other banks. Chodorow-Reich therefore begins his analysis of outcomes by asking whether firms that had been borrowing from banks that cut their lending by more in the crisis were less likely to obtain a loan from any bank. He finds strong evidence that they were. Then, turning to the key variable—employment—he finds an impact of reductions in credit supply to a firm on its employment that is statistically significant. Moreover, the economic magnitude is large: the point estimates imply that employment growth at a firm that had been borrowing from a syndicate at the tenth percentile of syndicate health (in terms of the loan growth of its constituent banks) is four percentage points lower than employment growth at a firm that had been borrowing from a syndicate at the ninetieth percentile of syndicate health.

Chodorow-Reich also examines heterogeneity in these effects. He finds that the employment effects are much larger at smaller firms and at firms that did not have access to the bond market (as measured by having a credit rating or having ever issued publicly traded debt). These are precisely the types of firms one would expect to have the most difficulty obtaining other funding if they were unable to borrow from their usual lenders. Thus the results about heterogeneity both help show the mechanism by which the baseline results come about and increase our confidence in those results.

Chodorow-Reich's final (and most tentative) step is to try to estimate the overall effects of the fall in credit supply to firms. He faces three major challenges. First, his results concern only a subset of firms, and those firms may not be representative of the entire economy. Second, his results are relative:

[20] A further complication is that the BLS data are confidential and their use by researchers is subject to numerous restrictions, including that they cannot be accessed remotely or taken off-site. An essential part of Chodorow-Reich's research was spending many days in the basement of the Department of Labor building in Washington, D.C.

his evidence is that if one firm faced a larger reduction in credit supply than another, its employment fell by more. But the overall effects depend on absolute impacts, not relative ones. Third (and somewhat related to the second challenge), there can be general-equilibrium effects. A reduction in employment and production at one firm can increase employment at other firms if it drives down prevailing wages, or if it causes demand for other firms' products to increase as consumers shift away from the firm. Alternatively, the fall in income from one firm's decline in production can reduce demand for other firms' products, and so decrease employment elsewhere.

Chodorow-Reich takes only a first pass at these issues. He argues that for reasonable parameter values, the negative general-equilibrium effects are probably larger than the positive ones, and thus that looking only at partial-equilibrium effects is likely to yield a conservative estimate of the aggregate effects. He then considers the implications of assuming that the small and medium-sized firms in his sample (which are where he finds the main employment effects) are representative of small and medium-sized firms in the economy, and that a firm that had been borrowing from a syndicate at the ninetieth percentile of syndicate health faced no change in its credit supply (which is probably a conservative estimate of the decline in credit supply in the 2008 crisis). He calculates that in this case, the shift back in credit supply to firms was the source of about 20 percent of the overall fall in U.S. employment in the year after Lehman's failure. The results are sensitive to the specifics of his various simplifications, however; and even accepting the simplifications, the confidence interval associated with this estimate is fairly wide. Importantly, however, this estimate involves only one type of credit to one part of the economy; most notably, it concerns only credit to firms, not households. And as we have just discussed, the estimate probably errs on the low side. Thus Chodorow-Reich's results are very supportive of the conventional view that disruptions of credit availability were central to the Great Recession.

Discussion

As discussed in Section 10.3, many studies find that financial-market imperfections affect firms' behavior, and Section 8.6 describes similar evidence for consumers. And Chodorow-Reich's paper is not the only one to find that credit disruptions affect microeconomic outcomes in times of financial crisis. Other examples include Campello, Graham, and Harvey (2010); Amiti and Weinstein (2011); Almeida, Campello, Laranjeira, and Weisbenner (2012); Edgerton (2012); Ziebarth (2013); Duygan-Bump, Levokov, and Montoriol-Garriga (2015); Mondragon (2015); and Benmelech, Meisenzahl, and Ramcharam (2017).

At the same time, as we have just seen, moving from microeconomic evidence to macroeconomic estimates is difficult. The types of calculations performed by Chodorow-Reich require numerous assumptions whose accuracy is difficult to assess. Thus there is also value in looking at more aggregate evidence.

One approach is to look not at the level of individual firms or households, but at regions within a country. Peek and Rosengren (2000) provide an early example of this approach. They examine the effects of Japan's financial crisis in the 1990s on U.S. states. They find that the crisis in Japan reduced lending in the United States by subsidiaries of Japanese banks, and, more importantly, that construction activity fell more in states where Japanese banks provided a larger fraction of lending. A more recent study looking at the regional level is Huber (2017). He examines variation across German counties in the dependence of their firms on a major bank that suffered large losses in the 2008 crisis from its international exposure that were unrelated to its domestic lending. He finds county-level effects that are not only large, but substantially larger than one would obtain by aggregating estimated firm-level effects. Thus his results point to large negative general-equilibrium effects at the county level.

