Delegated Monitoring and Legal Protection

Douglas W. Diamond
University of Chicago, GSB

June 2005, revised October 2006.

This is Chapter 1 of the 2005 Princeton Lectures in Finance, presented in June, 2005.

The material in Section 4 and in Sections 3.2 and 3.3 is an exposition of results from Diamond [1984, 1996], but the balance is new material.

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

Princeton Lectures in Finance, 2005
Chapter 1: Delegated Monitoring and Legal Protection

Douglas W. Diamond
Preliminary
Do Not Quote
June 2005, revised October 2006.

These lectures describe financial intermediation theory, which explains the functions of institutions such as banks. A financial intermediary issues financial claims to investors for the purpose of buying other financial claims. This indirect financing is very important both historically and today. The lectures describe and integrate previous analysis of intermediaries and develop some new theories of financial systems, integrating financial intermediation theory with recent research on the effects of law on finance. The field of law and finance addresses the effects of the legal system on financial contracting.

This is not a survey of all research in this area, nor a historical development of my own research on financial intermediation. I integrate my own research over then last twenty five years, some of which is coauthored with Philip Dybvig or Raghuram Rajan, with other related ideas and approaches, into some simplified and extended models. This can illustrate new links between the implications of these various theories.

Financial intermediation theory studies why indirect financing is used and explains the form of the contracts written by intermediaries. In addition, it explains which borrowers within a country will choose to borrow via intermediaries instead of raising their financing directly from investors. In retrospect, much of my own research and much of the literature has implicitly assumed strong legal protection of investors and creditors. This chapter develops a theory of financial intermediation for various degrees

of legal protection. It has new implications for the financial structure and contracts for countries with different legal systems.

The research area of law and finance, largely initiated by Shleifer and Vishny [1997], has produced many empirical and theoretical insights. The key empirical results are cross sectional predictions across countries about the effects of the legal system on the access to finance and the financial contracts used. Very little of this research has examined the contracts and structure of financial intermediaries. Legal systems vary in the consequences of fraud, the misappropriation of investors' funds, default on a debt contract, and in creditor property rights more generally. In addition to varying consequences, the costs of accessing courts and the level of corruption in the legal system differ across countries.

This first lecture examines the monitoring role of financial intermediaries. Monitoring is the observation of information or acquisition of skills that allow the monitor to deter an agent from taking a self-interested action. The self-interested action can be thought of very generally, but can most simply be referred to as theft. Monitoring allows theft to be deterred. This is most easily seen in the case of two parties, an agent who steals and a principal who can monitor to deter theft.

Monitoring is the observation of information that is not freely available to all, either because it is costly to observe or requires specialized skills to observe. I will distinguish monitoring from verification. Monitoring does not make information freely available to others, while verification makes the information available to all.

Before further describing the technology of monitoring, it is useful to set up the basic idea that if a firm needs monitoring and it raises funds from multiple investors, it may be beneficial to delegate the monitoring to one investor. This avoids duplication of

monitoring costs. This delegation of monitoring may give rise to problems of its own, which I refer to as delegation costs.

My earliest work on financial intermediation and optimal financial contracting focused on the role of diversification and on the use of debt contracts by borrowers and intermediaries. I implicitly assumed that legal protection was very strong in that it was possible to write and enforce a contract to deter a borrower's self-interested action as long the action could be exactly detected at no cost. The role of intermediaries when legal protection is strong is to reduce the cost of providing incentives to borrowers when information about the action is not freely available. When legal protection is weak**,** then even if it can be exactly detected without cost, it may not be possible to deter borrower theft (or other misdeeds). This occurs because the penalty is less then the spoils of the crime. In this environment, there are some differences in the role and structure of intermediaries and in the optimal form of their contracts.

The costs and benefits of diversification by banks depend on the strength of legal protection and the details of how it is applied to financial intermediaries.

The balance of this chapter proceeds as follows. Section 2 describes monitoring and how its benefit trades off against either the cost of monitoring directly or the cost of providing incentives for delegated monitoring. Section 3 defines the strength of legal protection and characterizes the best available financial contracts when legal protection is strong. Section 4 analyzes delegated monitoring and the role of banks when legal protection is strong and the cash flows received by borrowers are risky. Section 5, 6 and 7 study the effects of weak legal protection. Section 5 examines the role of monitoring when borrower cash flow is risk free and shows how to provide banks with incentives for delegated monitoring. Section 6 compares contracts that provide banks with incentives

to act as delegated monitors with contracts that impose joint liability on several borrowers. Section 7 analyzes delegated monitoring when the cash flows received by borrowers are risky. Section 8 presents conclusions.

Section 2: The Costs and Benefits of Monitoring

Before the monitoring based theory of financial intermediation was developed, the primary view was based on intermediaries as reducers of transaction costs. A production function for producing financial assets and liabilities was assumed, and banks and other intermediaries were analyzed using neoclassical production theory (for surveys see Baltensperger [1980] and Benston and Smith [1976]). We will see that there is a notion of reducing transaction costs in the model of delegated monitoring, but its primary focus is to understand why intermediation reduces costs, deriving the technology of the cost reduction.

Why add a layer of delegation between borrowers and lenders? The answer in a competitive market must be that the extra layer has benefits that exceed its costs. Another answer might be laws or other limitations that prevent direct access of borrowers to lenders, but this will not be the focus of my analysis. Before describing the details of how financial contracts should be written with and without a financial intermediary, it is useful to define some of the key issues.

I will focus on the benefits of monitoring, but the point applies more generally to net benefits of delegating another task that improves the efficiency of loan contracting. Investors and borrowers can contract directly without any monitoring. There are two other alternatives contracting arrangements to consider when there are many investors per borrower. One is for each investor to monitor the borrower. The other is for one lender to monitor the borrower on behalf of the other lenders, a situation that I will refer to as

delegated monitoring. The best contract is the best of these three. If monitoring is not worth its cost, then there will be direct lending without monitoring. If monitoring is worth its cost, then the question is whether direct monitoring by each investor is better than delegated monitoring. If there are costs, D, of providing the incentives for delegated monitoring, these must be subtracted from the increased benefit that the delegated monitor can provide. The increased benefit of delegated monitoring arises because of some combination of specialized skill (the agent delegated the monitoring is a better monitor and has a better monitoring technology) and reduction in the duplication of effort. This can be illustrated by the model of Diamond [1984]. Everyone is assumed to have access to the same monitoring technology, so any advantage of delegation of monitoring is due to reduced duplication of costly effort. Monitoring allows a gross improvement in contracting efficiency which is worth S and monitoring costs K. If there are n lenders, direct monitoring by each lender costs nK and dominates direct lending with no monitoring if S-nK>0.[1] Delegated monitoring is better than no monitoring if S–K-D>0, and it dominates direct monitoring if nK>K+D, which is equivalent to D<(n-1)K. Delegated monitoring is best if its net benefit, S-K-D, exceeds the larger of S-nK and zero.

Section 3: What is monitoring and how does it relate to the legal environment?

