

Banks, Runs and Liquidity Creation:

Chapter 2 of *Banks, Liquidity and Legal Protection*

Douglas W. Diamond

University of Chicago, G.S.B. and N.B.E.R.

January 2007, revised August 2007.

I am grateful to Elizabeth Cammack for many helpful comments. I gratefully acknowledge financial support from the National Science Foundation and the Center for Research in Security Prices at the University of Chicago, GSB.

© Douglas W. Diamond, 2007

This is based on the 2005 Princeton Lectures in Finance, which I presented in June, 2005.

This chapter explores the links between loan monitoring, legal sanctions for default, liquidity and the threat of runs on banks and other borrowers. Banks make loans which cannot be sold quickly at a high price. Banks issue demand deposits which allow depositors to withdraw at any time. This mismatch of liquidity, where a bank's liabilities are more liquid than its assets, has caused problems for banks when too many depositors attempt to withdraw (a situation referred to as a bank run). Banks have followed policies to stop runs and governments have instituted deposit insurance to prevent runs.

Diamond and Dybvig [1983] explains why banks choose to issue such deposits that are more liquid than their assets and analyzes some reasons why banks are subject to runs. The model has been widely used to understand bank runs and other types of financial crises, as well as ways to prevent such crises. Diamond and Dybvig [1983] argues that an important function of banks is to create liquidity, that is, to offer deposits that are more liquid than the assets they hold. Investors who have a demand for liquidity will prefer to invest via a bank, rather than hold assets directly. Investors demand liquidity because they are uncertain about when they need to consume and thus how long they wish to hold assets. As a result, they evaluate the proceeds of liquidating the position on the various possible dates, rather than on a single date. Section I describes this demand for liquid assets. Section II describes creating liquidity in the Diamond-Dybvig model.

The definition of an illiquid asset in this chapter is one where the liquidation proceeds available from physical liquidation or sale are less than the present value of its future payoffs. Chapter 3 also considers a more general definition that accounts for endogenous changes in the present values of the cash flows from some assets. In the

current chapter all the cash flows from all assets are certain, allowing the analysis to focus on the liquidity risks of assets. In Sections I and II, which are based on the Diamond-Dybvig [1983] model, asset illiquidity is exogenous. In the balance of the chapter, asset liquidity and a borrower's financial capacity to pay lenders are endogenous. The endogenous determination of these is based on the model of monitoring and legal protection from Chapter 1. There is no uncertainty and any diversion by a borrower is assumed to be reversible, allowing unfettered negotiation over the consequences of diversion.

Section III shows how delegated monitoring, described in chapter 1, can lead to the illiquidity of financial assets. The Diamond-Dybvig model focuses on the consequences of illiquid assets but is silent on the reasons why they are illiquid: illiquidity is assumed due to production processes that provide a low payoff if terminated early.

Section IV shows how illiquidity of bank loans made by delegated monitors may require the first-come-first-served demand deposit contracts that were assumed in the Diamond-Dybvig model. The analysis combines ideas from Diamond-Rajan [2001] and Diamond [2004] with the model from Chapter 1. The delegated monitoring model used in Chapter 1 and Section III of this chapter assume contracts that commit to impose legal sanctions contingent on low payments to lenders or depositors. In addition, they assume that monitors have all the bargaining power over borrowers: they capture the entire surplus when they use their monitoring to threaten to stop a crime in progress. As a result, monitoring is effective in deterring diversion if stopping a crime in progress can sufficiently reduce the spoils of diversion. Section IV allows for more equal bargaining

power and for legal sanctions on borrowers that must be imposed voluntarily by lenders or monitors (allowing for sanctions that are not automatically triggered by a default). Demand deposits provide depositors (or lenders more generally) an incentive to invoke sanctions that synthesize automatic sanctions. I show that although loans made by an undelegated monitor will be illiquid, the deposits of a bank which is a delegated monitor will be liquid.

Section V examines the financial structure of an economy where investors can lend directly with or without monitoring, and where monitoring can be delegated, and describes the use of short-term demandable debt by firms or banks. Section VI relates the results to previous literature, especially to Calomiris-Kahn [1991] and Diamond-Rajan [2001]. Section VII concludes this chapter.

## **I. The Demand for Liquidity**

This section first analyzes some important reasons for the demand for more liquid assets by investors who are consumers. It then provides an alternative motivation for liquid assets by entrepreneurs. When the assets that investors can hold directly are illiquid, there is a demand for creating more liquid assets.

An asset when the proceeds available from physical liquidation or sale on some date are less than the present value of its payoffs on some future dates. In the extreme, a totally illiquid asset is worthless (cannot be sold or physically liquidated for a positive amount) on some date but has a positive value on a later date. The lower the fraction of the present value of the future cash flow that can be obtained today, the less liquid is the asset. This definition is a sufficient condition for an asset to be illiquid, but I will describe more complicated situations in chapter three where changes in the discount rates

or present value (on an asset specific or systemic basis) can make assets vary in liquidity. The current definition holds discount rates constant and is the clearest notion of illiquidity.

Consider the following asset on three dates,  $T=0$ ,  $T=1$  and  $T=2$ . If one invests one unit at date 0, it will be worth  $r_2$  at date 2, but only  $r_1 < r_2$  at date 1. The lower is  $r_1/r_2$ , (holding constant discount rates) the less liquid is the asset.

### **I.A. The Uncertain Horizon of Investors**

Investors face an uncertain horizon to hold the asset. Each will need to consume either at date  $T=1$  or  $T=2$ . However, as of date 0, an investor does not know which date he or she will need to consume. Each begins with 1 to invest on date  $T=0$ . An investor cannot buy direct insurance against his or her need for liquidity because the need is their private information. However, contracts can be designed that indirectly provide this insurance by offering assets that are more liquid and offer a higher ratio  $r_1/r_2$  and result in a smaller loss when liquidated early. As a result, for some risks, investors can save and liquidate assets as needed. I will call an investor a "type 1" if he needs to liquidate at  $T=1$ , and "type 2" if he can wait until  $T=2$ . For this example, this means he will consume only at  $T=1$  if of type 1, and only at  $T=2$  if of type 2 (or he can store date 1 consumption goods and consume them at date 2).

As of date 0, an investor does not know which type he or she will be, but each investor has a probability  $t$  of being of type 1 and  $1-t$  of being of type 2. There is no aggregate uncertainty, and there will be a fraction  $t$  of investors of type 1. To be concrete, suppose that  $t = 1/4$  and that there are 100 investors. As a result, 25 will be of type 1, and 75 will be of type 2, but it is not known at date 0 which investors will be of

each type.

A type 1 investor with utility function  $U(c)$  who consumes  $c_1$  at  $T=1$  has utility  $U(c_1)$ . A type 2 investor who consumes  $c_2$  at  $T=2$  (this  $c_2$  includes any stored date-1 consumption goods) has utility  $U(c_2)$ . The utility function is the same for both types, but the date on which an investor wishes to consume depends on his type. An investor who holds the asset  $(r_1, r_2)$ , which gives a choice of  $r_1$  at date 1 or  $r_2 > r_1$  at date 2 consumes  $c_1 = r_1$  if of type 1 (with probability  $t$ ) or  $c_2 = r_2$  if of type 2 (with probability  $1-t$ ). The investor's expected utility is given by:

$$tU(r_1) + (1-t)U(r_2).$$

I assume that the investor has the constant relative risk aversion utility function of  $U(c) = -1/c$ . To simplify exposition, I add a constant of one to the utility (with no effect on any decisions) and use the utility function  $U(c) = 1 - (1/c)$ . This allows the expected utility calculations to yield positive numbers.

### **I.B. Comparing More And Less Liquid Assets**

Consider the following two assets, both of which cost 1 at date 0. The illiquid asset has  $(r_1=1, r_2=R)$ , and a more liquid asset with  $(r_1 > 1, r_2 < R)$ . Investors only have access to the illiquid asset. Later, I will show how banks can create the more liquid asset, although there is no physical asset with its payoffs, but for now I simply illustrate the demand for liquidity with the following numerical example. The illiquid asset has  $(r_1=1, r_2=R=2)$ . As a comparison, consider a hypothetical more liquid asset that has  $(r_1=1.28, r_2=1.813)$ . Section II explains why these particular numerical values are used. The expected utility from holding the illiquid asset is:

$$\frac{1}{4}U(1) + \frac{3}{4}U(2) = 0.375.$$

The expected utility of holding the more liquid asset is:

$$\frac{1}{4}U(1.28) + \frac{3}{4}U(1.813) = 0.391 > 0.375.$$

Each investor prefers the more liquid asset. A risk averse investor prefers this smoother pattern of returns: holding the illiquid asset is risky because it delivers a low amount when liquidated early, on date 1.

Note that if investors were not risk averse and had constant marginal utility of consumption, they would not prefer this particular liquid asset. That is, if  $U(c)=c$ , then the expected utility of holding any asset is equal to its expected payoff given the policy of liquidating when of type 1. For the illiquid asset, the expected payoff is:

$$\frac{1}{4}(1) + \frac{3}{4}(2) = 1.75.$$

The more liquid asset gives expected payoff of'

$$\frac{1}{4}(1.28) + \frac{3}{4}(1.813) = 1.68 < 1.75.$$

The more liquid asset has a lower expected rate of return. Sufficiently risk averse investors, but not risk neutral investors, are willing to give up some expected return to get a more liquid asset.

Investors have state dependent utility and sell for a reason that they have not purchased insurance against. In particular, a type 1 investor liquidates the asset at a time when the proceeds that he or she receives are especially valuable, because marginal utility of consumption is high. An investor's demand for liquidity is greater the higher is his or her (relative) risk aversion because to liquidate early implies low consumption and thus high marginal utility of consumption.

### **I.C. Entrepreneurial Liquidity Demand**

An alternative motivation for a large demand for liquid assets comes from an

entrepreneur who may have a sudden need to fund a very high return project at date 1 (which cannot be funded elsewhere). The entrepreneur only wishes to consume on date 2, but may choose to liquidate his assets on date 1 to fund his high return project. As a result, the entrepreneur places a higher value on date 1 liquidation proceeds in those states of nature where he chooses to liquidate early. Suppose that with probability  $t$ , the entrepreneur needs to fund the investment project, and that it returns  $\Psi > R$  per unit invested. With probability  $1-t$  he does not get this opportunity and has access only to storage (storing one unit of goods at date 1 returns one unit at date 2). The availability of the high return is private information. Consider an asset that costs 1 at date zero and offers either  $r_1$  at date 1 or  $r_2 > r_1$  at date 2. When the entrepreneur has access only to storage, he will not liquidate the asset, but when he needs to fund the high return project he will liquidate it if the project's return  $\Psi$  exceeds  $r_2/r_1$ , the rate of return from continuing to hold the asset. As of date 0, the entrepreneur values an asset that can be liquidated for  $r_1$  at date 1 or  $r_2$  at date 2, as follows:  $t r_1 \Psi + (1-t) r_2$ , if  $\Psi > r_2/r_1$ , and as  $t r_1 + (1-t) r_2$ , if  $\Psi \leq r_2/r_1$ . This is qualitatively similar to the risk averse consumer, because the entrepreneur liquidates when the value of the proceeds is very high. Suppose that  $\Psi=2.5$ ,  $R=2$ , and  $t=1/4$ , the entrepreneur then values the illiquid asset ( $r_1=1$ ,  $r_2=2$ ) as  $1/4 \Psi(1) + 3/4 2=2.125$ , and the liquid asset ( $r_1=1.28$ ,  $r_2=1.813$ ) as  $1/4 \Psi(1.28) + 3/4 (1.813)=2.160$ . The entrepreneur prefers the more liquid asset.

The entrepreneurial demand for liquidity will be even more similar to the investor/ consumer demand for liquidity if the high return project has decreasing returns to scale.

I do not continue to analyze the entrepreneurial demand for liquidity here, but

refer the reader to Diamond and Rajan [2001] and Holmström and Tirole [1998] . I now return to the consumer demand for liquidity.

## II. Bank Liquidity Creation

I now show that a bank can provide the more liquid asset by offering demand deposits, even though the bank invests in the illiquid asset ( $r_1=1$ ,  $r_2=2$ ). I assume a mutual bank without equity (purely for expositional simplicity). Suppose that in return for a deposit of 1 at date  $T=0$ , the bank offers to pay  $r_1=1.28$  to those who withdraw at  $T=1$  or to pay  $r_2=1.813$  to those who withdraw at  $T=2$ .

If the bank receives \$1 from each of the 100 investors, it gets \$100 in deposits on date  $T=0$ . If the bank invests in the illiquid asset, it will need to liquidate some of the illiquid asset at  $T=1$  to pay 1.28 to those who withdraw.

At  $T=1$ , the bank's entire portfolio can be liquidated for 100. Suppose 25 depositors withdraw 1.28 each, then  $25(1.28) = 32$  assets must be liquidated: (32% of portfolio must be liquidated). If 32 assets are liquidated, then 68 will remain until  $T=2$ , when they will be worth  $R=2$  each. On date 2, there remain 75 depositors, each of whom will receive :

$$\frac{[100-32] \cdot 2}{75} = \frac{[68] \cdot 2}{75} = 1.813$$

Depositors prefer the more liquid asset. A bank can provide a more liquid deposit (smaller loss from early liquidation) than is available from holding the assets directly. This liquidity transformation service is one of the most important functions of banks. It is an equilibrium (a Nash equilibrium) for 25 depositors to withdraw at  $T=1$ , because if all depositors expect 25 to withdraw at  $T=1$ , only type 1 depositors will withdraw because the 75 type 2 depositors prefer the 1.813 available at  $T=2$  to the 1.28 available at

T=1.