Ultimately, of course, much of our interest is in effects at the level of the entire economy; and just as extrapolating from effects at the level of individual firms and households to regional outcomes is difficult, so is extrapolating from regional to economy-wide effects. Thus there is value in looking directly at the aggregate evidence. Unfortunately, as we discussed at the start of this section, identifying the causal effects of crises using aggregate evidence is extremely challenging. Research using aggregate data has found that financial crises are on average followed by large declines in output and that there is considerable heterogeneity across crises, but thus far has had limited success in determining the causal impact of crises.[21]

Problems

10.1. Consider the model of Section 10.1. Assume that utility is logarithmic, that $\beta = 1$, and that there are only two states, each of which occurs with probability one-half. In addition, assume there is only one investment project. It pays R_G in state G and R_B in state B, with $R_G > R_B > 0$. We will refer to G as the "good" state and B as the "bad" state.

[21] Some examples of research in this area are Bordo, Eichengreen, Klingebiel, and Martinez-Peria (2001); Hoggarth, Reis, and Saporta (2002); Cerra and Saxena (2008); Reinhart and Rogoff (2009); Gilchrist and Zakrajsek (2012); Jordà, Schularick, and Taylor (2013); and C. Romer and D. Romer (2017). Jalil (2015) uses information from newspapers to identify a set of panics in the nineteenth-century United States that appear to have been due largely to idiosyncratic factors, and that may therefore shed light on the causal effect of crises. He finds large falls in output following these panics. Unfortunately, it is hard to determine the *external validity* of this finding (that is, the extent to which it applies to other settings, such as modern advanced economies).

(*a*) What are the equilibrium conditions?

(*b*) What are C_1, K, C_2^G, C_2^B, q_G, and q_B?

(*c*) Now suppose a new investment project is discovered. It pays off only in one state of the world. Let $R_{NEW} > 0$ denote its payoff in that state.

(*i*) What is the condition for there to be strictly positive investment in the new project?

(*ii*) Assume the condition in (*i*) is satisfied. What are the equilibrium levels of K (investment in the old project) and K_{NEW} (investment in the new project)?

(*iii*) Suppose the state in which the new project pays off is B. What is the condition for C_2^B to be greater than or equal to C_2^G, so that it is no longer reasonable to describe B as the "bad" state?

10.2. Consider the model of Section 10.1. Suppose, however, that there are M households, and that household j's utility is $V_j = U(C_1) + \beta_j^s U(C_2)$, where $\beta_j^s > 0$ for all j and s. That is, households may have heterogeneous preferences about consumption in different states.

(*a*) What are the equilibrium conditions?

(*b*) If the β's differ across households, can a situation where each agent owns an equal fraction of the claims on the output of each investment project, so that C_{2j}^s for a given s is the same for all j, be an equilibrium? Why or why not?

10.3. Consider the model of investment under asymmetric information in Section 10.2. Suppose that initially the entrepreneur is undertaking the project, and that $(1 + r)(1 - W)$ is strictly less than R^{MAX}. Describe how each of the following affects D:

(*a*) A small increase in W.

(*b*) A small increase in r.

(*c*) A small increase in c.

(*d*) Instead of being distributed uniformly on $[0, 2\gamma]$, the output of the project is distributed uniformly on $[\gamma - b, \gamma + b]$, and there is a small increase in b.

(*e*) Instead of being distributed uniformly on $[0, 2\gamma]$, the output of the project is distributed uniformly on $[b, 2\gamma + b]$, and there is a small increase in b.

10.4. A simpler approach to agency costs: limited pledgeability. (Lacker and Weinberg, 1989; Holmström and Tirole, 1998.) Consider the model of Section 10.2 with a different friction: there is no cost of verifying output, but the entrepreneur can hide fraction $1 - f$ of the project's output from the investor (with $0 \le f \le 1$). Thus the entrepreneur can only credibly promise to repay fraction f of the project's output.

(*a*) Consider a project with expected payoff γ that exceeds $1 + r$. What is the condition for the project to be undertaken?

(*b*) Suppose the condition you found in part (*a*) is satisfied with strict inequality. Is the contract between the investor and the entrepreneur uniquely determined? If so, what is the contract? If not, explain why.

(*c*) Limited pledgeability leads to inefficiency (relative to the case of no frictions) if $\gamma > 1+r$ but the project is not undertaken. Describe whether each of the following can cause a project with $\gamma > 1+r$ not to be undertaken:

(*i*) A fall in the entrepreneur's wealth, W.