3.1 Contracts without monitoring.

To understand how monitoring can resolve incentive problems between borrowers and lenders, I begin by examining the best contracts without monitoring. Assume that borrowers receive cash flows from business operations and must voluntarily turn them

[1] Another option is non-delegated monitoring with less duplication of effort, analyzed in Winton [1995]. Winton considers multiple prioritized debt contracts, only some of which need monitoring. Because there is still duplicated monitoring, it is qualitatively similar to monitoring by all m investors. This option is not considered here, to avoid complicating the analysis.

over to investors. In the course of business, the borrower deals with customers and suppliers, and this allows the possibility of diverting cash flow to himself. To be concrete, if the borrower receives a cash flow, H, and if he does not divert any of it, the cash flow is verifiable and investors have access to H, up to the amount that they are contractually owed, F. If the borrower diverts, the diverted amount is unavailable and unverifiable to investors or courts. Because diverting cash can be costly due to covering one's tracks, a fraction $t \geq 0$ of the amount diverted is destroyed. If $t=0$, as in Diamond [1984], the borrower can steal at no cost. If $t>0$, the costs of diverting may include payoffs to accomplices, such as suppliers (as in Lacker-Weinberg [1989] and Calomiris-Kahn [1991].

The legal system allows borrowers and investors to write contracts which depend on verifiable quantities, such as the amount of cash actually paid by the borrower to the investors. This amount is observed after the borrower has had the opportunity to divert an unverifiable amount of cash. Because the cash payment to lenders can be used in a contract, legal actions contingent on the payment can be specified. These legal actions or sanctions on the borrower reduce the borrower's payoff by $\phi H \geq 0$. Proposition 1 describes how the strength of legal protection influences a borrower's incentive to repay or to divert funds.

Proposition 1 If the borrower has a cash flow of H and is supposed to pay an amount F to the lender, and if there is sanction of ϕH for all payments less than F, then he will make the payment if and only if it costs him less to pay investors than to incur the legal sanctions and costs of diversion, $F \leq (t+\phi)H$. This implies that his payoff after paying investors, H-F, weakly exceeds his payoff from diversion, $H(1-t-\phi)$. If $F>(t+\phi)H$, the borrower will divert funds and default on the debt.

The ex-post sanctions can be thought of in several ways. First, they may represent a legal penalty for fraud or for default. Second, they may represent the value of lost reputation (which is very important with weak legal protection): the value of lost rents from future business due to revelations of diversion. Finally, they may represent ex-post interventions that reduce the borrower's proceeds from diversion. All of these are useful to deter diversion even if they do not benefit the lender. To focus on this incentive effect, I will assume that the sanction is costly to impose because it is a non-pecuniary penalty that hurts the borrower without any benefit to the lender's recovery. One such example is the liquidation of the borrower's assets with a recovery rate of zero to the lender (as in Lacker [1991] and Diamond [1996]). The sanction is only useful as a threat to punish the borrower and deter diversion.

Actual cash payments are observable and can be written into contracts. As a result, in the case where cash flows are certain, any payment of less then the promised amount F (which is assumed to be less than or equal to H) indicates diversion. I define **strong legal protection** as that which can deter all diversion, or $\phi+t=1$. I assume strong legal protection for the rest of this section. In contrast, when there is **weak legal protection,** $\phi+t<1$, it may not be possible to deter diversion even when it can be detected ex-post.

For this chapter, I assume that contracts which impose penalties contingent on the amount paid to investors are fully enforced. I also assume that there is no cash recovered by the lender when the sanction is imposed. In addition, everyone is risk neutral, there is no discounting, and riskless interest rates are zero.

Proposition 1 shows that with strong legal protection, $t+\phi=1$, diversion by the borrower can be deterred by imposing the penalty if less than F is paid. Most obviously

this is interpreted as a debt contract with face value F. Under certainty, debt is identical to profit-sharing equity, where the borrower must pay out a share of known profits. As long as there is a court system that will enforce the penalty when too little is paid, the borrower will not divert. When cash flow is certain, the penalty need not be imposed because its prospect deters diversion. Therefore, the costs of imposing the sanction need not be incurred. However, once there is uncertainty about the amount of cash obtained by the borrower, sometimes the penalty may need to be imposed.

3.2 Debt contracts with no monitoring, uncertain and unobservable cash flows, and strong legal protection.

When there is uncertainty about the amount of business cash flows, even strong legal protection ($\phi+t=1$) need not provide a perfect way to eliminate diversion. The amount actually paid to investors must come from the realized cash flows, and a low payment is not a perfect indicator of diversion, because a low realized cash flow can force the borrower to default. When realized cash flows are uncertain and unobservable, how does one specify an optimal financial contract between investor and borrower when the contract depends only on the amount paid to investors? That is, for which payments should the sanctions be imposed? Because the sanction (such as liquidation or costly bankruptcy) delivers no recovery to the lender, one wants to impose it as little as possible, but instead use the prospect of a sanction to deter diversion.

The borrower needs to fund an indivisible investment project and has no funds of his own. To raise sufficient funding to undertake the project, the borrower needs to offer outside investors an expected repayment of I (for example the project costs I to fund and lenders require an 0% expected rate of return or, more generally, the project requires initial capital of $I/(1+r)$ and investors require an expected rate of return of r). The

project's realized cash flow is a random variable with realization denoted by C. The probability distribution of C, the value of cash flow from the project, is known to all borrowers and lenders, and is given by H with probability P, and L with probability 1-P. A useful example is I=100, H=160, L=90, P=2/3 and the monitoring cost K=2.

The sanction is best used as a payment-contingent sanction in the following way. If the lender is ever to impose the sanction for a given payment, he also should impose it for all lower payments. Suppose instead that the lender writes a contract that does not impose the sanction if L is paid, but does impose it for some higher payment. Then, whenever then the borrower has at least L, he will never pay more than L. In the case of strong legal protection, the borrower would avoid the penalty by paying L, and keep the remainder for himself. The threat to impose the sanction given higher payments, is meaningless, because the payment will never exceed L.

If the borrower has sufficient cash and if legal protection is strong, he will pay the lowest amount that avoids the sanction, and he will be able to keep the rest, without needing to divert it. The only defaults occur when the borrower has insufficient funds to pay that amount. This implies a description of the optimal financial contract without monitoring: select a payment, F, that, if paid, the sanction is avoided. The lender commits to impose the sanction for all lower payments. This implies that the optimal contract when monitoring is impossible is a debt contract with face F. The face value includes the promised payment of principal and interest. On theories of debt and sanctions see Diamond [1984], Townsend [1979], Gale and Hellwig [1985] and Dubey, Geanakoplos and Shubik [2005].

3.2.1 Determination of the face value of non-monitored debt

This section determines the minimum face value, F, of non-monitored debt which

will induce the borrower to make debt payments with an expected value of I (this is the required expected repayment to induce the risk neutral investors to fund the project).

Suppose F≤L<I. When the business cash flow C=L, the borrower pays F (paying less would result in sanction and give a zero payoff) and gets L-F≥0. When C=H, the borrower pays F (to avoid sanction), and keeps H-F for himself. This implies that with face value of F, the lender gets F for sure, which is less than I and not acceptable.

If instead L>I, then riskless debt with face value of F=I can allow the firm to finance itself and avoid diversion at no cost. I assume L<I for the balance of this chapter.

Suppose that F is between L and H. Any face value of debt between L and H forces the borrower to default when the project returns L, but to pay in full when the project returns H (because paying F returns H-F>0). This gives the lender an expected return of PF, because the lender recovers nothing when the sanction is imposed. Solving for the face value of debt (between L and H) that gives lenders an expected repayment of I (PF=I) yields F=I/P. As long as H>I/P, the borrower can get a positive return, H-I/P, by borrowing with non-monitored debt. However there will be costs of financial distress with probability 1-P, and the expected cost of distress is (1-P)L. For the example of I=100, H=160, L=90, and P=2/3, the face value of unmonitored debt is F=150, and the expected cost of financial distress is (1-P)L=30.