When assets are illiquid and risk averse investors do not know when they will need to liquidate, the bank can create a more liquid asset that allows investors to share the risk of liquidation losses. The bank can give a fraction  $t$  of investors  $r_1$  at date 1 and a fraction  $1-t$  of investors  $r_2 = \frac{[1-t_1]R}{1-t}$  at date 2 because if a fraction  $t$  of the depositors get  $r_1$  in period T=1, this will leave a fraction  $[1-t_1]$  of the assets unliquidated and in place until date 2. Each of the remaining fraction  $(1-t)$  of depositors can receive  $r_2 = \frac{[1-t_1]R}{1-t}$  in period 2. Note that for the illiquid asset,  $r_1=1$  and  $r_2=R$ .

## II.A. The Optimal Amount of Liquidity

It is interesting to see (but not essential to understanding the points here) that the deposit contract that gives  $r_1=1.28$  to those who withdraw at T=1 or  $r_2=1.813$  to those who withdraw at T=2 is the optimum amount of liquidity to create. The optimal amount of liquidity to create is the amount that maximizes each investor's ex-ante expected utility, choosing  $c_1=r_1, c_2=r_2$  to maximize  $tU(r_1) + (1-t)U(r_2)$ , subject to  $r_2 \leq \frac{[1-t_1]R}{1-t}$ ,  $r_1 \geq 0$ ,  $r_2 \geq 0$ .

For an interior optimum, the optimal values satisfy  $U'(r_1)=RU'(r_2)$ , so the marginal utility is in line with the marginal cost of liquidity, and  $r_2 = \frac{[1-t_1]R}{1-t}$ , because no liquidity is wasted. For the case used in the example where  $U(c)=1-1/c$ , marginal utility is  $U'(c) =$

$1/c^2$  and the condition is  $\frac{r_2^2}{r_1^2} = R$ , or  $\frac{r_2}{r_1} = \sqrt{R}$  because both  $r_1$  and  $r_2$  are positive.<sup>1</sup> From

---

<sup>1</sup> For general constant relative risk aversion utility functions  $U(c)=c^{1-\rho}/(1-\rho)$ , marginal utility is  $U'(c)=c^{-\rho}$ . The optimal  $r_1$  is greater than 1 whenever the rate of relative risk aversion,  $\rho$ , is greater than 1 (as seems to

$r_2 = \frac{[1 - t_1]R}{1 - t}$ , this becomes  $r_1 = \frac{\sqrt{R}}{1 - t + t\sqrt{R}}$ . For the example of  $R=2$ ,  $t=1/4$ , this optimum

is  $r_1 = 1.28$ .

## II.B. An Extension

When long-term assets are even more illiquid, there is an additional way that banks can help investors share the risk of liquidation losses. Suppose that the illiquid asset is as before, except that it returns  $1 - \tau$  (instead of 1) if liquidated at date 1, and  $\tau > 0$ . In this case a short term liquid asset (equivalent to storage) that returns one unit per unit invested in the previous period offers a higher one period return than investing in a long-term asset and liquidating it at date 1. However, because a bank knows that a fraction  $t$  of depositors need to withdraw at  $T=1$ , it can obtain the same set of payoffs as before:

$r_2 = \frac{[1 - t r_1]R}{1 - t}$ , by investing in short term assets to finance all of the date 1 withdrawals.

If the bank pays  $r_1$  at date 1 or  $r_2$  at date 2, it puts a fraction  $(t r_1)$  of assets into short term assets and  $1 - (t r_1)$  into long term illiquid assets and achieves the same payoffs as in the previous case. This holding of an inventory of liquid assets is referred to as the asset management of liquidity.

Investors holding the assets directly cannot do as well. Returning to the example, one possibility is for the bank to offer  $r_1=1$  or  $r_2=R=2$ . If an individual were to directly hold assets that allowed 1 to be obtained at date 1 (to consume 1 if of type 1), he or she would need to put 100% into short term liquid assets and would consume only 1 if of type 2, by reinvesting in the short term asset at date 1. The individual cannot achieve

---

be true in practice). At an interior optimum,  $\frac{r_1^{-\rho}}{r_2^{-\rho}} = R$  or  $\frac{r_2}{r_1} = R^{1/\rho}$ . With  $\rho > 1$ , this implies  $r_2 < R$ ,  $r_1 > 1$ .

The example assumes  $\rho=2$ .

$r_1 > 1$  at all. To obtain  $r_2 = R$  he or she must hold only illiquid long term assets, implying that the largest  $r_1$  available is  $r_1 = 1 - \tau$ . When long term assets are more illiquid ( $\tau$  is positive), then banks not only allow the risk of liquidating an illiquid asset to be shared, but also reduce the opportunity cost of creating a liquid date 1 payoff,  $r_1$ . This advantage of banks is present in the models of Bryant [1980], Jacklin [1987], Haubrich and King [1990] and Cooper and Ross [1998]. An investor's opportunity set without the bank is worse than the bank's, because an investor needs all or none of his liquidity, while the bank knows that a fraction  $t$  of its depositors will need liquidity at date 1. To be precise, the individual investor without using the bank can put a fraction  $\alpha$  in short term assets and the remainder in long term assets to obtain a choice between  $r_1 = \alpha + (1 - \alpha)(1 - \tau)$  and  $r_2 = \alpha + (1 - \alpha)R$ . Substituting out  $\alpha$ , the individual's tradeoff between  $r_1$  and  $r_2$  is given by  $r_2 = 1 + (1 - r_1) \frac{(R - 1)}{\tau}$ . When the probability of being of type 1,  $t$ , is not equal to 1 or zero, this is dominated by the bank's opportunity set of  $r_2 = \frac{[1 - tr_1]R}{1 - t}$ . For example when  $t = 1/4$ ,  $\tau = 1/2$  and  $R = 2$ , then without the bank the investor can get  $r_2 = 1 - 2(r_1 - 1)$ , so  $r_1 = .9$  implies  $r_2 = 1.2$ . However, if the bank sets  $r_1 = .9$ , it can offer  $r_2 = r_2 = \frac{[1 - .25(.9)]2}{.75} = 2.07$ . This provides an extra reason for bank's creating liquidity when assets are illiquid. The ability for banks to offer a given amount of liquidity, and the problems that this can possibly cause, are identical to the original model with  $\tau = 0$ . As a result, for the balance of section II, I return to the original Diamond and Dybvig [1983] model with  $\tau = 0$ .

### II.C. Bank Runs

Banks can create liquidity, by offering deposits that are more liquid than their assets. If only the proper depositors withdraw, it works very well. However, creating

this liquidity leaves the bank subject to bank runs. The bank may have liquidity problems. If a depositor's need for liquidity (the depositor's type) were a verifiable characteristic that could be written into contracts, the contract could specify that a type 1 be given  $r_1$  at date 1, and a type 2 be give  $r_2$  at date 2. However, on date 1 when each depositor learns his or her type, this is unverifiable private information. If a bank offers liquid deposits that offer each depositor the opportunity to withdraw  $r_1$  on date 1 or  $r_2$  on date 2, the depositors may select the appropriate withdrawal date for their type. That is the type 1's take  $r_1$  and the type 2's take  $r_2$ , and if all are expected to do this, each will choose the option that is best for him or her. It turns out, however, that there are multiple equilibria. That is, there is more than one self-fulfilling prophecy about who withdraws at date 1. There is a good equilibrium where only the type 1 depositors withdraw and a bad equilibrium (a bank run) where all withdraw at date 1 because they all expect each other to do the same.

To see why there are multiple equilibria, consider how much is left to pay depositors who wait until date 2 to withdraw if a fraction  $f$  of initial depositors withdraw at date 1. Because each asset is worth 1 at date 1, a fraction  $f r_1$  of the total assets must be liquidated at date 1. This leaves  $r_2(f) = \frac{\{1 - [f \times r_1]\}R}{1 - f}$  for each of the fraction  $1-f$  who wait until date 2. In any equilibrium, at least a fraction  $t$  of deposits will be withdrawn, or  $f \geq t$ , because type 1's always withdraw at date 1. The type 2 depositors will chose to withdraw at date 1 as well if they believe that a fraction  $f$  will withdraw and  $r_2(f) < r_1$ . In the example with 100 depositors,  $t=1/4$ , or 25 are of type 1. If just the type 1 depositors withdraw, or  $f=t=1/4$ , and  $r_1 = 1.28$ , then  $r_2 = 1.813 > r_1$ , and the type 2 depositors will choose to wait until date 2 to withdraw. Depositors must choose simultaneously, before

they know the actions of others. Each needs a forecast of  $f$ , denoted by  $\hat{f}$ . Given a borrower's forecast, he or she chooses whether to withdraw at date 1. A Nash equilibrium is a self-fulfilling prophecy of  $\hat{f} = f$ , and in the good equilibrium,  $f = \hat{f} = t = \frac{1}{4}$ .

However, suppose all depositors forecast that everybody else will withdraw (i.e., 99 depositors, so  $\hat{f} \geq 0.99$ ). Then the bank will fail before  $T=2$ . If 79 depositors or more are expected to withdraw, then the bank will be worthless at date 2: the bank can be liquidated for at most 100 at  $T=1$ , and if 79 depositors were to each receive 1.28, at  $T=1$ , the bank would not have sufficient assets, because  $79 \times 1.28 = 101.12 > 100$ . Note that a prophecy of  $\hat{f} = 0.99$  is not self-fulfilling, because if it is believed by all, then every depositor will withdraw. The self-fulfilling prophecy of a bank run is  $f = \hat{f} = 1$ , where all rush to withdraw. Providing liquidity leaves the bank subject to runs. If a run is feared, it becomes a self-fulfilling prophecy

The first paragraph of Diamond and Dybvig [1983] follows. "Bank runs are a common feature of the extreme crises that have played a prominent role in monetary history. During a bank run, depositors rush to withdraw their deposits because they expect the bank to fail. In fact, the sudden withdrawals can force the bank to liquidate many of its assets at a loss and to fail. In a panic with many bank failures, there is a disruption of the monetary system and a reduction in production."

Bank runs disrupt production because they force banks to call in loans early. This forces the borrowers to disrupt their production. The model does not have an explicit model of loans from the banks; it simply models the bank loans as illiquid. See Diamond and Rajan [2001] for a description of why bank loans are illiquid.

These two possible equilibrium beliefs (self-fulfilling forecasts of  $f$ ) are locally stable. That is, if  $t=1/4$ , a type 2 depositor will not run given a forecast  $\hat{f}$  is just above  $1/4$ , for example  $\hat{f}=0.27$ . Similarly, a type 2 depositor would run given a forecast  $\hat{f}$  just below 1, for example,  $\hat{f}=.97$ . The tipping point for a run is a forecast implying that  $r_1 \geq r_2$  or  $r_1 > r_2(\hat{f}) = \frac{\{1 - [\hat{f} \times r_1]\}R}{1 - \hat{f}}$ , which in the example is  $\hat{f} > \frac{(R - r_1)}{r_1(R - 1)} = \frac{2 - 1.28}{1.28(2 - 1)} = 0.5625$ .

Because moving away from a good equilibrium requires a large change in beliefs, the initiation of a run when none was expected requires something that all (or nearly all) depositors see (and believe that others see). For example, a newspaper story that the bank is doing poorly could cause a run even if many knew that it was inaccurate, because those who know it is inaccurate can believe that the others will decide to withdraw based on the story. Even sunspots could cause runs if everyone believed that they did.

Using diversified funding sources can help insulate a bank from runs, if diversified means that there is no commonly observed information source that is seen by a large number of the diverse depositors. An older example is also useful. It would make sense for a bank to have a large lobby (or fast bank tellers), because if a line to withdraw extended out to the street, passersby may conclude that a run is in progress. Conversely, once a run is in progress, it will be important to be able to convince all depositors that it will stop and to have all the depositors know that all others have been so convinced.

When depositors do not all observe the same news or other information sources, then the depositors will not all have a way to tell if others are choosing to panic and run (they will have “incomplete common knowledge”). There are some very interesting analyses of runs in this context, see Morris and Shin [2003] and Goldstein and Pauzner

[2005]. Important analyses of bank policies when there is an unavoidable positive probability of a run are presented in Peck and Shell [2003], Ennis [2003], and Ennis and Keister [2006].

#### **II.D. Suspension of convertibility**

In this simple model, a bank can suspend convertibility of deposits to cash to stop a run. That is, suppose the bank does not allow more than a fraction  $t$  of deposits to be withdrawn (does not allow  $f > t$ , or in the example, allows only 25 to withdraw). Then no matter how many depositors attempt to withdraw at date 1, a type 2 will get

$$r_2(t) = \frac{\{1 - [t \times r_1]\}R}{1 - t} > r_1 \text{ at date 2.}$$

In the example, the type 2 would get 1.813 at date 2.