(*ii*) An increase in the fraction of the project the entrepreneur can hide, $1-f$ (that is, a fall in f).

(*iii*) An increase in idiosyncratic risk. Concretely, suppose that (as in part (*d*) of Problem 10.3), the output of the project is distributed uniformly on $[\gamma - b, \gamma + b]$ rather than uniformly on $[0, 2\gamma]$, and there is an increase in b.

10.5. (*a*) Show that in the model analyzed in equations (10.15)–(10.23) of Section 10.4, the unconditional distributions of C_{2t}^a and C_{2t}^n are not normal.

(*b*) Explain in a sentence or two why the analysis in the text, which uses the properties of lognormal distributions, is nonetheless correct.

10.6. Fundamental risk and noise-trader risk. Consider the following variant on the model of noise-trader risk in equations (10.15)–(10.23). There are three periods, denoted 0, 1, and 2. There are two assets. The first is a safe asset in perfectly elastic supply. Its rate of return is normalized to zero: one unit of the economy's single good invested in this asset in period 0 yields one unit of the good for sure in period 1, and one unit of the good invested in this asset in period 1 yields one unit for sure in period 2. The second is a risky asset. Its payoff, which is realized in period 2, is $1 + F_1 + F_2$, where F_t is distributed normally with mean 0 and variance V_t^F. F_1 is observed in period 1, and F_2 is observed in period 2. This asset is in zero net supply. Thus equilibrium requires that the sum across agents of the quantity of the asset demanded is zero.

There are two types of traders. The first are noise traders. They demand quantity N_0 of the risky asset in period 0 and $N_0 + N_1$ in period 1, where N_0 is exogenous and N_1 is distributed normally with mean 0 and variance V_1^N. F_1, F_2, and N_1 are independent. The second are arbitrageurs. A_0 are born in period 0 and A_1 are born in period 1. They live for two periods (0 and 1 for those born in period 0; 1 and 2 for those born in period 1). They only value consumption in the second period of their life and have utility $U(C) = -e^{-2\gamma C}, \gamma > 0$. They have no initial wealth.

(*a*) Consider first period 1.

(*i*) Consider a representative arbitrageur born in period 1. What is his or her second-period consumption as a function of P_1, F_1, and F_2, and his or her purchases of the risky asset, X_1^a? What is the mean and variance of his or her second-period consumption as a function of P_1, F_1, X_1^a, and V_t^F? What is the first-order condition for his or her choice of X_1^a?

(*ii*) What is the condition for equilibrium in period 1?

(*iii*) Use the results in (*i*) and (*ii*) to find an expression for $P_1 - (1 + F_1)$, the departure of the price in period 1 from its fundamental value.

(*iv*) Do your results support the statement in the text that greater fundamental risk mutes sophisticated investors' willingness to trade to offset departures of asset prices from fundamentals, and so leads to larger departures of asset prices from fundamentals?

(*b*) Now consider period 0.

(*i*) What is the first-order condition for X_0^a (purchases of the risky asset by the representative sophisticated investor) in terms of $E_0[P_1]$ and $\mathrm{Var}(P_1)$ and the parameters of the model?

(*ii*) Use the results from part (*a*) to find $E_0[P_1]$ and $\mathrm{Var}(P_1)$ in terms of exogenous parameters.

(*iii*) Use the results in (*i*) and (*ii*) to find an expression for $P_0 - 1$, the departure of the price in period 0 from its fundamental value.

(*iv*) Do increases in fundamental risk (V_1^F and V_2^F) increase departures of asset prices from fundamentals? Do increases in noise-trader risk? Are there interactions—that is, does an increase in noise-trader risk increase, decrease, or have no effect on the effect of fundamental risk?

10.7. A simple model of agency risk. Consider the previous problem. For simplicity, assume $A_0 = 0$. Now, however, there is a third type of agent: hedge-fund managers. They are born in period 0 and care only about consumption in period 2. Like the sophisticated investors, they have utility $U(C) = -e^{-2\gamma C}$ and no initial wealth. There are A_H of them. They participate in the market for the risky asset in period 0, and do not make any additional trades in period 1. However, they face a cost if they incur short-term losses, and gain a reward if they obtain short-term gains. Specifically, if a hedge-fund manager purchases amount H of the risky asset, he or she receives $aH(P_1 - E_0[P_1])$ in period 1, where $a > 0$ and where P_1 is the price of the risky asset in period 1. The manager then holds this payment in the safe asset from period 1 to period 2, and so it adds to (or subtracts from) his or her period-2 consumption.