3.3 The Value of Undelegated Monitoring

This section explains the value of undelegated monitoring, continuing with the assumptions and model from section 3.2. Here and in this entire chapter, I assume that lenders possess all of the bargaining power and capture all of the efficiency gains in surplus from any renegotiation of contracts. In practice, this means that contracts will be renegotiated only if it helps the lenders. Lenders will make a take it or leave it offer to

the borrower in any renegotiation. If the contract specifies a penalty contingent on a payment, the lender can commit to deliver the penalty. Chapter 2 examines less extreme lender bargaining power, but the extreme assumption in this chapter both simplifies the explanation and delivers the most important results. In this section, I will use "liquidation" to mean imposing the legal sanction.

Monitoring of the realized cash flow, and using the information to liquidate only when there is actually diversion, is potentially better than incurring costs of imposing costly liquidation as a function of actual payments to lenders. Monitoring may possibly deter diversion at a lower cost.

Suppose that the lender monitors the value of the borrower's operations and observes the actual cash flow that accrues to the borrower. Then, instead of liquidating when less than the face value of debt is paid, the lender can use the sanction threat and offer to refrain from liquidation as long as the borrower repays as much as possible. The lender can commit to impose the sanctions if the borrower pays less. Instead of always writing a contract that leads to liquidation when less than F is paid, the lender can offer to accept L when C=L, but continue to require a payment of F when it when C=H. This policy leads the borrower to pay F when C=H and L when C=L. The lender has all of the bargaining power and will offer to accept less than F only when C=L.

The gross value of monitoring (ignoring all costs) to risk neutral agents is the expected savings of financial distress costs from imposing sanctions, which is equal to (1-P)L. This is the savings from monitoring, S, described in section 2. This benefit must be compared with the cost of monitoring. The cost of monitoring the cash flows of the borrower's project is K. If there were a single lender, then monitoring would cost K. Duplicated monitoring by each of n lenders would cost nK and would be equivalent to a

single lender facing a monitoring cost of nK. I assume that the cost of monitoring is incurred before a loan is repaid. This implies that the lender must learn in advance about the borrower's business to properly interpret any ex-post data about the project's return. In this case, the monitor must establish a costly relationship with the borrower. When I discuss weak legal protection in section 5 of this chapter, more general types of ex-ante monitoring are considered. In chapter 2, I discuss relationship lending in more detail.

When there are many lenders per borrower, undelegated (direct) monitoring becomes very costly. A reduction of these costs may be achieved by delegating the monitoring to a single monitor. This is examined in the next section.

Section 4: Delegated Monitoring and Financial Intermediation with Strong Legal Protection

4.1 A Role for an intermediary.

If all lenders are wealthy enough and willing to lend I to each borrower, then there is one loan per borrower (n=1) and there will be no duplication of monitoring effort. However, if there are not enough large lenders to satisfy borrowers' demands for financing, borrowers must borrow from n>1 small lenders. If the small investors cannot delegate monitoring and n is large, monitoring costs, nK per loan, are prohibitive. Diversified financial intermediaries can serve as delegated monitors and act like "synthetic" large investors.

Suppose that there are only small investors each with I/n to lend, and n small lenders are needed to finance I. If the cost of monitoring is K>I/n for each, then its cost would exceed I, which is prohibitive, and no one would monitor. When monitoring costs are prohibitive, the optimal contract is widely-held debt with face value I/(1-P) (see section 3.2.1). Delegating monitoring to one agent avoids duplication of effort, but

causes incentive problems for the agent delegated the monitoring task. Small lenders will not observe the information monitored by the agent, and they may not even observe that any effort was put into monitoring. The agent (called "the banker") has a conflict of interest with the small lenders. The conflict is similar to the conflict of interest between the borrower and the small lenders. How can the monitoring task be delegated without the need for each lender to monitor the monitor at a prohibitive cost? The solution is for the banker to face sanctions as a function of the amount paid to the n small lenders (depositors). This gives the banker incentives in the same way it does the borrower: when legal protection is strong, the banker is always better off paying a sufficient amount to avoid the sanction. If the banker writes a debt (deposit) contract with face value B and faces sanctions (liquidation) whenever he pays less than B to depositors, he will choose to make the payment B whenever it is feasible, and he will never pay more than B to depositors.

Because there is strong legal protection, imposing the sanction eliminates any benefit to the banker (monitor) from diverting or agreeing to share diversion proceeds with the borrower; the banker gets a zero payoff if the sanction is imposed. As assumed before, there is no cash recovery when the sanction is imposed (the depositors get nothing). There are several ways to interpret this high cost of actually imposing the sanction. One interpretation is that when too little is paid to the depositors, the bank's assets (loans) are liquidated, consuming all of the bank's assets. Another interpretation, a bit outside the model, is that because the banker gets zero when he defaults on deposits, the banker eliminates any discretionary component of monitoring if he anticipates bank failure. The reduced monitoring will decrease the value of bank assets. The assumption

that borrowers and lenders get zero when the bank fails serves as a simple shorthand for these more complicated aspects of bank failure costs.

4.2 Delegated Monitoring Without Diversification Does Not Succeed

The face value of bank debt, B, is the largest amount the banker ever chooses to pay depositors. Paying B avoids the sanctions from defaulting on deposits. Whenever the banker cannot pay B, the sanctions are imposed and depositors get nothing. Collectively, depositors require an expected repayment of I, implying that $B \geq I$.

Suppose that the banker monitors a single loan (manages a one-loan bank) on behalf of the small lenders, implying there is no diversification across loans. When the borrower's project returns $L<I$, the banker can monitor, threaten to liquidate, and collect the L without actually liquidating. However, the bank itself fails and is liquidated, because the face value of the bank's debt, B, is at least I ($B \geq I > L$).[2] As a result, the bank is liquidated whenever the borrower would have been liquidated, had the borrower used widely-held (and thus unmonitored) debt. Unless the n lenders each monitor the banker at a prohibitive total cost exceeding I, the one-loan bank will default and be liquidated just as often as the borrower. This one-loan bank example seems to imply that delegating the loan monitoring to the banker does not succeed.

4.3 Can the banker use diversification to reduce delegation costs?

Suppose the banker monitors a diversified portfolio of loans. A very simple way to show the value of diversification is to examine the two-loan bank. In particular, suppose the banker monitors the loans of two borrowers, whose returns are independently

[2] In the text I ignore the I/n of capital that the banker can contribute, to simplify the explanation. One can slightly lower the face value of debt issued to small outside lenders, but the complication is not very informative. The banker has capital of his or her own to invest. The bank need not raise I, but only I(1 - 1/n). This is equivalent to the case where the banker has none of his own capital but outside investors require a -1/n expected return per unit of investment. The one-loan bank is not viable even when only a -1/n return must be given to outside depositors.

distributed but are otherwise just like that of the single borrower (each loan has a P probability of returning H and a 1-P probability of returning L). The banker attracts 2I in deposits from 2n investors and lends it out to two different borrowers.[3] The banker gives each borrower a debt contract with face F and collects F when the borrower has H and monitors and uses the threat of legal sanctions to collect L when the borrower has L. As a result, the banker does not need to use costly liquidation to enforce his loan contract with either borrower. The banker issues non-monitored debt deposits that are widely-held, and the bank is liquidated whenever it pays less than face value of its deposits. This requires no monitoring by the 2n small investors. Let B denote the face value of bank deposits per loan, implying that the two-loan bank has total deposits of 2B, and each deposit of $\frac{I}{n}$ has face value $\frac{1}{n}$B.