As a result the depositors never panic, and a run would never start. In this case, the suspension is only a threat that need not actually be carried out. The problem is to convince potential participants in a run that convertibility will be suspended at the proper time. In the days before deposit insurance, banks regularly suspended convertibility to stop runs (see Friedman and Schwartz [1962]). In a more general model where the fraction of type 1 depositors fluctuates sufficiently (and the realized fraction cannot be written into contracts), suspension cannot be used only as a threat. Some suspension would actually occur and would be unpopular. If suspension occurred regularly, depositors would desire another way of stopping runs caused by panics. In practice, government-provided deposit insurance has been instituted following many financial crises. Its effects are described in the next section.

#### **II.E. Deposit Insurance**

An alternative way to stop and prevent runs is deposit insurance, a promise to pay the amount promised by the bank no matter how many depositors withdraw, without

suspension of convertibility. In the example, this is a promise of 1.28 to those who withdraw at  $T=1$  and 1.813 to those who withdraw at  $T=2$ . How can this be accomplished if everyone withdraws? Unless there are outside resources that we did not account for in the model, the only way is to take some resources away from those who run and withdraw. Governments have taxation authority, the ability to take resources without prior contracts. This gives government deposit insurance an advantage over private deposit insurers who might themselves fail in a run, or who would need to hold sufficient liquid assets to prevent the financial system from creating liquidity.

In our example with  $t=1/4$ , where exactly 25 people ought to withdraw, suspension of convertibility works as well. However, if there is aggregate uncertainty about  $t$ , the fraction of type 1's (withdrawals needed when there is not a run), then suspension is costly. Suspension may prevent some type 1 depositors from withdrawing. Deposit insurance can stop runs and avoid suspension of convertibility.

A bank with deposit insurance can credibly promise not to have runs. Government deposit insurance works because the government has taxation authority and, unlike most insurance companies, can provide a guarantee against large losses that are usually off the equilibrium path without holding reserves to back up their promise. In addition, a deposit insurance law commits the government to insure banks, which is an advantage over discretionary policies if self-fulfilling prophecies of runs need to be eliminated. Suspension of convertibility is usually a discretionary policy, see Gorton [1985]. Another discretionary policy to prevent banks from liquidating illiquid assets and avoiding self-fulfilling runs is central bank lending, financed by implicit taxation or money creation authority. The extent of the Great Depression in the United States in the 1930s has been

blamed on the lack of Federal Reserve discount window lending by Friedman and Schwarz [1963]. Deposit insurance will solve this problem of discretionary lending, but its guaranteed bailout of depositors may cause incentive problems if bank regulation is poorly structured (see Barth, Caprio, and Levine [2006]).

### **III. Illiquid Financial Assets**

The definition of an illiquid asset in this chapter is one where the liquidation proceeds available from physical liquidation or sale are less than the present value of its future payoffs. This section uses a generalization of the model of Chapter 1 to examine why bank loans (and related financial assets) are illiquid and to show that this provides added insights into the way that banks create liquidity for depositors. I show that although loans made with undelegated monitoring will be illiquid, the deposits of a bank which is a delegated monitor will be liquid.

The Diamond-Dybvig model described earlier in this chapter focuses on the consequences of illiquid assets but is silent on the reasons why they are illiquid: illiquidity is assumed due to production processes that provide a low payoff if terminated early. The following analysis explains the illiquidity of bank loans combining the ideas in Diamond-Rajan [2001] with the model from chapter 1.

The model in this section explains the illiquidity of bank loans using the idea of asset specificity, which is often cited as the motivation for illiquid real (non-financial) assets. Some real assets are illiquid because their value is highly specific to the original owner (their *future* payoffs to the original owner are larger than to the buyer, and the owner will only sell them if current liquidity is needed). Real assets with high specificity

include machines that are best used in a particular firm or which require specialized skills to operate, or houses with very unusual characteristics, where it may be difficult and costly to find buyers who value them highly. One way to motivate the illiquid asset in the Diamond-Dybvig model is as a totally specific real asset that can never be operated by another entrepreneur at date 2 (and has a zero operating value) but which can be physically liquidated for a low amount on date 1.

A financial asset (a loan), seems less specific than a real asset (a machine) because loans appear not to require production skills to maximize their value. However, if specialized monitoring or loan collection skills are required to induce borrowers to pay a loan, then the value of a loan is specific to the monitoring or collection skills of its current owner. Undelegated monitoring requires that the owner of the loan have skill in monitoring. I use the model from chapter 1 to illustrate this issue. A borrower who owes a large amount to a lender will divert funds if not monitored. Selling a loan to a buyer who cannot monitor (or who is a less skilled monitor) will result in low sale proceeds because the buyer will pay only the present value of the amount that he or she can collect from the borrower. The loan buyer will collect very little if the borrower diverts funds, or the buyer might need to offer to accept a lower payment from the borrower to deter diversion.

Suppose that monitoring is required to deter a borrower from diverting on date 2. I use a two period extension of the model from chapter 1 that assumes a risk free cash flow of  $H$  and assume that borrower diversion is reversible. If the loan is sold on date 1 to someone who cannot monitor and has access only to legal sanctions for default, then if

sold and thus not monitored, the borrower will pay at most  $H(t+\phi)$  on date 2.<sup>2</sup> If a loan is not sold and is thus monitored, the borrower will pay up to  $H(t+\phi+m)$ .<sup>3</sup> As a result, if the loan with face value  $F=H(t+\phi+m)$  is sold, the borrower's future payment declines by  $Hm$ , and a buyer will pay at most the present value of  $H(t+\phi)$  for the loan at date 1. The future added value of the monitoring,  $Hm$ , cannot be sold and is illiquid, reducing the amount for which the loan can be sold.

This illiquidity of loans made by undelegated monitors suggests an added value of delegated monitoring. Under the assumptions of chapter 1, in particular where there is an automatic legal sanction on the bank if it pays less than promised to depositors, and where the bank has all bargaining power, bank deposits will be liquid even though bank loans are not. Deposits are liquid because the banker's incentives to monitor are provided by the threat of sanctions triggered by bank failure. Because no specialized monitoring by depositors is required, the value of deposits does not depend on the skill of the depositor; they are worth the same to all holders and are liquid. The deposits could be resold, or the bank can borrow from new depositors (who need no monitoring skill) if old depositors need to withdraw because they need liquidity.

Banks create liquid deposits from illiquid loans by acting as delegated monitors. This is valuable not only to investors whose investments are too small to justify direct monitoring, but also to investors who wish to lend sufficiently large amounts to justify direct monitoring but who may have uncertain future liquidity needs. An investor who could directly monitor a loan, but who might need liquidity, would rather hold a claim on

---

<sup>2</sup> This is Proposition 1. Recall that diversion destroys a fraction  $t$  of output, and the legal penalty reduces diverted output by a fraction  $\phi$ .

<sup>3</sup> This is Proposition 2. Recall that if a monitor stops a crime in progress, the diverted output is reduced by a fraction  $m$ .

a bank which is delegated the monitoring of loans, than personally monitor a loan. This benefit occurs even though the investor could monitor the loan personally without duplication of effort (which removes the cost benefit of delegated monitoring). Let  $PV(Z_2)$  denote the date 1 present value of a claim that pays  $Z_2$  on date 2. An investor who personally monitors and is not a delegated monitor can choose between selling the loan at date 1 for  $r_1 = PV(H(t+\phi))$  or holding it until date 2 and receiving  $r_2 = H(t+\phi+m)$ . If the investor instead holds a claim on a delegated monitor, the choice is between date 1 proceeds of  $r_1 = PV(H(t+\phi + \min\{\phi_M, m\}))$ , or holding on until date 2 and receiving  $r_2 = H(t+\phi + \min\{\phi_M, m\})$ .<sup>4</sup> If monitoring is delegated, the deposits will be liquid, and the date 1 proceeds equal the present value of the date 2 proceeds ( $r_1 = PV(r_2)$ ). For a sufficiently high probability of a liquidity shock that forces an undelegated monitor to sell his loan, the liquid deposit from the delegated monitor will be preferred. This illiquidity of loans is an additional reason (in addition to avoiding duplication of effort in initial monitoring) that delegated monitoring will be preferred to direct monitoring. The next section shows how this notion of illiquidity relates to that in the Diamond-Dybvig model. It can be skipped without loss.

### **III.A Relation to the numerical example of the Diamond-Dybvig model**

It is instructive to link this liquidity value of delegated monitoring to the numerical example of the Diamond-Dybvig model presented above, where bank deposits offer a choice between  $r_1 = 1.28$  at date 1 or  $r_2 = 1.813$  at date 2, while holding illiquid assets directly offers  $r_1 = 1$  or  $r_2 = 2$ . In the Diamond-Dybvig model, these payoffs are produced by a cross-subsidy of those who withdraw at date 1 by those who wait until date

---

<sup>4</sup> See Proposition 2.

2. In the monitoring based model described above, these payoffs are produced by the bank offering fully liquid deposits that remove the need to directly hold claims which are owner specific and thus illiquid (the bank offers an  $r_1$  which is the date 1 present value of  $r_2$ ). Suppose that the loan is held by an undelegated monitor who can collect  $H(t+\phi+m)=2$  at date 2, or it can be sold at date 1 for 1 (and that 1 is the present value of the amount that a lender who cannot monitor can collect at date 2:  $H(t+\phi)$ ). The bank, as the delegated monitor, can commit to pay depositors  $H(t+\phi+\min\{\phi_m, m\})=1.813$  at date 2 and can commit to pay the present value of this amount at date 1 (because the deposit is fully liquid). Because the bank deposits are liquid, this requires that 1.28 be the date 1 present value of 1.813 at date 2; equivalently, the date 1 present value of one unit of date 2 goods is  $1.28 / 1.813 = 0.706$ . This is consistent with an undelegated monitor selling the loan for a price of 1 at date 1 to someone who cannot monitor if that buyer will collect 1.416 at date 2; the date 1 price of 1 is the present value of  $1.416=H(t+\phi)$ , because  $(1.28/1.813)(1.416)=1$ . If  $H\phi_M=1.813-1.416=0.397$  and  $Hm=2-1.416=0.584$ , then all the numbers line up with the earlier example.

#### **IV. Monitoring, limited commitment and bargaining**

The delegated monitoring model from Chapter 1 and section III above assumes contracts that commit to impose legal sanctions contingent on low payments to lenders or depositors. In addition, monitors have all the bargaining power over borrowers: they capture the entire surplus when they use their monitoring to threaten to stop a crime in progress. As a result, monitoring is effective in deterring diversion if stopping a crime in progress can sufficiently reduce the spoils of diversion. This section allows for legal sanctions on borrowers that must be imposed voluntarily by lenders or monitors

(allowing for sanctions that are not automatically triggered by a default). This generalization shows how delegating monitoring may require the first-come-first-served demand deposit contracts that were assumed in the Diamond-Dybvig model. These deposits can provide depositors (or lenders more generally) an incentive to invoke sanctions that synthesize automatic sanctions. In addition, this section separately gives borrowers some bargaining power (so they get some surplus in negotiations with lenders or monitors), whether or not sanctions for default are automatic. This implies that skills other than the ability to observe a borrower are important to be an effective monitor (asset redeployment skills matter).

Because only the monitor observes diversion, the monitor must initiate discretionary actions or negotiations based on attempted diversion. Intervention must be a voluntary choice of the monitor's; there is no way to specify automatic intervention. As a result, his asset redeployment skills are important if the borrower has any bargaining power, even if sanctions for subsequent default are automatic.

When the borrower has some bargaining power, an effective monitor must be able to achieve a high recovery if he actually intervenes: the monitor needs a high outside option in the bargaining. This required high recovery or outside option can be interpreted as skill in asset redeployment or liquidation or more generally a specific skill possessed by a relationship lender. This generalization gives an added reason why monitoring ability is lender specific and why monitored loans are illiquid.

Section IV.A examines monitoring by an undelegated monitor with no bargaining power, when automatic legal sanctions are imposed for default, but where the monitor can choose to intervene to stop a crime in progress. This is followed by an analysis of

delegated monitoring in the case where depositors and lenders have no bargaining power and where automatic sanctions for default on deposits are also possible. Section IV.B examines the consequences of making all sanctions, including those for default, voluntary, for both undelegated and delegated monitors when depositors and lenders have no bargaining power.