(*a*) Consider first period 1. Find an expression for $P_1 - (1 + F_1)$ taking the period-0 purchases of the hedge-fund managers, $X_0^h A_H$, as given. (Hint: In period 1, the demand of the hedge-fund managers is fixed and does not respond to the price of the asset. As a result, in period 1 their demand enters in the same way as that of the noise traders.)

(*b*) Now consider period 0.

(*i*) Find an expression for the representative hedge-fund manager's period-2 consumption as a function of $P_0, F_1, F_2, a, P_1 - E_0[P_1]$, and X_0^h.

(*ii*) What are the mean and variance of his or her second-period consumption as a function of P_0, a, V_1^F, V_2^F, γ, A_1, V_2^N, and X_0^h?

(*iii*) What is the first-order condition for his or her choice of X_0^h?

(*iv*) Use your results to find an expression for $P_0 - 1$.

(*v*) Does greater agency risk (a higher value of a) increase the impact of the noise traders in period 0 on the price of the asset?

10.8. Consider Problem 10.6. Suppose, however, that the demand of the period-0 noise traders is not fully persistent, so that noise traders' demand in period 1 is $\rho N_0 + N_1, \rho < 1$. How, if at all, does this affect your answer in part (*b*)(*iii*) of the problem for how the noise traders affect the price in period 0? What happens if $\rho = 0$?

10.9. This problem asks you to show that with some natural variants on the approach to modeling agency risk in Problem 10.7, consumption is not linear in the shocks, which renders the model intractable.

(*a*) Consider the model in Problem 10.7. Suppose, however, that the representative hedge-fund manager, rather than receiving a payment or incurring a cost in period 1, is forced to sell quantity $b(E_0[P_1] - P_1)H, b > 0$, of the risky asset in period 1. Show that in this case, the manager's consumption is not linear in F_1.

(*b*) Consider the model in Problem 10.7. Suppose, however, that A_1 is not exogenous but depends on the success of the period-0 sophisticated investors: $A_1 = \overline{A} + b(P_1 - E_0[P_1])X_0^a, b > 0$. Show that in this case, the consumption of the sophisticated investors born in period 0 is not linear in F_1.

10.10. Prices versus quantities in the DeLong–Shleifer–Summers–Waldmann model.[22] Consider modeling the noise traders in the model of equations (10.15)–(10.23) of Section 10.4 in terms of shocks to the quantity they demand of the risky asset rather than to their expectations of the price of the asset. Specifically, suppose the demand of a representative noise trader is $X_t^a + \omega_t$, where X_t^a is the demand of a representative arbitrageur (see equation [10.17]), and ω_t is an i.i.d., mean-zero, normally distributed shock with variance V_ω. Intuitively, one might expect that: (1) there is some value of V_ω that yields the same equilibrium as that of the model of Section 10.4 (equation [10.18] seems to suggest this—but note that V is not a primitive parameter of the model); (2) the V_ω that does this is an increasing function of V_η (intuitively, both a higher V_ω and a higher V_η correspond to "noisier" noise traders); and (3) for any V_ω, there is also an equilibrium where $P_t = 1$ for all t (see n. 8). Are (1), (2), and (3) in fact all true? If not, try to explain intuitively the economics of why one or more of these conjectures is wrong.

10.11. Consider the Diamond–Dybvig model described in Section 10.6, but suppose that $\rho R < 1$.

(*a*) In this case, what are c_1^{a*} and c_1^{b*}? Is c_1^{b*} still larger than c_1^{a*}?

[22] This problem provides less of a step-by-step guide to working out its solution than most other problems, and as a result is more challenging. (Indeed, when I first assigned this problem to my students, I was not entirely sure of the answer.) One goal is to move in the direction of showing you what it is like to tackle a model whose implications are unknown rather than working through a structured problem.

(*b*) Suppose the bank offers the contract described in the text: anyone who deposits one unit in period 0 can withdraw c_1^{a*} in period 1, subject to the availability of funds, with any assets remaining in period 2 divided equally among the depositors who did not withdraw in period 1. Explain why it is not an equilibrium for the type a's to withdraw in period 1 and the type b's to withdraw in period 2.

(*c*) Is there some other arrangement the bank can offer that improves on the autarchy outcome?

10.12. Consider deposit insurance in the Diamond–Dybvig model of Section 10.6.

(*a*) If fraction $\phi > \theta$ of depositors withdraw in period 1, how large a tax must the government levy on each agent in period 1 to be able to increase the total consumption of the nonwithdrawers in the two periods to c_2^{b*}? Explain why your answer should simplify to zero when $\phi = \theta$, and check that it does.

(*b*) Suppose the tax is marginally less than the amount you found in part (*a*). Would the type b's still prefer to wait until period 2 rather than try to withdraw in period 1?