Suppose the banker monitors both loans. If both borrowers pay in full, the bank will receive 2F. If one defaults but not the other, the bank will receive L+F. If both default, the bank will receive L from each, or 2L. The distribution of payments to the bank, if the banker monitors, is:

Payment	**Probability**		**Explanation**
2F	P^2	[4/9]	both pay F
F+L	2(P)(1-P)	[4/9]	one pays F, one L
2L	$(1-P)^2$	[1/9]	both pay L

The example in the square brackets assumes P=2/3.

The total face value of bank deposits is 2B. If the bank must fail (be liquidated) when it collects face value of F from one borrower and L from the other, it will be liquidated whenever one loan defaults, and there will be no possible savings in costs of

[3] This assumes that the cost of monitoring is a labor cost of the monitor. If monitoring consumes inputs, then the bank would need to raise an additional K per loan, and offer slightly higher promised repayments.

financial distress compared to the borrowers issuing unmonitored debt directly. Alternatively, if the bank can and will pay its deposits when one loan defaults, it fails only when both loans default, and can reduce the probability of liquidation to $(1\text{-}P)^2$.

Continuing the example from section 3.2 with P=2/3, I=100, H=160 and L=90, we see that when payment of all deposits is possible when just one loan defaults, the total payment received by all depositors will be 2B with probability 8/9 and 0 with probability 1/9. The expected payment is 8/9(2)B. To raise the initial capital needed to make two loans requires an expected repayment of 2I=200, implying that 8/9(2)B=200, or 2B=225 which is the face value of deposits. Equivalently, let the promised interest rate on bank deposits be r_B, then because $2B=225=2I(1+r_B)=200(1+r_B)$, the promised interest rate on the bank deposit is $r_B = 12.5\%$.

In summary, if the bank fails with probability $(1\text{-}P)^2$, the constraints on the face values F and B are as follows:

$F + L \geq 2B$ (do not fail when exactly one loan defaults and pays L),

$B \geq \frac{I}{1-(1-P)^2}$ (provide a market expected return, when the bank fails only when both loans default).

Combining these two constraints:

$F \geq \frac{2I}{1-(1-P)^2} - L$ (avoid failure with one default and provide a market return).

In addition, $F \leq H$ is required to allow the payment to be feasible when the borrower obtains H. Finally, the monitor must be willing to take the job and receive an expected return of at least K per loan to cover the monitoring costs. All of the monitor's return comes from keeping the residual claim after repaying depositors. If the monitoring is not observable, the monitor might shirk, and then there is another incentive constraint, described in the next section.

Returning to the numerical example, if the bank is to be able to pay 225 when one loan defaults (paying L=90) and the other does not default (paying F), then L+F must be at least 225, and the face value of each loan must satisfy F≥135. If the bank made loans with this face value, it could avoid liquidation with probability 8/9. In summary, if the bank monitors its loans, it will have the cash and the incentives to pay bank deposits in full with probability 8/9 as long as F≥135, or the interest rate on bank loans is at least 35%. The banker keeps the residual cash after paying depositors. This has value only when neither loan defaults (which has a probability of 4/9). The expected value of the banker's cash is $\frac{4}{9}(2F-2B)=\frac{4}{9}(270-225)$ =20, or 10 per loan. This exceeds the per loan cost of monitoring, K=2, because the constraint that the bank not fail when just one loan defaults is binding. The banker earns a small rent of 8 per loan. This rent, plus the expected cost of bank failure per loan, or $(1\text{-}P)^2L$= 1/9 (90)= 10, add up to a total delegation cost of D= 8+10 = 18 .

We consider three types of contracting arrangements:

1. No monitoring: a directly issued debt contract with face value=I/P=150 for each borrower (and distress costs of S=L(1-P)=30),

2. Direct monitoring by investors, which avoids distress costs of S=L(1-P)=30 but costs nK=2n,

3. Delegated monitoring of loans with face value=135 by an intermediary, which avoids distress costs of S=L(1-P)=30 but has monitoring costs plus delegation costs, K+D=2 + 18 = 20 (in the two loan case).

For delegated monitoring to dominate direct monitoring, the monitoring costs plus delegation costs, K+D, must be less than or equal to the cost of direct monitoring, nK.

For the numerical example, this is $20 \leq 2n$, implying, n, the number of depositors per loan is greater than or equal to 10.

Diversification within the intermediary makes delegated monitoring the best option because it reduces the delegation costs of providing incentives for the bank to monitor loans and pay depositors. The previous example shows that diversification from a bank making only two loans was sufficient to give the bank reduced delegation costs. However, the result is generally true for financial intermediaries with a very large number of loans. This is shown in section 4.5.

4.4 Implications of Strong Legal Protection.

When legal protection is strong, borrowers with risk free projects or projects that can issue risk free debt which pays lenders a return of I (these are projects with a cash flow of at least I in all circumstances) can be financed directly with no need for monitoring. Risky borrowers where cash flow can be less than I can finance with unmonitored debt only if they incur a positive probability of costly financial distress whenever they cannot pay as scheduled. Monitoring and renegotiation can reduce this cost, but if monitoring must be delegated, reasonably well diversified institutions (banks) that issue unmonitored debt (deposits) will be required. Poorly diversified banks will not survive, and regulations that limit bank diversification can make banks very unstable.

With strong legal protection, sufficient sanctions are available to deter diversion. The problem solved by monitoring is the appropriate state contingent application of the (costly) sanctions. The same strong legal sanctions that can deter diversion by the borrowers will serve to deter diversion by bankers. Banks will serve as safe places to invest as long as they are sufficiently well diversified.

4.5 Unobservable monitoring expenditure

Suppose that the bank's expenditure on monitoring (or relationship building) is not observable and that the bank can choose not to monitor, saving its cost. I follow the analysis of Diamond [1996] here (see also related analysis in Holmstrom and Tirole [1997]). Without monitoring, the bank would not know the borrower's realized cash flow. Thus, it would not be able to offer to take less than face value F only when the borrower has cash of L. It would instead leave in place the commitment to liquidate (impose the legal sanction, with a zero recovery to the bank) when less than F is paid. Monitoring provides no benefit to the banker when all loans pay in full (monitoring is not needed to force a borrower to pay F) nor when all loans default (because the bank fails then and is liquidated). The entire increase in the bank's return comes from increasing the return when just one loan defaults. If loan and deposit interest rates are set such that all loan collection proceeds are used to pay deposits when only one loan defaults, none of the benefit of monitoring accrues to the banker and there will be no incentive to monitor.

If the bank must be given an incentive to monitor, the expected value of the banker's residual claim must increase at least by the cost of monitoring the two loans, 2K. Diamond [1996] shows that providing an incentive to monitor adds another constraint on F, the face value of bank loans: $2P(1-P)(F+L-2B) \geq 2K$.

The constraint on the face value of deposits per loan, B, is unchanged from section 4.3 and remains $B \geq \frac{I}{1-(1-P)^2}$. Substituting this into the previous constraint, which implies that the banker has an incentive to monitor, yields:

$$F \geq \frac{K}{P(1-P)} + \frac{2I}{1-(1-P)^2} - L \quad \text{(Incentive to monitor)} .$$

The term $\frac{K}{P(1-P)}$ in the above constraint represents the additional payment required to give the bank an incentive to monitor. The constraint can be compared to that in section 4.3, which did not include an incentive to monitor.