#### **IV.A Monitoring when there are Automatic Legal Sanctions for default and the borrower has all the bargaining power**

To illustrate the importance of a monitor's redeployment skills when the borrower has some bargaining power, consider an undelegated monitor with no bargaining power (where the borrower gets the entire surplus from negotiations). The borrower's debt has a face value of  $F$ . If automatic legal sanctions are triggered for payments less than  $F$  (and these cannot be renegotiated), we saw in Proposition 1 that the borrower, in order to avoid the sanctions, will pay up to  $H(t+\phi)$  without diverting if there is not monitoring. If legal protection is weak (low  $\phi$ ), this will be too low an amount to make lending profitable. Suppose that the face value of debt,  $F$ , exceeds  $H(t+\phi)$  and the borrower diverts. If the monitor threatens to stop the crime in progress, the borrower will offer the monitor a payment to refrain from stopping the crime. Because the borrower has all the bargaining power, the monitor will accept an amount that gives him a total payoff equal to what he can obtain by rejecting the offer (his outside option). The monitor's outside option is the larger of two amounts: the recovery achieved from intervening immediately to stop the crime (e.g., redeploying assets immediately), which is  $X_m$ , and the recovery,  $X_\phi$ , from allowing the crime in progress to proceed, leading to a subsequent default and automatic sanction. The monitor will prefer to intervene unless offered a payment of at

least  $\max\{X_m - X_\phi, 0\}$ . Stopping the crime in progress would reduce the borrower's diversion proceeds by more than this payment, so the borrower will prefer to pay the monitor to not intervene. The borrower's offer of  $\max\{X_m - X_\phi, 0\}$  gives him a payoff of  $H(1-t-\phi) - \max\{X_m - X_\phi, 0\}$ .<sup>5</sup> If the borrower does not divert and pays  $F$ , his payoff is  $H-F$ . As a result, if  $F$  exceeds  $H(t+\phi) + \max\{X_m - X_\phi, 0\}$ , diversion remains attractive even if monitored.

When sanctions on default are automatic, an undelegated monitor with no bargaining power can force the borrower to pay up as much as  $F \leq H(t+\phi) + \max\{X_m - X_\phi, 0\}$ . The monitor's redeployment skill, reflected in a high value of  $X_m$ , is essential to be effective at deterring diversion. If  $X_m \leq X_\phi$ , an undelegated monitor with no bargaining power is no better at deterring diversion than a lender who does not monitor. This is in marked contrast to an undelegated monitor who has all the bargaining power (as in chapter 1), who can induce the borrower to repay a face value  $F$  up to  $H(t+\phi+m)$  without diverting, by capturing all the surplus from threatening to stop a crime in progress.

A skilled monitor, one with  $X_m - X_\phi > 0$ , can serve as a delegated monitor when structured as a bank if automatic legal sanctions are triggered by default on deposits (as in Chapter 1). A delegated monitor who issues deposits with face value  $B$  will monitor loans and pay depositors up to  $B = H(t+\phi) + \min\{H\phi_M, \max\{X_m - X_\phi, 0\}\}$  without allowing diversion.<sup>6</sup> This is identical to the case of automatic legal sanctions in chapter 1, except that the monitor's redeployment skill,  $X_m$ , is important. The final part of the next section shows a very different result when legal sanctions for default are not automatic.

---

<sup>5</sup> The borrower could also offer no payment when  $X_m - X_\phi > 0$  and allow the monitor to stop the crime in progress, giving the borrower a payoff of  $H(1-t-\phi-m)$ . Because  $Hm \geq X_m$  and  $X_\phi \geq 0$ , this is dominated.

<sup>6</sup> The legal sanction on the bank is  $H\phi_M$  if it defaults.

#### **IV.B Monitoring when Legal Sanctions for default are voluntarily imposed and the borrower has all the bargaining power**

When legal sanctions are not automatic, a contract will be renegotiated to avoid inefficiently imposing voluntary sanctions. This can reduce or remove the ability to use sanctions to provide incentives. If the threat of sanctions is needed to deter diversion, bargaining power and asset redeployment skills will be important even for lenders (or depositors) who do not monitor. As in the previous section, consider a borrower with all of the bargaining power (who can successfully offer a lender the recovery from imposing the legal sanction). A lender who does not monitor recovers  $X_\phi$  from imposing the legal sanction. If the face value of debt exceeds  $X_\phi$ , the borrower will divert and threaten to reverse the diversion only if the face value is reduced to  $X_\phi$ . The lender will accept. Therefore, the most that the lender who does not monitor will receive is  $X_\phi$ .

A lender who is an undelegated monitor can get a recovery of  $X_m$ , if he stops a crime in progress. The borrower can successfully induce him to neither stop the crime nor impose the legal penalty for default by offering a payment of  $\max\{X_m, X_\phi\}$ . Stopping the crime in progress will also reduce the borrower's diversion proceeds by  $H_{m \geq X_m}$ , but the monitor cannot use this threat to get the borrower to pay more than  $\max\{X_m, X_\phi\}$ . As in the previous section, effective monitoring also requires asset redeployment skills: a high value of  $X_m$ . If the monitor's recovery is not larger than  $X_\phi$ , monitoring will have no deterrent effect. If the skills to achieve a high value of  $X_m$  are lender specific, then effective undelegated monitoring will be lender specific.

When no automatic sanctions are possible and the depositors have no bargaining power, the delegation of monitoring will not be possible. The delegated monitor can

threaten not to use his monitoring skills to collect the loan and leave the depositors to collect the loan without monitoring. A depositor who does not monitor achieves a recovery of  $X_\phi$  from imposing the legal penalty if the monitor does not intervene to stop a borrower from diverting. This is the same recovery achieved when there is direct lending to the borrower, without a monitor. If the delegated monitor issues deposits (that are not first-come-first served demand deposits) and the depositors have no bargaining power, the coalition of the delegated monitor and the borrower will pay the depositors no more than  $X_\phi$ . Monitoring will allow the monitor to collect larger payments from the borrower but they will not be paid to depositors. This implies that delegated monitoring requires a way to commit depositors not to accept payments as low as  $X_\phi$ . The absence of automatic sanctions will make delegated monitoring unviable.

Depositors need to commit not to renegotiate to accept less than  $B$ , and this can be achieved by demand deposits paid on a first-come-first-served basis. This will be a self-enforcing equivalent to contracts that achieve automatic sanctions and thus will automatically deter renegotiation. The next section shows that delegated monitoring is viable for a delegated monitor who can freely renegotiate the face value of loans,  $F$ , but who faces automatic sanctions from default on deposits. This motivates results shown thereafter where first-come-first-served demand deposits will work as effectively as automatic sanctions to provide incentives for delegated monitoring.

#### **IV.B.1 Automatic Sanctions for deposit defaults when the delegated monitor has all the bargaining power**

When depositors can negotiate to remove their threat to impose the legal sanctions for payments less than  $B$ , and have no bargaining power, they receive no more than  $X_\phi$ ,

the amount that they recover from imposing the sanction. If instead there are automatic sanctions for all payments less than  $B$  (made by the delegated monitor to depositors), the delegated monitor will choose to pay  $B$  to depositors unless the legal sanctions are insufficient. In addition, automatic sanctions imposed on deposit defaults can allow a delegated monitor to collect an incentive compatible payment from the borrower that exceeds  $X_m$ , the amount that an undelegated monitor could collect.

**Proposition 3.** A delegated monitor with deposits with face value  $B$  who faces automatic legal sanctions for deposit default and who monitors a loan with a sufficiently high face value  $F$  with voluntary sanctions for loan payments below  $F$  will be willing to pay depositors up to  $B$  for all  $B \leq H(t+\phi) + \min\{\max\{0, X_m - X_\phi\}, \phi_M H\} \equiv \bar{B}$ . A delegated monitor with deposits of  $B \leq \bar{B}$  can enforce an incentive compatible payment from the borrower of up to  $B + \max\{0, X_m - X_\phi - \phi_M H\}$  if  $B \geq X_m$  and  $\min\{X_m, X_\phi\}$  if  $B < X_m$ .

Proof: See Appendix.

An undelegated monitor without bargaining power collects a loan for his own account and can collect only his outside option,  $X_m$ , while a delegated monitor who owes deposits  $B > X_m$  faces default penalties for making payments less than  $B$  to depositors. This discussion assumes that  $X_m > X_\phi$ , so the monitor can collect more than a lender who does not monitor. The delegated monitor will find it unattractive to accept a low verifiable loan payment less than  $B$  from the borrower because the payment accrues to depositors (because it is verifiable) and triggers default penalties on the delegated monitor (because it is less than  $B$ ). The only way to induce the delegated monitor to accept a payment of less than  $B$  and not intervene is to offer to share diversion proceeds

which is costly and which triggers deposit default sanctions. As a result, when deposits exceed  $X_m$  but do not exceed  $\bar{B}$  (given in Proposition 3), the borrower will be willing to make verifiable payments in excess of  $X_m$  to the delegated monitor, and the delegated monitor will accept them and repay deposits. The automatic sanctions on deposit defaults allow the delegated monitor to collect as much as an undelegated monitor monitoring a loan that has automatic sanctions for *loan* defaults. In addition, the delegated monitor can collect more than  $X_m$  and more than the value of its deposits when its outside option is positive ( $X_m - X_\phi - \phi_M H > 0$ ). The delegated monitor's use of deposits with automatic sanctions affects its ability to bargain with its borrower and to commit to pay its deposits. This illustrates the role of preexisting non-renegotiable debt as a way of committing to turn down low offers, as in Brander and Lewis (1998), Spier and Perotti (1993), or Diamond –Rajan [2000].

The ability to renegotiate contracts to avoid the sanctions adds a role for demand deposits to commit depositors not to renegotiate. Diamond-Rajan [2001] presents the idea that the threat of bank runs on demand deposits can commit depositors not to renegotiate. The model in Diamond-Rajan [2001] is based on Hart and Moore [1994] and thus is somewhat different from that in this section. I compare the models below, and their relation to the threat of runs on firms in Diamond [2004] and Von Thadden, Berglof, and Roland [2003] in Sections IV.C.1 and VI.

#### **IV.C Demand deposits commit depositors to not make concessions**

When the imposition of legal sanctions for default is voluntary, a delegated monitor who raises funds from depositors with no bargaining power will be able to deter them from imposing legal sanctions for default by offering them their outside option: the

recovery from imposing default sanctions,  $X_\phi$ . If the borrower and monitor collude to divert and offer to reverse the diversion in return for a reduction in the face value of total deposits, the depositors will accept any reduction to a total value of at least  $X_\phi$ . This assumes that the outside option of each depositor is his pro-rata share of the recovery from declaring a default,  $X_\phi$ . First-come-first-served demand deposits can provide important externalities that increase individual outside options sufficiently to induce depositors to reject offers greater than  $X_\phi$  from the delegated monitor and to impose the legal sanction instead. This works because each depositor has an incentive to run to the bank, demanding full payment of his deposit, whenever a loss in the future is anticipated. If the delegated monitor does not intervene after the borrower diverts, the total value of the bank will be  $X_\phi$ .

If demand deposits are issued to the depositors and if any depositors demand payment, the bank must pay them in full or an observable default will occur, triggering legal sanctions. Those who withdraw first get paid in full, until the bank fails and sanctions are triggered. As a result, any call for reductions in the amount owed to depositors will cause the bank to fail, even though a successful forgiveness of debt (reduction in the amount owed to depositors) would have avoided the failure and legal sanctions. Bank runs reward those who get out first and demand full payment, and they punish those who instead reduce their claims and leave their money in the bank. Alternatively, if the bank takes an action that will impose a future loss on depositors, but does not ask for debt forgiveness, each depositor will demand payment immediately. No one will leave their money in the bank if all anticipate a loss.

The incentive effects of potential runs on demand deposits are most easily demonstrated by a bank with two identical depositors who hold demand deposits with total face value of  $B$  (each has face value of  $\frac{1}{2}B$ ). Each depositor's payoff depends on his own decision (accept the delegated monitor's offer of less than  $\frac{1}{2}B$  or withdraw, demanding payment of  $\frac{1}{2}B$ ) and on the decision of the other depositor. The new face values offered by the bank need not be equal for the two depositors; that is, the new face values for depositors 1 and 2 are  $B_1'$  and  $B_2'$  respectively, such that  $B_1' + B_2' = B' < B$  (if they are equal  $B_1' = B_2' = \frac{1}{2}B'$ ). If the monitor offers depositors a revised total face value  $B' < B$  that is at least  $X_\phi$ , it is in their ex-post collective interest to accept. However, if  $B < 2X_\phi$ , the only Nash equilibrium is for both depositors to withdraw, which forces the legal sanctions with recovery  $X_\phi$  to be imposed. This is Proposition 4.

**Proposition 4.** If there are two depositors with first-come-first-served demand deposits with face value of  $\frac{1}{2}B$  each, such that  $B > X_\phi$  and  $B \leq 2X_\phi$ , where  $X_\phi$  is the recovery from imposing the legal sanction for default, then if offered any reduction in payments (which sum to less than  $B$ ) in return for not imposing the legal sanction, the unique Nash equilibrium is for each to demand payment of  $\frac{1}{2}B$ , forcing the legal sanction to be imposed.

*Proof:* Their individual payoffs are:

	#2 Accepts	#2 Withdraws
#1 Accepts	$(B_1', B_2')$	$(\min\{B_1', \max\{0, X_\phi - \frac{1}{2}B\}\}, \min\{X_\phi, \frac{1}{2}B\})$
#1 Withdraws	$(\min\{X_\phi, \frac{1}{2}B\}, \min\{B_2', \max\{0, X_\phi - \frac{1}{2}B\}\})$	$(\frac{1}{2}X_\phi, \frac{1}{2}X_\phi)$

If a lender believes that the other will withdraw, the best response is to do the same because  $B > X_\phi$  implies  $\frac{1}{2}X_\phi \geq X_\phi - \frac{1}{2}B$ . If both believe that the other will accept,

the best response for a depositor offered less than his original claim,  $\frac{1}{2}B$ , is to withdraw because  $X_\phi > \frac{1}{2}B$ , and withdrawing will allow full payment. If the offer reduces total debt to less than  $B$ , so  $B_1' + B_2' = B' < B$ , then either  $B_1'$  or  $B_2'$  must be less than  $\frac{1}{2}B$ . Thus at least one depositor has a dominant strategy to withdraw and the unique Nash equilibrium is for both to withdraw.

