10.6 The Diamond-Dybvig Model

A key characteristic of financial markets is that they are subject to sudden, convulsive changes. Such 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, such sudden changes 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). 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 show how such 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 only want to consume 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.

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

$$U^a = \ln c_1^a, \qquad (10.D1)$$

$$U^b = \rho \ln(c_1^b + c_2^b), \qquad (10.D2)$$

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 project to period 2 and have period-1 consumption of 0 and period-2 consumption of *R*. Thus expected utility under autarchy is

$$\begin{aligned} U^{AUTARCHY} &= \theta \ln 1 + (1-\theta)\rho \ln R \\ &= (1-\theta)\rho \ln R. \end{aligned} \tag{10.D3}$$

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 c_1^a (period-1 consumption per individual who consumes in period 1) times θ (the fraction of individuals who consume 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

$$c_2^b = \frac{(1 - \theta c_1^a)R}{1 - \theta}. \tag{10.D4}$$

A representative individual's expected utility is $\theta \ln c_1^a + (1 - \theta)\rho \ln c_2^b$. Using the budget constraint, (10.D4), to substitute for c_2^b, we can write this as

$$\mathrm{E}[U] = \theta \ln c_1^a + (1 - \theta)\rho[\ln(1 - \theta c_1^a) + \ln R - \ln(1 - \theta)]. \tag{10.D5}$$

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 above, the autarchy outcome is $c_1^a = 1$. Thus,[9]

[9] [NOTE TO COPYEDITOR/COMPOSITOR: SEE EQ. (1.31) ON P. 26 OF THE FOURTH EDITION FOR A SENSE OF HOW (10.D6) SHOULD BE FORMATTED.]

$$\frac{\partial \mathrm{E}[U]}{\partial c_1^a}\Big|_{c_1^a = c_1^{a,AUTARCHY}} = \frac{\partial \mathrm{E}[U]}{\partial c_1^a}\Big|_{c_1^a = 1}$$

$$= \theta + \frac{(1-\theta)\rho}{1-\theta}(-\theta)$$

$$= \frac{(1-\theta)\theta - \theta(1-\theta)\rho}{1-\theta} \tag{10.D6}$$

$$= (1-\rho)\theta$$

$$> 0.$$

It is easy to check that $\partial^2\mathrm{E}[U]/\partial {c_1^a}^2 < 0$. 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.D5) implies that the first-order condition for the optimal level of c_1^a under full information is

$$\frac{\theta}{c_1^{a*}} + \frac{(1-\theta)\rho}{1-\theta c_1^{a*}}(-\theta) = 0, \qquad \text{(10.D6½)}$$

which implies

$$c_1^{a*} = \frac{1}{\theta + (1-\theta)\rho}. \tag{10.D7}$$

$$> 1.$$

And substituting this expression into (10.D4) gives

$$c_2^{b*} = \frac{\rho R}{\theta + (1-\theta)\rho}. \qquad (10.D8)$$

$$< R.$$

Notice that although c_2^{b*} is less than its level under autarchy, it is greater than c_1^{a*}.

A Bank

One of Diamond and Dybvig's key insights 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.[10]

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

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

than $1/c_1^{a*}$) ask 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 that 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 when only the type *a*'s withdraw in period 1 is

$$c_2 = \frac{(1-\theta c_1^{a*})R}{1-\theta}$$

(10.D9)

$$= c_2^{b*},$$

where the second line uses the economy's resource constraint, (10.D4). 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. Thus there is a Nash equilibrium where the economy attains the first best even though individuals' types are unobserved and there is no government intervention.

Notice that the bank is providing something that can be described as liquidity: it makes long-term investments but allows depositors to access funds before the investments mature. Moreover, 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 to allow for early withdrawals. Depositors who withdraw early are paid interest even though the bank earns no interest on its holdings of cash.

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.[11]

[11] This analysis ignores one complication. Because utility is logarithmic (see [10.D1] and [10.D2]), the fact that individuals face some chance of having zero consumption when there is a run means that their expected utility in the

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, and 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

---

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.

Fortunately, there are at least two straightforward 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 an equilibrium with a bank that attracts deposits and where there is a positive probability of a run can exist, 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.D1) and (10.D2) 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.

underlying assets they are investing in is greatly reduced. In 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 many 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). Then a run may take the form of many wholesale investors simultaneously refusing to roll over their loans or 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 things 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 the bank, 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 off 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.

This discussion raises two issues. The first is how the central bank's offer to lend resources to the 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.

A second (and 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

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

$$= \frac{c_2^{b*}/c_1^{a*}}{R} \qquad (10.D10)$$

$$< 1,$$

where the last 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 above, being willing to lending 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 $(1-\theta)c_1^{b*}$ to the period-2 withdrawers, leaving it with a profit of $R - c_2^{b*}$ at the expense of the central bank.[12]

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.

[12] 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.

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 response 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 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. The economy's resource constraint makes it impossible for all individuals to have c_1^{a*} in period 1. Thus there is no private entity that can provide insurance or make loans to the entire financial system. A government with the power to use taxes or create inflation is needed.