Returning to the numerical example, the constraint is F≥144. This adds an additional $\frac{K}{P(1-P)} = \frac{2}{2/9} = 9$ to the required face value F obtained in section 4.3. It gives the banker an additional expected rent of $P\frac{K}{P(1-P)} = \frac{K}{(1-P)} = \frac{2}{1/3} = 6$ per loan. The total expected rent to the banker becomes 6+8=14.

An incentive to monitor requires that monitoring increase the bank's expected payment by at least K per loan. As long as F≥144 (the interest rate on bank loans exceeds 44%), the banker is willing to invest 2K=4 to monitor both loans because it increases the value of his residual claim on the bank by this amount.[4]

The need to provide incentives puts a floor on the banker's expected profit, and gives a rent to the banker. If further diversification is not possible, either because there are just two loans or because a two-eyed banker can only monitor two loans, bank profits cannot be driven to zero by competition. The two-loan bank has the following profits. The banker gets the residual claim above 2B=225, or:

2F – 2B = 288 - 225 = 63 with probability 4/9, when neither loan defaults,

F+L -2B= 234 - 225= 9 with probability 4/9, when one loan defaults, and

0 with probability 1/9, when both loans default.

[4] The banker could monitor just one loan, but will not prefer this. The condition to prefer to monitor one loan (versus none), P(1-P)(F+L-2B)≥K, automatically implies an incentive to monitor both loans.

This works out to a total expected payment of 28, or (63)4/9 + (9)4/9=32. This is a return to the banker of 16 per loan, which is in excess of 2, the cost per borrower of monitoring, and the banker earns a rent of 16 - 2 = 14.

The delegation cost per borrower, D, equals the cost of bank failure of 1/9(90)=10, plus rent to the banker of 14 for a total of 24. All parties are better off with the banker as delegated monitor. The borrower prefers to borrow at 44% (F=144) from the bank, versus at 50% (F=150) by borrowing directly without monitoring. The investors get an expected repayment of I=100 in either situation. The banker is happy with any claim with an expected payment above 4 and, in this case, ends up with an expected payment of 28.

The rent to the banker is due to uncertainty about the amount that the bank's loans will be able to repay. The law of large numbers implies that if the bank gets sufficiently diversified across independent loans with expected repayments in excess of the face value of bank deposits, then the chance that it defaults on its deposits gets arbitrarily close to zero. In the limit of a perfectly diversified bank, the bank would never default and would face no costs of bank failure.[5] In addition, the rent needed to provide incentives to monitor approaches zero. The delegation cost for the bank approaches zero, and the only cost of intermediation is the (unavoidable) cost of monitoring. Competitive and fully-diversified intermediation would drive borrowers' expected cost of borrowing down to the cost of capital, I=100, plus the cost of monitoring, or K=2. In the limit of perfect diversification, the face value of bank loans approaches F= 108 which is the solution to PF + (1-P)L=I+K, or 2/3F + 1/3(90)= 100 + 2; it exactly covers the bank's cost of monitoring (K=2). This is a bit too strong because in practice the number of loans in the

[5] For a formal limiting argument about well-diversified intermediaries, see Diamond [1984], and for a generalization see Kraska-Viramil [1992].

bank's portfolio is limited, and it is likely that the default risk of borrowers is not independent, but is positively correlated.

The general message is that diversification allows banks to transform monitored bank loans into deposits that do not need monitoring, delegating the monitoring to bankers. These contractual forms minimize the sum of monitoring and delegation costs.

Section 5. Weak Legal Protection and Small Sanctions

If the legal system provides sanctions which are too weak to deter some borrowers from diverting funds, then more detailed ex-ante monitoring of actions and cash flows may be needed. Responding to an actual default, which occurs at the end of a period, is too late. Early monitoring is needed in order to quickly intervene to reduce the borrower's payoff from diversion, stopping a crime in progress, as in Calomiris-Kahn [1991] and Diamond [1991]. The value of early monitoring can be due to knowledge of the location of diverted funds, the ability to expose secret side deals or just the ability to impose costs on the borrower if and only if he is diverting funds. This ability to reduce the spoils of a crime in progress gives the monitor the ability to extract a larger cash payment from the borrower. Section 5.1.1 looks at the case where monitoring is not delegated and the monitor is the investor. This could be thought of as lending by a wealthy family. Section 5.1.2 examines delegated monitoring with weak legal protection. The weak legal sanctions that lead to the need for ex-ante monitoring can complicate the delegation of monitoring. For simplicity, I suppress the cost of monitoring from the analysis of all of the cases with weak legal protection.

5.1 A Model of monitoring with weak legal protection.

The borrower cash flows and the time line of borrower and monitor actions are the same for undelegated and delegated monitoring.

The model of monitoring in this section returns to the case where the borrower's cash flow is H for sure. Recall from section 3.1, that legal sanctions can reduce the borrower's payoff from diversion by ϕH. Also recall that if the borrower diverts the cash flow H, he can only obtain proceeds H(1-t), where t is the fraction of cash destroyed by covering one's tracks. Therefore, imposing legal sanctions on a diverting borrower reduces the proceeds available to the borrower to $H(1-t-\phi)$. In this section, I examine the case of weak legal protection, implying $\phi<1-t$. In monitoring the borrower, if the monitor observes the act of diversion sufficiently early, he can intervene and stop the crime in progress (the details of timing are discussed below). This reduces the borrower's diversion proceeds by Hm. Monitoring is useful for eliminating or reducing the borrower's spoils of diversion. This ability to reduce the borrower's diversion proceeds gives clout to the monitor. If m=1-t, monitoring eliminates all of the borrower's spoils. If m<1-t, the spoils are reduced and the additional effects of the legal sanctions for default, ϕH, are relevant. This implies that the cash available to the borrower is $H(1-t-m-\phi)$ if the borrower diverts, the monitor stops the crime in progress, and legal sanctions are incurred.

The time line of borrow and monitor actions is as follows. First, the borrower chooses whether to divert cash. If the borrower does not divert cash, it is available to the lender, and the borrower will use it to make the promised loan payment, $F \leq H$. If the borrower diverts the cash, there is not a verifiable default or a full payment of F at this stage. However, the monitor observes the borrower's action. If the borrower is in the process of diverting cash, the monitor can commit to stop the crime in progress (reducing the diversion proceeds by Hm) unless the borrower makes a payment specified by the monitor. Then the borrower accepts or rejects the monitor's offer. For simplicity, until

chapter 2, I continue to assume that the monitor has all the bargaining power and can make this brief commitment to stop the crime in progress if the borrower rejects his offer.[6] Stopping the crime will reduce the diversion proceeds to H(1-t-m); the borrower defaults and is still subject to the legal penalty, ϕH. The borrower's payoff if the monitor actually stops the crime in progress is H(1-t-m-ϕ). The borrower's outside option, his payoff if he rejects the offer, is at least this amount.