Q.E.D.

A depositor anticipating a loss from the delegated monitor's offer (or from the delegated monitors's lack of intervention) will choose to run to demand payment if the other depositor is anticipated not to do so. Depositors who get out early when a loss is anticipated become senior to those who do not demand payment. If other depositors are anticipated to withdraw, a depositor will still want to withdraw immediately because when paid first-come-first served, withdrawing immediately gives either full payment (if one succeeds in getting to the bank before other depositors) or results in the bank running out of assets (in which case one is no worse off than leaving money in the bank). This commits depositors to run, which invokes the legal penalty for default by the delegated monitor.

#### **IV.C.1 The required number of depositors to deter all concessions**

If  $X_\phi \geq \frac{1}{2} B$ , then with two depositors each depositor will run rather than make any concession at all; if the other depositor is anticipated to make a concession, a depositor can avoid any loss by withdrawing. If  $X_\phi < \frac{1}{2} B$ , then two depositors are not enough for this to be true. However, as shown in Diamond [2004], if  $X_\phi > 0$  then a number of depositors  $N \geq B/X_\phi$  will suffice to induce each depositor to run rather than make any concession, because if  $X_\phi \geq B/N$ , then it never pays to make a concession if other deposits

will make concessions. A depositor's payoff from being the first to withdraw is  $\min\{B/N, X_\phi\}$  if his deposit has face value  $B/N$  and if all others are expected to accept. In addition, it will be better to withdraw rather than make a concession if any other depositors are expected to withdraw. Because deposits must be paid on demand if default is to be avoided, if leaving money in the bank has a positive value, one can achieve full payment by withdrawing, and this exceeds the payoff from making a concession. If, instead, leaving money in the bank has no value, it will clearly be better to withdraw (quickly before the money runs out).

#### **IV.C.2      Should Banks or Borrowers Issue Short-Term Demandable Debt?**

The payoffs from first-come-first-served short-term deposits can replicate automatic sanctions because for any offer less than face value  $B$ , it is a dominant strategy for all depositors to run. Because a run will cause an observable default, running invokes the voluntary legal sanction, making its imposition effectively automatic. This was illustrated here by first-come-first-served short-term deposits of a delegated monitor.

The importance of monitoring is illustrated by considering what would happen if a borrower issued first-come-first-served short-term debt directly to lenders who do not monitor. This short-term debt would replicate unmonitored debt with automatic imposition of legal sanctions. This will not be a viable financing choice under the assumptions used in sections IV.A and IV.B because there is weak legal protection. Monitoring is needed because the legal sanction for default is too small to deter diversion, even if automatically imposed.<sup>7</sup> Instead, demand deposits issued by the delegated monitor allow depositors to commit to impose default penalties on the

---

<sup>7</sup> The borrower needs to commit to pay  $I$  to lenders without diverting, and it is assumed that  $H(t+\phi) < I$ .

delegated monitor. As in Chapter 1, these penalties provide incentives to monitor loans on behalf of depositors. When legal protection is weak and automatic sanctions are not possible, a delegated monitor can commit to not allowing diversion by a borrower by financing itself with demand deposits from multiple depositors. This will also remove the problem of illiquidity which would be present if there were an undelegated monitor. The deposits of the delegated monitor will be liquid when its loans are not.

In contrast, when the legal environment is strong but enforcement is costly (and  $X_\phi$  is low), unmonitored short-term demandable debt is a viable means of direct financing for borrowers. Diamond [2004], which analyzes this case (albeit in a different setup), can be illustrated with this model. Suppose that legal protection is strong but legal penalties for default are voluntary and offer lenders a low recovery.<sup>8</sup> Monitoring would not be needed if default sanctions were automatic. If default sanctions are not automatic, monitoring remains unnecessary if the borrower synthesizes automatic sanctions by issuing short-term demandable debt. This induces lenders to voluntarily impose the sanctions for default. Without the incentives from the demandable debt, lenders without bargaining power could not force the borrower to pay more than what they recover from imposing the legal penalty.

More generally, first-come-first-served short-term debt can synthesize automatic default sanctions, but only monitoring can provide a borrower with a stronger incentive to pay than the incentive provided by automatic default sanctions. When sanctions are voluntary, short-term debt will be used by a borrower or a delegated monitor who needs to synthesize a contract with automatic sanctions. This has interesting predictions about when monitoring is used and when first-come-first-served short-term debt is used. Section V describes more generally the amounts that can be committed to pay to lenders or depositors through the use of monitoring and of short-term debt.

---

<sup>8</sup> Formally,  $H(t+\phi) \geq I$ , default penalties are voluntary and give a low recovery,  $X_\phi < I$ .

**V. Incentive compatible payments with given bargaining power, with and without automatic sanctions.**

The maximum amount that a borrower will choose to pay without diversion depends on how large his payoff is from diverting or threatening to divert if a lender does not reduce the required payment. The borrower's payoff depends on the strength of default sanctions, their application (either voluntary, automatic or effectively automatic due to short-term debt) as well as the possible added deterrence of monitoring. Chapter 1 describes what happens when all sanctions are automatic and the borrower gets none of the surplus in negotiation, and the previous section describes voluntary and automatic sanctions where the borrower gets the entire surplus in negotiations. This section describes the intermediate cases, where the surplus is shared by the borrowers and lenders (and in the case of delegated monitoring, the depositors negotiating with the monitor).

I introduce a parameter,  $\mu$ , between zero and one, which measures the bargaining power of a borrower (or  $\mu_{DM}$ , measuring the power of a delegated monitor bargaining with depositors). Each party to negotiation gets at least his outside option, the payoff if no agreement is reached. In addition, a borrower gets a fraction  $\mu$  of the increased surplus in negotiation with a lender or monitoring lender. For example, a lender who does not monitor and threatens to impose the default sanction gets the remaining fraction,  $1-\mu$ , of the surplus. Likewise, a lender who monitors and observes a crime in progress and threatens to intervene obtains a fraction  $1-\mu$  of the surplus from removing this threat. Similarly, a depositor negotiating with a delegated monitor gets a fraction  $1-\mu_{DM}$  of the surplus in negotiating whether to impose the default sanction on the delegated monitor. The bargaining power of borrowers negotiating with lenders,  $\mu$ , could differ from the

bargaining power of delegated monitors negotiating with depositors, (e.g.,  $\mu_{DM} \neq \mu$ ).

These intermediate cases where surplus is shared can be described using the results with  $\mu=0$  and  $\mu=1$  where one party gets the entire surplus from negotiations and the other receives his outside option. Table 1 describes these cases, using results shown previously. If negotiation breaks down, the legal sanction is imposed, or in the case of a monitor, a crime in progress is stopped. The lender's recovery from actually imposing the legal sanction without monitoring is  $X_\phi$ , and the recovery from a monitor actually stopping a crime in progress is  $X_m$ . The maximum incentive compatible payments for the various regimes of sanctions (automatic or voluntary) with  $\mu=1$  and when  $\mu=0$  are given in Table 1.

Table 1: Maximum Incentive Compatible Payments that deter both diversion and renegotiation. The borrower gets a fraction $\mu$ of the surplus from renegotiation with a lender (or a delegated monitor gets a fraction $\mu_{DM}$ and depositors get a fraction $1-\mu_{DM}$ ). This table provides the face values when $\mu$ and $\mu_M$ are both either 0 or 1.				
	$\mu=\mu_{DM}=1$ (no automatic sanctions)	$\mu=\mu_{DM}=1$ automatic default sanctions ( for deposits and loans)	$\mu=\mu_{DM}=1$ automatic default sanctions on deposits only	$\mu=\mu_{DM}=0$ , for all cases of sanctions (automatic or voluntary)
No monitoring	$F_1=X_\phi$	$F_1=H(t+\phi)$	$F_1=X_\phi$	$F_0=H(t+\phi)$
Undelegated Monitoring	$F_1=\max\{X_m, X_\phi\}$	$F_1=H(t+\phi)+\max\{X_m-X_\phi, 0\}$	$F_1=\max\{X_m, X_\phi\}$	$F_0=H(t+\phi+m)$
Delegated Monitoring (maximum payment to depositors)	$B_1=X_\phi$	$B_1=H(t+\phi)+\min\{\max\{X_m-X_\phi, 0\}, \phi_M H\}$	$B_1=H(t+\phi)+\min\{\max\{X_m-X_\phi, 0\}, \phi_M H\}$	$B_0=H(t+\phi+\min\{m, \phi_M\})$

Table 1 is a convenient summary of the maximum incentive compatible payments lenders can receive in the cases involving default sanctions when one party gets the entire bargaining surplus. The payoffs with  $\mu=\mu_{DM}=0$  are from Chapter 1, Propositions 1 and 2. The payoffs with  $\mu=\mu_{DM}=1$  and no automatic sanctions are from Section IV.B, and the

payoffs with automatic sanctions for both borrower and delegated monitor defaults are from section IV.A. The payoffs from delegated monitoring with automatic sanctions on deposits only are from section IV.B.1.

The following lemma allows a simple way of describing these payments when the borrower has an intermediate amount of bargaining power.

**Lemma 1:** The maximum incentive compatible payment that a borrower with cash flow of  $H$  will make without preferring diversion is the product of  $\mu$  and the payment which gives the lender his outside option plus the product of  $1-\mu$  and the payment that gives the borrower his outside option.

Proof: Let  $O_{lender}$  be the lender's outside option and  $O_{borrower}$  be the outside option of the borrower. The borrower's payoff from diversion is  $O_{borrower} + \mu(H - O_{borrower} - O_{lender})$  because the total surplus is  $H$  if an agreement not to impose the inefficient sanction is reached while imposing the sanction gives total surplus of  $O_{lender} + O_{borrower}$ . Let  $F_\mu$  denote the  $F$  such that  $H-F$  equals the payoff from threatening diversion, thus  $F_\mu = \mu O_{lender} + (1-\mu)(H - O_{borrower})$ . To show that  $F_\mu = \mu F_1 + (1-\mu)F_0$  note that  $F_1 = O_{lender}$  and  $F_0 = H - O_{borrower}$ . The result follows immediately.

QED.

## V.A Feasible financial structures

Borrowers need to raise capital by committing to pay lenders an amount  $I$ . A comparison of the amount that must be paid,  $I$ , to the largest incentive compatible payment using results in Table 1, Lemma 1, and Proposition 3 determines how lending must be structured. Any structure that will not allow the borrower to commit to pay  $I$  is infeasible. Proposition 5 characterizes the maximum incentive compatible payments for

general bargaining power. It adds one additional category, for completeness, that up to now has been assumed to be infeasible: an all equity capital structure without rights to impose default penalties. A borrower can get  $H(1-t)$  from diverting if there is no default penalty, and therefore he will pay up to  $Ht$  without diverting. If investor protection laws are very strong or diversion is very costly, this form of financing will be available.

**Proposition 5:** The possible financial structures and their largest incentive compatible payments to ultimate lenders (lenders when there is not delegated monitoring and depositors when there is delegated monitoring) when the borrower has bargaining power  $\mu$  (over lenders) and the delegated monitor has bargaining power  $\mu_{DM}$  (over depositors) with appropriately selected face values  $F$  (and  $B$  for delegated monitors) are as follows:

- a) All equity:  $tH$
- b) Non-demandable unmonitored debt:  $\mu X_\phi + (1-\mu) H(t+\phi)$
- c) Demandable unmonitored short-term direct debt:  $H(t+\phi)$
- d) Undelegated monitored debt:  $\mu \max\{X_\phi, X_m\} + (1-\mu) H(t+\phi+m)$
- e) Delegated monitor without demandable deposits (maximum payment to depositors):  $\mu_{DM} X_\phi + (1-\mu_{DM}) H(t+\phi + \min\{m, \phi_M\})$
- f) Delegated monitor with demandable deposits (maximum payment to depositors):  $H(t+\phi) + \min\{ \phi_M H, \mu \max\{X_m - X_\phi, 0\} + (1-\mu) (mH) \}$ .

Proof: Parts a through d follow directly from Lemma 1 and the results summarized in Table 1. Part e follows from these results and Proposition 3 because a delegated monitor who monitors a loan with a sufficiently high face value will be able to negotiate with depositors who cannot run to reduce their deposits to  $\mu_{DM} X_\phi + (1-\mu_{DM})$

$H(t+\phi + \min\{m, \phi_M\})$  (the sufficiently high face value of the loan must be at least

$H(t+\phi+\min\{m,\phi_M\})$ ). A delegated monitor with deposits less than or equal to  $\mu_{DM} X_\phi + (1-\mu_{DM}) H(t+\phi+\min\{m,\phi_M\})$  will not be able to get concessions from depositors. A delegated monitor with deposits of  $\mu_{DM} X_\phi + (1-\mu_{DM}) H(t+\phi+\min\{m,\phi_M\})$  can collect a payment of this amount from a borrower by Proposition 3, independent of  $\mu$ . Part f follows from Proposition 3 because a bank with this level of demandable deposits cannot get depositors to reduce their face value, and this level of deposits implies that a delegated monitor can collect a loan with this face value independent of  $\mu$ .

Q.E.D.

The largest incentive compatible payments are ranked as follows:  $a < b \leq c$ ,  $e \leq f$ , and  $e \leq d$  if  $\mu_{DM} \leq \mu$ . These results describe the constraints on available financial structures in the economy. If multiple structures are feasible, then other considerations are needed to deliver a prediction of the choice in an economy. These other considerations include uncertainty, the risk of financial crises, costs of monitoring and providing incentives for delegated monitoring under uncertainty. These are considered in chapter 3.

In the case of certainty we consider here, it is useful to show how much a delegated monitor can commit to pay investors when some but not all liabilities are demand deposits. Corollary 1 describes the maximum incentive compatible payments to depositors of financial structures that combine demand deposits with longer term, non-demandable claims.

**Corollary 1.** A delegated monitor with outstanding demandable deposits with face value  $B$  such that  $B \leq H(t+\phi) + \min\{ \phi_M H, \mu \max\{X_m - X_\phi, 0\} + (1-\mu)(mH) \}$ , and  $B \geq \mu X_\phi + (1-\mu) H(t+\phi)$  as well outstanding liabilities to investors that are non-demandable debt or equity can commit to pay total liabilities (including deposits) of up to

a maximum of  $B + (1-\mu_{DM})(H(t+\phi+\min\{m,\phi_M\}) - B)$ .

Proof: Demandable deposits of this amount imply that deposits cannot be reduced by negotiation. Any negotiation about payments from the delegated monitor takes place with the owners of non-demandable capital. The outside option of the delegated monitor (if he follows through with his threat not stop the borrower from diverting) is to default and incur penalties. A total verifiable payment of  $H(t+\phi+\min\{m,\phi_M\})$  to depositors and owners of non-demandable capital provides the monitor the identical payoff, because this is the largest verifiable payment the monitor can collect from the borrower without preferring to default and accept an unverifiable side payment from the borrower. The outside option of the owners of non-demandable capital is zero because if the monitor does not intervene they have access only to default penalties  $X_\phi$  when negotiating with the borrower. They will collect from the borrower a total of either  $\mu X_\phi + (1-\mu)H(t+\phi)$ , if  $H(t+\phi) < B$ , or  $B$ , if  $B \leq H(t+\phi)$ . By lemma 1, the total payment to capital holders is  $(1-\mu_{DM})(H(t+\phi+\min\{m,\phi_M\}) - B)$ .

Q. E. D.

Corollary 1 provides payoffs to outside investors for capital structures which include both demandable short-term deposits and longer term bank capital. A similar result follows immediately for borrowers who borrow directly without monitoring, issuing both demandable claims and longer term non-demand debt or equity claims. I do not provide the proof, but a nearly identical argument implies that a borrower with demandable debt  $F$  which satisfies  $F \leq H(t+\phi)$ , and  $F \geq \mu X_\phi + (1-\mu)H(t+\phi)$  as well as outside non-demandable capital can commit to pay investors up to a total of  $F + (1-\mu)(H(t+\phi) - F)$ .

Capital structures with these combinations of short and longer term claims are useful under uncertainty. These are analyzed in Chapter 3.

The model in this Chapter uses important related theories developed previously. It is useful to explore the relation between my results and the previous results. This is analyzed in Section VI. Readers interested only in the empirical implications about financial structure and financial systems can proceed to the conclusion and then to Chapter 3.

## **VI. Related Models and Approaches**

### **VI.A The Diamond-Rajan [2001] Model**

Diamond-Rajan [2001] presents a model that is similar to that in Sections IV and V, but is based on Hart-Moore [1994], instead of the monitoring model from Chapter 1. It has very similar implications to the model in section IV with high levels of borrower and delegated monitor bargaining power. Diamond-Rajan [2001] does not model legal systems, but their approach yields similar predictions to the monitoring model with strong legal protection and this high bargaining power of borrowers. In Diamond-Rajan [2001] there is no diversion or monitoring, and the timing of bargaining is slightly different.

Diversion is assumed to be impossible but borrowers can renegotiate debt to seek concessions from lenders. Instead of diverting, a borrower can choose not to fully repay debt by threatening to quit, withholding the skills needed to produce the cash flow  $H$  from the firm's assets. I use the notation of this book rather than that in Diamond-Rajan [2001]. Without the borrower's skills (human capital), a lender indexed by  $i$  can produce only  $X_i < H$  if he forecloses and redeploys the borrower's physical capital. If the borrower

has all the bargaining power ( $\mu=1$ ), his threat to quit will force the lender to reduce the required payment debt to  $F=X_i$ . With more general bargaining power, where the borrower obtains a fraction  $\mu$  of the surplus, the threat could reduce the required payment to  $\mu X_i + (1-\mu)(H)$ .

Diamond and Rajan [2001] model the timing of bargaining somewhat differently, following Hart and Moore [1994]. They assume that if the borrower threatens to quit, the lender must first decide whether to redeploy the assets (liquidate for  $X_i$ ), before negotiating with the borrower. If the lender chooses not to liquidate, then this period, the project cannot be operated without the borrower's skills. As a result, the lender then negotiates to divide the surplus which is the total payoff,  $H$ , from the borrower operating the project. In this case, negotiations allow the borrower to force the lender to reduce the required debt payment to the larger of  $X_i$  and  $(1-\mu)H$ . This timing assumption has similar implications to that made in Section IV. Under either assumption, if borrower bargaining power  $\mu$  is high, then liquidation value matters most; if it is low, then it hardly matters. In Diamond-Rajan [2001], if one assumes that bargaining power,  $\mu$ , is sufficiently high, then the specific liquidation value of investor  $i$ ,  $X_i$ , determines how much a borrower is willing to pay without renegotiating. The payoffs from the Diamond-Rajan model are then identical to the model in section IV when the borrower has all the bargaining power ( $\mu=1$ ). Because it is assumed that borrowers cannot divert funds, but only threaten to not produce output, a borrower gets a payoff of zero if assets are liquidated. This is equivalent to the assumption of strong legal protection ( $\phi+t=1$ ) in the monitoring model (where the borrower's payoff from threatening diversion is  $H(1-t-\phi)$ ).

In the Diamond-Rajan model, lenders can extract larger payments from borrowers

by learning how to achieve high liquidation values. This “relationship lending” plays the role of monitoring by increasing the amounts that borrowers will choose to repay. The relationship lender, indexed by  $m$  (to use the same notation as the monitoring model), has the highest liquidation value,  $X_m$ , for the asset. Similarly, if the liquidation values,  $X_i$  differ across lenders, an asset is illiquid if it must be sold to a lender with a lower liquidation value than the original lender. The equivalent of delegated monitoring in this framework is the delegation of relationship lending to acquire a high value of  $X_m$ . A relationship lender learns about the firm to achieve a high  $X_m$  but does not have funds of his own to lend. The relationship lender borrows from unskilled lenders with low values of  $X_i < X_m$ . However, if the borrower threatens to quit, the relationship lender can threaten the unskilled lenders to not use his loan collection skills and leave them to collect  $X_i$ . Because there is not automatic liquidation on default, without a device to commit to liquidation, the unskilled lenders will not collect more from the relationship lender than directly from the borrower. If unskilled lenders could commit to take ownership of the relationship lender’s loans (and collect from the borrower the low value that they can obtain) if the relationship lender ever threatens to quit, the relationship lender would not threaten and would instead pay whatever amount he originally promised. Demandable deposits can commit depositors to run and not make concessions.

However, the mechanism in Diamond-Rajan by which the threat of runs serves as a commitment device differs from the monitoring model of the previous sections of this Chapter. The way that it differs also has implications for the use of demandable claims by banks versus short-term debt issued directly by borrowers. This is explored in the next section.

### **VI.A.1      Should Banks or Borrowers Use Short-Term Claims Subject to Runs?**

The externality across lenders imposed by demandable debt may allow lenders (or depositors) to use the threat of runs to commit not to make concessions. In the monitoring model described in this chapter, a run invokes a legal sanction that reduces a borrower's (or delegated monitor's) gain from diversion. In the Diamond-Rajan model, the only enforcement device is the ability to liquidate assets. Diamond and Rajan assume that runs can reallocate ownership of claims on a first-come-first-served basis, but can not cause immediate liquidation of physical assets *during* a run. If, instead, a run commits lenders to immediate and irreversible asset liquidation during the run, then the threat of runs is a very powerful commitment device that deters borrowers from seeking unneeded concessions (enforced by the borrower's threat to quit). If lenders liquidate the asset, the borrower gets a payoff of zero. In this case where immediate liquidation is caused by a run, short-term demandable debt issued directly by a borrower would deter the borrower from seeking any concessions, avoiding the need for a relationship lender. An interesting model that uses this approach is presented in Von Thadden, Berglof, and Roland (2003). If runs can only reallocate ownership of claims on a first-come first-served basis, as Diamond and Rajan assume, then demandable debt is useful only for banks (relationship lenders).

If a run only reallocates ownership of claims on a borrower, then a run will not directly cause liquidation. After a run, negotiations can occur. The original borrower is the best person to run the firm's assets and will end up retaining and managing the assets. The amount paid to a given lender will depend only on the lender's liquidation value  $X_i$  and bargaining power. The run does not change the liquidation value  $X_i$  or the

bargaining power. As a result, runs on the firm's short-term debt do not discipline entrepreneur borrowers (this is similar to the inability of runs to commit borrowers to not divert when legal protection is weak). If there is not a relationship lender (that is, if all lenders have identical liquidation values), then a run that reallocates the ownership of debt (and its rights to liquidate) has no effect on the borrower. However, runs can discipline a relationship lender (a banker) with asset redeployment skills (high  $X_i$ ) by removing his rights to bargain to collect loans.

A run forces the loans, and thus the rights to negotiate with the borrower, to be reallocated away from the relationship lender. If the loans themselves are given directly to those who run, they are transferred to those who get there first. If the loans are sold to finance payments to those who run, then the loans are allocated to whoever buys them. In either case, the run removes the relationship lender's right to negotiate with the borrower. The run may or may not destroy the relationship lender's loan collection skills (high  $X_i$ ). If the skills are directly destroyed, then a run will discipline the relationship lender (who will get a payoff of zero after a run). If the skills persist, the next paragraph shows that he still will get a payoff of zero after a run.

Absent a run or a call for reductions in  $B$ , the banker has the exclusive right to negotiate with the borrower, and he keeps the residual amount collected in excess of the amount owed to depositors,  $B$ . After a run, the depositors or loan buyers own the loans and have the exclusive right to negotiate with the borrower, but they have the option of later entering into negotiation with the banker to collect the loan (or inviting the banker into multiparty negotiations with them and the borrower). If the depositors/loan buyers choose to wait and enter these negotiations that include the banker, any offers that the

borrower made previously are off the table. The payoffs to each party, in particular the payoff to depositors/loan buyers, from these later negotiations is known in advance, and they serve as an outside option in the negotiation with the borrower. As a result, the depositors/loan buyers and the borrower reach an agreement without going to the final stage, and this gives the banker zero.

The relationship lender does not add value but can only transfer value from the borrower. As a result, the threat of a run which removes the relationship lender's ownership rights serves to commit the relationship lender to collect the loan on behalf of depositors. In contrast, the threat of runs does not commit borrowers issuing debt directly unless the runs cause immediate liquidation.

#### **VI.A.2 Diamond Rajan: The Liquidity Risk of Undelegated Monitoring/ Borrowing from a Relationship Lender who is not a Bank**

Relationship lending (or in this Chapter's terminology, delegated monitoring) by banks financed by demand deposits allows banks to create liquidity. Relationship lending from a non-bank lender (undelegated monitoring) will not allow this liquidity creation. Diamond-Rajan [2001] analyzes the consequences of a liquidity shock received by a non-bank relationship lender. We consider a relationship lender who might experience an entrepreneurial liquidity shock, as in section I.C, and may have access to an investment opportunity in the future with a high rate of return, but which he cannot finance externally. A good example is trade credit, a relationship loan made by an entrepreneurial business lending its own money. The lender's business may need funding, and the lender must have some rights to intervene due to need to commit the borrower to repay the loan.

If the non-bank lender does not experience a liquidity shock, then he like all others in the economy has a discount factor of 1, and the present value of one unit of future consumption is one unit today. If the lender gets a liquidity shock, it occurs on date 1, and he discounts the future such that one unit of date 1 consumption is worth  $R > 1$  units of date 2 consumption. The lender has access to an investment in his business that has a rate of return  $R$  and which cannot be financed externally. This shock occurs with probability  $\theta$  ( $\theta$  plays the role of  $t$ , the probability of a liquidity shock, in the Diamond-Dybvig (1983) model).