There is a subtlety in the time line regarding the timing of the borrower's act of diversion relative to when the monitor must appear in order to be effective. The primary focus is on the case where the monitor appears before the crime is actually completed. In this case, the diversion is about to occur, but the borrower can rethink his decision to divert after the monitor threatens to stop the crime in progress. This is referred to as the reversible case. The significance of this case is that the borrower has the option of paying F, the face value of the loan, rather than suffer the threats and consequences imposed by the monitor. In this case, the borrower's outside option is the larger of H-F and H(1–t–m–ϕ), where H-F is the payoff from reversing diversion and H(1-t-m–ϕ) represents the payoff if the crime is actually stopped. This reversible case is more realistic than the irreversible case, because it allows negotiation between the banker and the borrower, letting them decide how to spilt the spoils. It includes the situation of irreversible diversion where the bank and the borrower negotiate a deal before the diversion has occurred.

The less interesting case is when the monitor appears just after the crime (diversion) has occurred. After diverting funds, the borrower has cash of only H(1-t),

[6] Due to this ability to commit and obtain all surplus, the amount of the payoff to the monitor from actually stopping the crime matters very little (only in off the equilibrium path payoffs). I could instead assume that stopping the crime in progress gives a payment of mH to the monitor. This would remove the need for short-term commitment to the stop the crime if the offer was rejected.

thus his outside option to fully repay gives him a payoff of $H(1-t)-F$. If he does not fully repay, the borrower remains subject to the threats and consequences imposed by the monitor. This is referred to as the irreversible case. The borrower's outside option is the larger of $H(1-t)-F$ and $H(1-t-m-\phi)$.

The undelegated monitoring section and the delegated monitoring section analyze both the reversible and irreversible cases. It turns out that they are identical for undelegated monitoring, and the reversible case is more compelling for delegated monitoring.

5.1.1 Undelegated monitoring

Proposition 1 showed that an unmonitored borrower with cash flow H diverts funds and defaults on his debt whenever the face value of F of his debt exceeds $H(t+\phi)$. If $H(t+\phi)$ is less than I, the borrower cannot finance his project. Now suppose that a lender is ex-ante monitoring the actions of a borrower whose loan he alone financed directly. With this undelegated monitoring, I will show that the borrower diverts funds and defaults on his debt only when F exceeds $H(t+m+\phi)$. In other words, because monitoring reduces the proceeds from diversion by mH, borrowers are willing to pay this larger face value to lenders. Undelegated monitoring increases the amount of financing available to borrowers by Hm.

For the reversible case, if the monitor observes diversion by the borrower, he will demand a payment that drives the borrower's payoff down to the borrower's outside option. The outside option is the larger of H-F and $H(1-t-m-\phi)$, as shown above. Anticipating this payoff if he attempts to divert versus a payoff of H-F if he does not attempt to divert, he will not attempt to divert if $H\text{-}F \geq H(1-t-m-\phi)$, or $F \leq H(t+m+\phi)$.

For the irreversible case, the only difference is that if the monitor observes diversion by the borrower, the borrower's outside option is the larger of H(1-t)-F and H(1−t−m−ϕ), as shown above. Anticipating this payoff if he attempts to divert versus a payoff of H-F if he does not attempt to divert, he will not attempt to divert if H-F ≥ H(1−t−m−ϕ), or F≤ H(t+m+ϕ). This is identical to the reversible case.

In the case of reversible diversion, if the monitor observes attempted diversion, he demands a payment of the smaller of F and H(t+m+ϕ), forcing the borrower to reverse the diversion. The monitor could instead require a borrower who has diverted funds not to reverse the diversion and pay the monitor a share of diversion proceeds H(1-t). However, the monitor is never better off doing so when he is the only lender (he owns the loan). It will turn out to be relevant in the case of delegated monitoring.

Figure 1 shows the outside options in stage 2 of the figure and the equilibrium payoffs of the borrowers and the monitoring lender in stage 3, given that the borrower attempts to divert at stage 1. It accounts for the non-negativity constraints glossed over in the text.

Figure 1: Outside Options and Payoffs if the Borrower Diverts

Stage 1	Borrower diverts	
Stage 2	Monitor can threaten to stop a crime unless a specified payment is made	**Borrower's outside option: If diversion is irreversible:** $\max\{H(1-t)-F,((1-t-m-\phi)H)\}$ for $m< 1-t-\phi$ $=\max\{H(1-t)-F,0\}$ for $m=1-t-\phi$ **Borrower's outside option: If diversion is reversible:** The larger of that with irreversible diversion and H-F
Stage 3	Payoffs when all surplus over the outside option goes to the monitoring lender	

$(B,L)=(\max\{0,(1-t-m-\phi)H\}), \min\{H(t+m+\phi),H(1-t)\})$
if diversion is irreversible

$(B,L)=\max(\{0,(1-t-m-\phi)H, H-F\},), \min\{H(t+m+\phi), H, F\})$
if diversion is reversible

The notation (B,L) denotes the payoffs of the borrower, B, followed by that of the monitoring lender, L.

5.1.2 Delegated Monitoring with Weak Protection.

In a weak legal environment, borrowers are more likely to divert funds, making monitoring more important. However, delegated monitoring is less effective than in a strong legal system. This is because the borrower and monitor may find collusion attractive.

The framework for this analysis is similar to the case of delegated monitoring in a strong legal environment. The information monitored is unobservable and unverifiable by other lenders, and the monitor needs to raise all funding from many small investors. The borrower and monitor can share the diversion proceeds if they wish. Outside investors do not monitor the monitor, so they will not be able to stop this joint crime

while in progress. Outside investors provide all capital and require a repayment of I. Investors can write contracts which impose legal sanctions on the borrower and the monitor as a function of observable cash payments. The key difference is that legal sanctions may be insufficient to deter diversion and/or collusion. As seen in the case of undelegated monitoring with weak legal protection, ex-ante monitoring to allow the stopping a crime in progress is needed to deter borrower diversion. Also, in a strong legal environment, delegated monitoring relied on strong sanctions to deter collusion. Weak legal sanctions are less able to deter collusion.

The borrower and the monitor (but no others) observe any act of diversion by the borrower. If the borrower does not divert funds, he can make a verifiable payment, $V \leq H$, to the monitor. If a verifiable payment is made, the monitor cannot divert it. The monitor can make a verifiable payment to lenders, Z, up to the amount, V, which is paid by the borrower. Investors can write contracts where the imposition of a penalty of up to ϕH on the borrower and $\phi_M H$ on the monitor, is contingent on the amount of the payments Z and V.[7] If there is a fixed maximum sanction the legal system imposes, then these two sanctions are equal ($\phi=\phi_M$), but in general they could be different. As before, let F be the face value of a debt contract (loan) owed by the borrower to the monitor (bank). Payments of V less than F trigger a legal penalty ϕH and give the monitor the right to stop a crime in progress. Let B be the face value of a debt contract (bank deposits) owed by the delegated monitor (bank) to the outside investors, where a legal penalty of $\phi_M H$ is imposed on the monitor for payments of Z smaller than B. The bank must collect at least B from the borrower to avoid bank failure (default on deposits).

[7] If we changed the contract slightly to require the monitor to force a payment, V=B, to be made directly by borrower to outside investors, we could allow the legal sanctions to depend only on V.

If the borrower diverts funds, the borrower can make either unverifiable side payments, denoted by U, to the monitor up to the diversion proceeds, H(1-t), or verifiable loan payments, V, up to H if diversion is reversible and up to H(1-t) if it is irreversible.

If the monitor observes diversion, he can threaten to stop it and reduce the proceeds from diversion by mH unless the borrower makes a specified payment to the monitor. The monitor sets the specified payment to make the borrower indifferent between making the payment and letting the monitor actually stop the crime in progress. The monitor can specify either a verifiable payment, V, or an unverifiable payment, U, from the diversion proceeds. All of the unverifiable payment U will accrue to the delegated monitor. If the monitor receives a verifiable payment V from the borrower that satisfies $V \geq F \geq B$, neither is penalized and the monitor's payoff is V-B and the borrower's is H-V. If the monitor receives a side payment U from the diversion proceeds, the delegated monitor's payoff is $U-\phi_M H$ and the borrower's is $H(1-t-\phi)-U$. Legal sanctions, ϕH and $\phi_M H$, are incurred because the borrower and monitor default when they share the proceeds of diversion.

The amount of financing available to the borrower is limited by the largest amount that he will pay investors instead of diverting and defaulting. The limit is the largest incentive-compatible value of F. With no monitoring, $F \leq H(t+\phi)$, as shown in Proposition 1. Section 5.1.1 shows that with undelegated monitoring, $F \leq H(t+m+\phi)$, an increase of Hm. Hm is the amount by which a monitor can reduce the borrower's diversion proceeds. Hm represents the clout the monitor has over the borrower. In these two cases, since there is no intermediary (bank) involved, the payment F is made directly to investors. For delegated monitoring, the payment to investors is what depositors receive from their deposits with face value B. Proposition 2 characterizes the largest

values of B that the bank will pay without default. It shows that for reversible diversion by the borrower, this maximum value exceeds the amount an unmonitored borrower is willing to pay by the smaller of Hm and $H\phi_M$. The bank's clout, Hm, can be used for good or evil purposes. The bank can force the borrower to make a larger verifiable loan payment (good) or force the borrower to share diversion proceeds (evil). If the legal sanctions, $\phi_M H$, on the bank are too low, the bank will collude and use its clout for evil purposes. If diversion is irreversible, this largest value of B exceeds the maximum payment of an unmonitored borrower by Hm, just like for undelegated monitoring.

Proposition 2 A delegated monitor can reduce a borrower's diversion proceeds by Hm. If diversion is reversible, the delegated monitor will collude with the borrower, allow diversion and default on deposits with face value B if $B \geq \min\{ H(t+\phi+\phi_M), H(t+m+\phi)\}$. If diversion is irreversible, the delegated monitor will collude, allow diversion and default on deposits with face value B if $B \geq H(t+m+\phi)$.

The case of reversible diversion is most relevant because it does not artificially restrict the borrower and the bank from reaching a mutually beneficial deal. The borrower can negotiate a deal with the bank when he can still use all of his cash flow to avoid a default. This implies that if the bank and borrower negotiate a deal to jointly divert, only some of the spoils will go to the bank. In this case, Proposition 2 implies that with very weak legal protection, $\phi_M=0$, delegated monitoring does not work. If diversion is irreversible, the borrower will not divert in the first place because all of the spoils would go to the banker. Therefore, the borrower repays the bank loan with a verifiable payment, V, which accrues to depositors. In this case even when $\phi_M=0$, delegated

monitoring works as well as undelegated monitoring, but only because the borrower and the monitor cannot reach a deal to their joint benefit.

Proof of Proposition 2:

For the cases of reversible and irreversible diversion, the borrower's payoff from not diverting at all is H-F and the monitor's payoff is F-B. If diversion is irreversible, the borrower can pay at most H(1-t) once he diverts. If F>H(1-t), and the monitor commits to stop the crime if his offer is rejected, the borrower's outside option is H(1-t-m-ϕ) because he receives a legal penalty. The borrow will not divert if $F \leq H(t+m+\phi)$. If $t+m+\phi \geq 1$, irreversible diversion is always unattractive to the borrower, who will not attempt to divert, so long as it is possible to fully repay ($F \leq H$).

Similarly, if diversion is irreversible and $F \leq H(1-t)$, borrower's outside option after diversion includes the right to fully repay F after initially diverting funds and is $\max\{H(1-t)-F, H(1-t-m-\phi), 0\}$. Only in the case where the larger expression is H(1-t)-F is this different from the previous paragraph. In this case, the borrower's payoff from diversion is H(1-t)-F . However, the borrower will prefer not to divert in the first place because it yields the payoff $H-F>H(1-t)-F$.

If diversion can be reversed after it is detected, the borrower's outside option given diversion is at least H-F, because the borrower can still fully repay F and remove the monitor's right to stop the crime. The borrower can find it advantageous to negotiate with the delegated monitor to share diversion proceeds. The borrower and monitor will choose a diversion decision that maximizes their total joint payoff. If the borrower diverts, the delegated monitor can propose a verifiable payment of up to F or a non-verifiable payment U (which will accrue to the monitor) from the diversion proceeds of H(1-t). The borrower's outside option is $\max\{H-F, H(1-t-m-\phi)\}$, because if cash remains

diverted, the borrower will receive the legal sanction ϕH. If the outside option is the second term, $H(1-t-m-\phi)$, then $F > H(t+m+\phi)$, and default and diversion are unavoidable because the monitor cannot force the borrower to make a verifiable payment in excess of $H(t+m+\phi)$. If instead the outside option is H-F ($F \leq H(t+m+\phi)$), the borrower's payoff from shared diversion is $H(1-t-\phi)-P$, and he will reject any offer of $U>F-H(t+\phi)$. The monitor's payoff from receiving a diversion share $U=F-H(t+\phi)$ (which leaves the borrower's payoff equal to his outside option) is $U-\phi_M H = F-H(t+\phi+\phi_M)$. The monitor's payoff from a verifiable payment of F is F-B. The monitor prefers a verifiable payment of F if $F-B \geq F-H(t+\phi+\phi_M)$, or $B \leq (t+\phi+\phi_M)H$. Instead, if B exceeds $H(t+\phi+\phi_M)$, the monitor shares in the diversion proceeds and delegating monitoring does not succeed. Q.E.D.

Bank contracts (delegated monitoring contracts) impose consequences on the monitor for default on deposits, and these defaults are caused by defaults on the borrower's loans. This is similar to joint liability of several borrowers for each other's default. I explore this in the next section.

Section 6: The link to joint liability: Differential Payoffs from Diversion

Financial intermediary contracts subject both borrower and delegated monitor to sanctions when the borrower defaults. This is similar to a group penalty sometimes used in the military, where all soldiers are punished if any of them perform poorly. Even more similar is debt with joint liability, sometimes used in village economies or by the Grameen bank, where all borrowers are sanctioned if any default.

It is helpful to compare the results on delegated monitoring under weak legal protection in the previous section with a model of joint liability for two borrowers, neither a pure monitor, who are subject to joint liability. The joint liability is a penalty

of ϕH on each if either default on their individual obligations to each pay F to lenders. This is useful to understand the difference between diversified banks (delegated monitors) and conglomerates (joint liability of multiple operating divisions). It also illustrates advantages of separating banking from commerce, so that banks as delegated monitors do not have operating divisions with large potential for independent diversion. As a result, they are more likely to use their clout to extract verifiable payments instead of colluding to jointly divert cash.