There are three time periods. On date 0, the borrower needs to raise capital of 1. If there is no renegotiation (or diversion), the borrower will produce cash flows of  $H_1$  at date 1 and  $H_2$  at date 2 with the capital. A relationship lender (monitor) can redeploy capital (achieve a recovery if he intervenes) for  $X_{M1}$  at date 1 or  $X_{M2}$  at date 2. Any other lender can redeploy assets for lower amounts:  $X_{\phi 1}$  at date 1 or  $X_{\phi 2}$  at date 2. For simplicity, lenders and depositors have none of the bargaining power,  $\mu = \mu_M = 1$ , and their redeployments values (outside options) determine the amount they are paid. It is assumed that  $\max \{X_{M1}, X_{M2}\} > 1$  and  $\max \{X_{\phi 1}, X_{\phi 2}\} < 1$ . As a result, relationship lending (monitoring) is needed to commit the borrower to pay a sufficiently large amount.

If the borrower threatens to quit (divert) at any time between dates 1 and 2, the borrower can force the relationship lender to reduce the required payment to  $X_{M2}$  if it exceeds this amount. In addition, the borrower can threaten to quit (divert) at any time before date 1. If the monitor intervenes and liquidates the borrower's assets before date 1, it destroys all future cash flows the borrower can produce. The assets can be redeployed for  $X_{M1}$  on date 1, or the relationship lender can retain the assets and redeploy

them for  $X_{M2}$  on date 2. When the relationship lender does not get a liquidity shock, he can redeploy assets for a present value of  $\max \{X_{M1}, X_{M2}\}$ , by choosing the best time to redeploy the assets. As a result, the borrower will pay up to a total of  $\max \{X_{M1}, X_{M2}\}$  in total payments rather than threaten to quit (divert). In this case, the borrower can commit to pay up to  $F_2 \leq X_{M2}$  on date 2 and a total on both dates of  $F_1 + F_2 \leq \max \{X_{M1}, X_{M2}\}$ . If the monitor is not subject to a liquidity shock, this allows the borrower's investment to be financed.

If borrowing from someone other than the monitor, the borrower can commit to pay on date 2 only up to  $F_2 \leq X_{\phi 2}$  and a total on both dates of  $F_1 + F_2 \leq \max \{X_{\phi 1}, X_{\phi 2}\} < 1$ .

### **VI.A.3 Borrower Liquidity Risk if the Relationship Lender Gets a Liquidity Shock.**

If the relationship lender gets a liquidity shock, he discounts the future such that one unit of date 1 consumption is worth  $R > 1$  units of date 2 consumption. To deter him from liquidating and redeploying assets, the borrower must offer him payment equal in present value to redeploying assets, or  $\max \{X_{M1}, X_{M2}/R\}$  in date 1 value. Note that if immediate liquidation yields the highest present value ( $X_{M1} > X_{M2}/R$ ) and if the borrower's project generates too little cash on date 1, the borrower may be liquidated because he cannot pay enough at date 1. To see why, consider the three ways that the lender can deal with the loan. The borrower can pay  $H_1 + X_{M2}/R$  in present value by paying  $H_1$  and promising  $X_{M2}$  at date 2. Alternatively, the relationship lender (undelegated monitor) could sell the date 2 part of the loan to a non-relationship lender for  $S = X_{\phi 2}$  today and receive at total of:  $H_1 + X_{\phi 2}$ . If all of these are less than the smaller of  $X_{M1}$  and the amount that the borrower owes at date 1, the borrower will be liquidated at date 1 for

$X_{M1}$ . If the last option occurs, it illustrates the borrower's liquidity risk of borrowing (with trade credit) from a relationship lender (undelegated monitor).

If the borrower had borrowed from a bank (delegated monitor), he would not face liquidation because the bank could borrow  $X_{M2}$  from someone without a liquidity shock, rather than liquidate. The bank would never liquidate to invest in another loan because he could borrow against the old loan (or the new loan) with demand deposits. A bank (delegated monitor) does not invest in projects other than relationship (monitored) loans, and these can be financed with demand deposits. Banks' deposits commit them to collect loans and give them access to liquidity. This allows them to insulate borrowers from liquidity shocks and provide a stable source of funds.

This illustrates an important point about separating banking from commerce. Banks that invest only in loans and are solvent will be more reliable than lenders with fleeting in-house business opportunities to finance. This is related to the point in Section 5.4 of Chapter 1 on the differences between conglomerates and banks.

#### **VI.A.4      Liquidity Risk from the Lender's Point of View**

A non-bank lender making a relationship loan at date 0 who is subject to a liquidity shock at date 1 will require either a high promised expected repayment if the loan cannot be called and liquidated at date 1 or the right to liquidate the loan as needed when he gets a liquidity shock on date 1. Loans are illiquid assets and deliver low returns when a relationship lender needs liquidity on date 1. A high unconditional expected return is required on an illiquid loan that will never be called and liquidated at date 1. When liquidation yields the largest proceeds on date 1, but the lender makes a loan that

does not allow this liquidation, he must be compensated because the lender places a higher value on proceeds at date 1 when he gets a liquidity shock.

If the relationship lender's loan collection skills are sufficiently important, the loan will be illiquid. Either the borrower or the lender will bear the consequences if a relationship lender is subject to liquidity shocks. The borrower will prefer to give the lender his required return without being liquidated (no matter what the probability of the liquidity shock) if the rate of return of the project, relative to its liquidation value, exceeds the relationship lender's rate of return given a liquidity shock, or  $\frac{H_1 + H_2}{X_{M1}} > R$  (which is assumed). Under this assumption, borrowers will issue debt contracts to monitors that lead to liquidation on date 1 only if this is required to give the lender his required return.

#### **IV.B Relation to Calomiris Kahn (1991)**

Calomiris and Kahn (1991) was the first study to link first-come-first-served demand deposits to providing incentives for the monitoring of banks. Calomiris and Kahn (1991) views demand deposits as a way to reward individual depositors who collect costly information about the local economy that predicts when a bank will fail. Those who can predict bank failure get out first and achieve a higher return than those who leave their money in a failing bank. The banker might steal the depositors' money (equivalent to diverting funds as in Chapter 1), and this theft can be deterred only if the bank's assets are liquidated *before* the diversion occurs. Liquidation deters a crime about to occur, rather than stopping a crime in progress. Monitoring involves the costly observation of variables that predict future payoffs of the bank's investments, because the banker will steal when future payoffs are low and his residual equity claim is of little

value. The role of short-term or demandable claims is to allow liquidation when bad news about the bank indicates that theft would soon occur.

If one depositor monitoring the state of the local economy produces enough information to close banks at the appropriate time, then giving that depositor alone a demandable claim (or a payment for revealing that bad times were coming) suffices to stop the banker from stealing. Calomiris and Kahn (1991) explains first-come first-served deposits with multiple depositors by the need for more information than one depositor can obtain. Each depositor observes a different noisy signal about the bank's future opportunities, and many depositors must monitor signals about future opportunities in order to get a good prediction of the future. A large number of depositors attempting to withdraw indicates many negative signals which predicts such poor future opportunities that the banker is likely to abscond. As a result, the bank should be liquidated.

The information role of short-term demandable deposits is to determine when depositors in aggregate have bad news about the future, similar to the information aggregation role of secondary markets for equity in Grossman (1976), Hellwig (1980) and Diamond-Verrecchia (1981). Demand deposits can provide stronger incentives for depositors to monitor information than can highly informationally efficient markets for bank equity. This incentive is important because monitoring is costly and unobservable. The first-come-first-served aspect gives a higher return to those who monitor the signal about the bank's future prospects because when they choose to withdraw they get out, on average, before depositors who did not monitor (and thus did not withdraw). This is because all who withdraw before the bank is closed get full value, while the remaining

depositors take a loss due to the costs of liquidating the bank's assets. This provides an incentive for many depositors to monitor signals about the bank's prospects, but only for the depositors with low monitoring costs.

The mechanism differs from that later proposed in Diamond-Rajan (2001), although it is similar in spirit. Calomiris and Kahn (1991) propose a way to provide incentives for the costly monitoring by depositors of the local economic conditions faced by bankers. Calomiris and Kahn (1991) assumes that the closing of the bank increases the depositors' ex-post return, because if not closed the banker will be likely to steal all of the deposits. Diamond and Rajan (2001) focus on the need to provide an incentive for depositors to run on the bank whenever it looks like they will bear a loss, even when this will increase their collective loss. Diamond and Rajan (2001) examines bank withdrawals caused by public information about the banker's actions, but the two models could be combined to cover the case of runs started by withdrawals due to costly monitoring by large depositors.

The model in Calomiris and Kahn (1991) does not have bank borrowers (bank assets are real investment projects). As a result, the logic of their results also applies to demandable short-term debt issued directly by a firm, if the firm is expected to misbehave if bad times are likely. A combination of their results with this book's model of monitoring and legal protection could help refine this prediction.

#### **IV.C Debt Priority and Liquidation Incentives as an Alternative to Runs**

Diamond [1993, 2004], and Park [2000] examine contracts where debt contract priority influences the liquidation incentives of lenders. If some lenders cannot negotiate with the borrower or the other lenders, then assigning debt priority is an alternative to

demandable debt as a way to improve a monitor's incentive to intervene. These models do not have bargaining (because the borrower cannot commit not to misbehave, even briefly). The monitor decides to liquidate (or intervene more generally) only after the borrower has already taken his irreversible action. The borrower will choose the efficient action (and not misbehave) if and only if the monitor will liquidate conditional on misbehavior. If the recovery from liquidating when the borrower misbehaves is  $X_m$  and if the payoff from not liquidating exceeds  $X_m$ , then a lender, who is the only lender or one with the same priority as others, will not liquidate. Anticipating this, the borrower will misbehave. Providing a senior claim to a monitoring lender (giving priority over the recovery from liquidation) and a junior claim to the lenders who cannot negotiate increases the incentive for a monitor to liquidate, as compared with equal priority claims, or sole funding by the monitor. The payoff that a monitor with senior debt of face  $F$  achieves from liquidating is  $\min\{F, X_m\}$ . The losses from intervention are borne by the other lenders (or by the borrower if the only junior claim is retained by the borrower). Making the monitor provide only part of the capital and hold a senior claim improves the ability of the monitor to deter diversion. The capital structure needs to be set carefully because the senior claimant intervenes to avoid future losses to his claim. Park [2000] shows that the senior claimant should face some risk of loss in the future if the borrower misbehaves.

Berglof and Von Thadden (1994) show a related result describing how seniority can improve a lender's bargaining position. They assume that lenders negotiate only sequentially with the borrower: the first to negotiate is effectively senior to the others, and the borrower cannot negotiate with all lenders simultaneously. This has the borrower

negotiate with a series of lenders each with a higher outside option than they have collectively, and this commits the borrower to pay more than if he negotiated with all lenders simultaneously.

In contrast to these priority-based models, the threat of a run makes the subset of lenders who refuse to make concessions senior to those who make them. As a result, none make concessions even though all are available to bargain with the borrower.

### **III. Conclusion**

The model of monitoring and legal protection has implications for an economy's financial structure, the importance of banks in some economies, and the use of short-term debt or deposits. It explains why banks who monitor loans issue demand deposits subject to runs and why bank loans are illiquid. If one assumes that bank assets are illiquid, Diamond-Dybvig [1983] shows that banks can create assets that provide investors with more liquidity than holding illiquid assets directly. When there is a demand for more liquid assets from investors or entrepreneurs, demand deposit contracts serve as a means for quick access to liquidity. Demand deposits work very well when investors forecast that banks will survive, but can cause severe damage if investors lose faith in banks. There is scope for banks to write more refined contracts, such as deposits with suspension of convertibility of deposits to cash. In addition, there may be a role for government policies that eliminate self-fulfilling runs on banks. The role of government is due to its taxation authority that is not available to private firms.

Explaining why bank loans are illiquid gives a much fuller view of banks and bank runs. A financial asset is illiquid if it cannot be sold for the present value of the future cash flows that would accrue to the seller. When there is lender specificity of loans due

to specialized monitoring or asset redeployment skills, loans will be illiquid. The monitoring and legal protection model developed in Chapter 1 shows that when legal protection of investors and creditors is weak, specialized monitoring skills will be required to allow borrowers to commit to repay lenders. It also shows how this monitoring can be delegated: the imposition of legal sanctions for default and the loss of loan negotiation authority when a delegated monitor defaults on deposits. In this Chapter, depositors may have difficulty committing to impose these legal sanctions of making the bank fail precisely because the loans are illiquid: the loans become worth less once the bank fails and is not monitoring and collecting the loans. Demand deposits issued by the bank build in an externality (a manufactured collective action problem) which commits depositors to run whenever they anticipate any loss of value, as in Diamond-Rajan (2001).

The monitoring and legal protection model shows that short-term debt can be important at the firm level (to induce firm runs), but that this is unlikely in economies with poor investor and creditor legal protection, where banks are required. The model predicts important causal links between capital structure, bargaining power and short-term debt subject to the threat of runs. This model does not require any uncertainty for its predictions. Detailed analysis of enforcement of contracts and the incentives to use the legal system delivers a simple theory of financial structure which delivers its predictions based on the characteristics of the legal system.

Two related aspects of the determination of financial structure are missing from the analysis of Chapter 2: the roles of uncertainty and of financial crises with causes other than panics. These are analyzed in Chapter 3.

## References

- Barth, James R, Gerard Caprio, and Ross Levine [2006] *Rethinking Bank Regulation: Till Angels Govern*, New York: Cambridge University Press, 2006.
- Berglof, Erik, and Ernst-Ludwig von Thadden, 1994, Short-term versus long-term interests: Capital structure with multiple investors, *The Quarterly Journal of Economics* 109, (4), 1055-1084.
- Bryant, John [1980], "A Model of Reserves, Bank Runs, and Deposit Insurance," *Journal of Banking and Finance*, 4: 335-344.
- Cooper, Russell and T.W.Ross [1998] "Bank Runs: Liquidity Costs and Investment Decisions." *Journal of Monetary Economics* 41(February):27–38.
- Diamond, D. W. [1984], "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies* 51 (July): 393-414.
- Diamond, D. W. [1996] "Financial intermediation as delegated monitoring: a simple example," *Economic Quarterly*, Federal Reserve Bank of Richmond, issue Sum, pages 51-66.
- Diamond, Douglas W. "Liquidity Banks and Markets." *Journal of Political Economy*, 1997, 105(5), pp. 928-56.
- Diamond, D. W. [2007], "Banks, Liquidity and Legal Protection," The 2005 Princeton Lectures in Finance.
- Diamond, Douglas W. and Dybvig, Philip H. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy*, 1983, 91(5), pp. 401-19.
- Diamond, Douglas W. and Rajan, Raghuram. "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking." *Journal of Political Economy*, April 2001, 109(2), pp. 287-327.
- \_\_\_\_\_. "Liquidity Shortages and Banking Crises," *Journal of Finance*, April 2005, 60 (2), pp. 615-647.
- Diamond, Douglas W. and Robert E. Verrecchia, 1981, "Information Aggregation in a Noisy Rational Expectations Economy," *Journal of Financial Economics* 9, (September): 221-235.
- Ennis, Huberto M. [2003] "Economic Fundamentals and Bank Runs," *Federal Reserve Bank of Richmond Economic Quarterly* 89/2: 55-71.
- Ennis, Huberto M. and Keister, Todd, [2006] "Bank runs and investment decisions revisited," *Journal of Monetary Economics* 53(2): 217-232, March.
- Friedman, Milton and Anna Schwartz [1962] *A Monetary History of the United States 1867-1960*. Princeton: Princeton University Press, 1963.
- Hellwig, Martin 1980, "On the aggregation of information in competitive markets," *Journal of Economic Theory* 22:3 (March).

- Itay Goldstein and Ady Pauzner, "Demand Deposit Contracts and the Probability of Bank Runs," *Journal of Finance*, vol. 60(3), pp. 1293-1328, June 2005.
- Gorton, Gary, [1985] "Bank suspension of convertibility," *Journal of Monetary Economics* 15(2)177-193, March.
- Grossman, Sanford [1976] "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information" *The Journal of Finance*, Vol. 31, No. 2 (May,1976): 573-585.
- Holmström, B. and J. Tirole [1998], "Private and Public Supply of Liquidity," *Journal of Political Economy* 106 (February): 1-40.
- Jacklin, C. J. [1987], "Demand Deposits, Trading Restrictions, and Risk Sharing," in: E. C. Prescott, N. Wallace (eds.), *Contractual Arrangements for Intertemporal Trade*, Minneapolis: University of Minnesota Press, 26-47
- Haubrich, Joseph G. and Robert G. King [1990] "Banking and Insurance," *Journal of Monetary Economics* 26: 361-386.
- Morris, Stephen and Hyun Song Shin [2003] "Global Games: Theory and Applications," in *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)*, edited by M. Dewatripont, L. Hansen and S. Turnovsky. Cambridge, England: Cambridge University Press.
- Peck, James, and Karl Shell [2003] "Equilibrium Bank Runs," *Journal of Political Economy*: 111(February):103–23.
- Spear, Kathleen and Enrico Perotti, "Capital Structure as a Bargaining Tool: The Role of Leverage in Contract Renegotiation," *The American Economic Review*, Vol. 83 (5) (1993), pp. 1131-1141.
- Von Thadden, Ernst-Ludwig, Erik Berglof, and Gerard Roland, 2003, Optimal debt design and the role of bankruptcy, Working Paper, (August) Universite de Lausanne.

Appendix: Proof of Proposition 3.

The delegated monitor is owed a loan with face  $F \leq H$  and owes deposits  $B \leq F$  with automatic default sanctions of  $\phi_M H$ . If the borrower diverts, he will either offer to reverse diversion if the monitor accepts a lower face value  $F'$ , or he will offer the monitor an unverifiable side payment  $U$  from the diversion proceeds, if the delegated monitor agrees not to intervene (not stop the crime). If the delegated monitor accepts the new face value,  $F'$ , his payoff is  $F' - B$  if  $F' \geq B$  and is  $-\phi_M H$  if  $F' < B$ . If the borrower offers a share of diversion proceeds  $U$ , and the monitor accepts, his payoff is  $U - \phi_M H$ . The delegated monitor's payoff from rejecting an offer and stopping the crime with recovery  $X_m$  is  $\max\{X_m - X_\phi, 0\} - \phi_M H$  if  $B > \max\{X_m, X_\phi\}$ , because the recovery from ex-post default,  $X_\phi$ , is verifiable and accrues to depositors, but  $\max\{X_m - X_\phi, 0\}$  accrues to the delegated monitor (the case of  $B \leq \max\{X_m, X_\phi\}$  is discussed below). The monitor's outside option is positive if  $X_m - X_\phi - \phi_M H > 0$ .

The delegated monitor requires a payoff at least equal to his outside option. The monitor will not stop the crime if the borrower offers either a verifiable payment which satisfies both  $F' \geq B$  and  $F' - B \geq \max\{X_m - X_\phi\} - \phi_M H$ , which is  $F' \geq B + \max\{0, X_m - X_\phi - \phi_M H\}$ , or an unverifiable side payment which satisfies  $U - \phi_M H \geq \max\{0, X_m - X_\phi\} - \phi_M H$  or  $U \geq \max\{0, X_m - X_\phi\}$ .

The borrower will threaten to divert if making one of these payments is attractive or will actually divert if this is best without negotiation. The borrower prefers to pay the original  $F$  rather than get the monitor to reduce it to the lowest  $F'$  a borrower will accept if  $H - F \geq H - (B + \max\{0, X_m - X_\phi - \phi_M H\})$ , or  $F \leq B + \max\{0, X_m - X_\phi - \phi_M H\}$ . This gives the borrower a payoff of  $H - (B + \max\{0, X_m - X_\phi - \phi_M H\})$ . Another option is to offer a side

payment  $U = \max\{X_m - X_\phi, 0\}$  to the delegated monitor giving the borrower a payoff of  $H(1-t-\phi) - \max\{X_m - X_\phi, 0\}$ . Paying face value lowest face value acceptable to the delegated monitor,  $F' = B + \max\{0, X_m - X_\phi - \phi_M H\}$  is preferable to making this side payment if  $H - [B + \max\{0, X_m - X_\phi - \phi_M H\}] \geq H(1-t-\phi) - \max\{X_m - X_\phi, 0\}$ , which is true if  $B$  satisfies  $B \leq H(t+\phi) + \min\{\max\{0, X_m - X_\phi\}, \phi_M H\} \equiv \bar{B}$ .  $\bar{B}$  is the maximum incentive compatible  $B$ . Without the automatic sanctions on deposit default, only  $X_\phi$  would be paid to depositors. Actually diverting is not best if  $F' \leq H(t+\phi+m)$ , which is assumed.

The delegated monitor can collect a loan with face value  $F$ , if it satisfies  $F \leq \bar{B} + \max\{0, X_m - X_\phi - \phi_M H\}$ . As a result, if  $X_m - X_\phi > \phi_M H$  the delegated monitor can collect more than  $B$  from the borrower (because the delegated monitor has a positive outside option).

When  $B \leq \min\{X_m, X_\phi\}$ , the following applies. The amount that the delegated monitor can commit to pay is at least  $X_\phi$  because this is the depositors' outside option. If  $X_m > X_\phi$  and  $B \in (X_\phi, X_m)$ , the delegated monitor's payoff from accepting  $F' \geq B$  is  $F' - B$ , from accepting a side payment is  $U - \phi_M H$  and the outside option of rejecting one of these offers and stopping the crime is  $\max\{X_m - B, X_m - X_\phi - \phi_M H\}$ . If  $B < \min\{X_m, X_\phi\} \leq X_\phi + \phi_M H$ , the outside option is  $\min\{X_m, X_\phi\} - B$  and the borrower will pay up to  $F' = \min\{X_m, X_\phi\}$  and with  $B = \min\{X_m, X_\phi\}$ , the delegated monitor will pay up to  $B = \min\{X_m, X_\phi\}$  to depositors. If instead  $X_\phi + \phi_M H \leq B < X_m$  or if  $B < X_\phi + \phi_M H < X_m$ , the outside option is  $X_m - X_\phi - \phi_M H$  and the borrower will pay up to  $F' = B + X_m - X_\phi - \phi_M H$  instead of  $U = X_m - X_\phi$  if  $H(t+\phi+\phi_M) \geq B$ . Setting  $B = H(t+\phi+\phi_M)$  implies that the borrower will pay up to  $F' = H(t+\phi + X_m - X_\phi)$

QED.

The maximum date-1 present value the impatient relationship lender can extract (i.e., the largest renegotiation proof amount) is  $E^I = \text{Max}\{X_{M1}, S, \frac{X_{M2}}{R}\}$  while the maximum amount she can extract if patient is  $E^{-I} = \text{Max}\{X_{M1}, X_{M2}\}$ .

(i) *The entrepreneur will be financed at date 0 and the loan will be a liquid asset if*

$$\text{Min}\{H_1 + S, E^I\} \geq 1. \quad (1)$$

*If  $\text{Min}\{H_1 + S, E^I\} < 1$*

(ii) *The entrepreneur will be financed at date 0 but the loan will be illiquid and he will have to pay a positive illiquidity premium if either*

$$H_1 + S < E^I \quad \text{and} \quad E^{-I} \geq 1 + \frac{\theta}{(1-\theta)} R(1 - (H_1 + S)) \quad (2)$$

*or*

$$H_1 + S \geq E^I \quad \text{and} \quad E^{-I} \geq 1 + \frac{\theta}{(1-\theta)} R(1 - E^I) \quad (3)$$

*or*

$$E^{-I} \geq 1 + \frac{\theta}{1-\theta} \left( R(1 - \text{Min}\{H_1, E^I\}) - \text{Min}\{\text{Max}\{(E^I - H_1)R, 0\}, X_{M2}\} \right) \quad (4)$$

(iii) *If none of (2), (3), or (4) hold, the entrepreneur will be financed at date 0 but only with the asset being liquidated when the lender is impatient if*

$$E^{-I} > 1 + \frac{\theta}{(1-\theta)} R(1 - X_{M1}) \quad (5)$$

(iv) *The entrepreneur will not be financed at all at date 0 otherwise.*

A delegated monitor with demand deposits can make finance the loan if  $\text{Max}\{X_{M1}, X_{M2}\} \geq 1$ , which is the condition for a undelegated monitor without a liquidity shock ( $\theta=0$ ).

### E. An Example.

Let  $X_{M1} = 0.9, X_{M2} = 1.1, H_1 = 0.05, H_2 = 1.5, R = 1.4, S = X_{\phi 2} = 0.8$ . Note that the date 1 cash flow. Because  $Max[X_1, X_2] = 1.1$ , the entrepreneur cannot commit to paying more than 1.1 even though he generates 1.55 from the project. Moreover, when hit by a liquidity shock, the relationship lender gets more in present value by liquidating ( $X_{1M} = 0.9$ ) than by selling the loan and collecting all the date 1 cash ( $H_1 + S = 0.85$ ) or holding it to maturity ( $\frac{X_{2M}}{R} = 0.79$ ).

When the probability of the liquidity shock,  $\theta$ , is low, the relationship lender will write a loan contract with a low short-term payment equal to  $H_1=0.5$  and will sell the loan when hit by the shock in preference to retaining it. We have  $V_1^I = V_1^{\sim I} = P_1 = 0.05, V_2^I = S = 0.8$ , and  $V_2^{\sim I} = P_2$ , where  $P_2$  is set to satisfy the lender's IR constraint, and will rise from 0.95 to 1.1 as  $\theta$  increases from 0 to 0.26. Since, the entrepreneur cannot commit to paying the relationship lender more than  $P_2=1.1$ , when  $\theta$  increases beyond 0.26, the only way the entrepreneur can satisfy the lender's rationality condition is by allowing her to liquidate at date 1 and get 0.9 if she suffers the shock. So if  $\theta > 0.26$ , the entrepreneur will offer  $V_1^I = V_1^{\sim I} = 0.05, V_2^I = 0.9$ , and  $V_2^{\sim I} = P_2$ , where  $P_2$  is again set to satisfy the lender's IR constraint. Since liquidation generates more than a loan sale, the relationship lender will again find it rational to lend for the range  $0.26 < \theta < 0.42$ . But when  $\theta > 0.42$ ,  $P_2$  exceeds 1.1 and is again not collectible. At this point, the probability of a liquidity shock for the lender is so high that lending is not individually rational.