One important way that joint liability is thought to work is by the diversification effect illustrated in the section on strong legal protection: it allows cross subsidy from successful to unsuccessful borrowers which avoids default and its costs, see Diamond (1984), Beasley and Coate (1995), Guinnane and Ghatak (1999), Prescott (1997) and Bond (2004). In the model of delegated monitoring in the previous section, there is no uncertainty, and thus no role for diversification. Despite this, joint liability of borrower and delegated monitor matters because if the borrower diverts cash and the bank fails, a sanction is imposed not just on the borrower but also on the delegated monitor, whose incentive to divert is less than the borrower's. Suppose that joint liability is imposed instead on two identical borrowers each with a project delivering a riskless cash flow of H. Each can use reversible diversion, and each can monitor the other (possessing a threat to reduce the other's diversion proceeds). In this case, joint liability would not provide increased incentives not to divert. Each alone could commit to pay $H(t+\phi)$ and together they can pay twice as much . If their total joint obligation, 2F, exceeds $2H(t+\phi)$, and each diverts, they will agree to share the proceeds of diversion, and jointly incur the penalty because diverting increases their shared total payoff ("total surplus") because $2F>2H(t+\phi)$.

However, if the two borrowers have very different incentives to divert funds, then joint liability and mutual monitoring can allow a borrower with a very low incentive to divert to convince the other not to divert, because diversion does not increase their total surplus. The increased incentive to pay without diversion is even stronger when diversion is irreversible (as it was for the delegated monitor in Proposition 2). Joint liability allows the incentive to divert to be eliminated when the party who would not divert unilaterally has sufficient bargaining power and monitoring is sufficiently effective.

Single liability is each borrower borrowing individually, with sanctions imposed only for one's own default. Joint liability is defined as subjecting each borrower to a penalty of ϕH unless each pays F. In addition, the borrowers are allowed but not required to pay off each other's obligations (this means that if a borrower has paid his loan of F he can always keep any remaining verifiable cash flow if he chooses). I assume that each borrower can monitor the other and can reduce the other's payoff from diversion by Hm. This is sufficiently large to deter diversion if desired (Hm is sufficiently large).

Proposition 3 Consider two borrowers, A and B, who each need to provide a return of I to investors to finance their projects. Their projects each produce a cash flow of H, but they have with different incentives to divert ($t_A<t_B$). Under single liability and without monitoring, only A would divert funds if F=I, that is $I\geq H(t_A+\phi)$ but $I\leq \min\{H(t_B+\phi),H\}$. Imposing joint liability on A and B if either pays less than F=I has the following effects.

i) If diversion is reversible, borrowers make diversion decisions which maximize their joint payoff. If $I=F<\min\{ H(t_A+2\phi),\ H(\frac{t_A+t_B}{2}+\phi)\}$, then neither divert;

if $H(\frac{t_A + t_B}{2} + \phi) \leq I \leq H(t_A+2\phi)$, then only A diverts (identical to single liability);

if $I=F>\max\{H(t_A+2\phi),\ H(\frac{t_A + t_B}{2} + \phi)\}$, then both divert.

ii) If diversion is irreversible and B has a sufficient fraction of the bargaining power, there is no diversion by either borrower.

iii) If diversion is irreversible and A has a sufficient fraction of the bargaining power, then A will divert but B will not (as with single liability).[8]

Proof: See appendix.

A financial intermediary has no diversion opportunities independent of those of its borrowers, which is equivalent to the intermediary being borrower B with $t_B=1$. A borrower with a low ability to divert unilaterally (a high value of t_B) is better able to commit to deter another borrower from diversion. If banking and commerce are not separated, or if we consider joint liability in conglomerates, then the value of t_B will low, and joint liability may not reduce diversion and may even increase it. In weak legal environments, separating banking from commerce increases the amount of funding that is available to borrowers by deterring diversion.

Section 7: Uncertainty, Diversification and Weak Legal Protection

This section examines the viability of well diversified financial intermediaries as delegated monitors when legal protection is weak. With strong legal protection, Section 4 shows that diversification is an essential part of delegated monitoring. When cash flows are uncertain, diversification is needed to make the cost of providing incentives for

[8] If a borrower were required to pay all of H (not just F) to lenders whenever the other borrower did not pay F, and not simply be subject to penalty under those circumstances, then joint liability can lead B to divert when he would not under single liability if A has sufficient bargaining power.

delegated monitoring less than the benefit of monitoring. With riskless cash flows, there is no need for monitoring because the strong sanctions triggered by a default deter borrowers from diversion. How do these results carry over to economies where legal protection is weak?

I introduce weak legal protection into a model that is otherwise identical to that in Section 4. The key features of that model are as follows. Each borrower has a project with an independent and identically distributed cash flow. Actually imposing legal sanctions provides a zero recovery to lenders, making them very costly to impose. The realized cash flow of a borrower is either C=H (with probability P) or C=L<H (with probability 1-P). Borrower cash flow is not observable and contingencies in contracts can depend only on the amount of cash actually paid by borrowers or banks. Borrowers can divert cash, making it unverifiable, but this destroys a fraction t of the diverted cash. Diversion is assumed to be reversible. Exactly like Section 4, the delegated monitor writes a deposit contract such that legal sanctions are imposed on the monitor if the bank pays less than the face value of deposits (B per loan). The monitor has the right to stop a borrower's crime in progress if the borrower pays less than F. The bank monitors each borrower's attempt to divert and also monitors borrower cash flow, C, to determine how much the borrower can afford to repay.

I assume that the monitor (bank) has the right to waive the sanctions on a borrower even if a borrower pays less than F, if and only if the monitor pays a total of B per loan to investors. The bank loses its right to waive sanctions for default if the bank fails. This realistic feature helps deter the bank from colluding with the borrowers to share the proceeds of diversion. It is important for the bank to be able to waive the legal

sanctions to avoid the costs of financial distress when borrowers have cash of only C=L to pay (because the face value of bank loans, F, must exceed L).

In Section 5, where it was certain that the borrower's cash flow was C=H, there was no loss of generality in describing the maximum total sanctions as a fraction of the cash flow. With uncertain cash flows, I allow the maximum total sanctions to be a general function of C. The maximum total sanction that can be imposed on a borrower is denoted as $\Phi_C \geq 0$. Under certainty in Section 5, where C=H, I assumed that the maximum total sanction was $\Phi_H = \phi H$. Note that making the sanctions a function of the actual cash flow does not mean the actual cash flow is observable or contractible. If the sanction is imposed on the basis of something that all can observe, the effect is to subtract a total of Φ_C from the borrower's payoff.

The maximum total legal sanction *per loan* that can be imposed on the bank depends on the realized cash flows of its borrowers (because the borrower can share cash with the monitor if the monitor allows diversion) and on N, the number of loans that the bank makes. The maximum *per-loan* sanction that can be imposed on a delegated monitor who monitors N borrowers all with cash flow C is $\Phi_{MCN} \geq 0$.[9] If a bank monitors N borrowers, and a fraction f have cash of H and 1-f have cash of L, I assume that the maximum per-loan sanction that can be imposed on the bank is $f\Phi_{MHN} + (1-f)\Phi_{MLN}$. This allows a fairly general specification for a fixed number of borrowers, N, per bank (do not confuse N with lower case n, the number of lenders needed to raise I and fund a borrower's project). I will examine the implications when the per-loan sanctions on the bank are either increasing or decreasing functions of N, the size of a bank.

[9] This implies a total sanction on the bank of $N\Phi_{MCN}$, but all of the analysis that I present for multiple loan banks is on a per-loan basis.

As in Section 4.5, diversification over sufficiently many independently distributed loans converges to a situation where a fraction P of the borrowers have a cash flow of H and 1-P have a cash flow of L. This can (essentially) eliminate the probability of bank failure, as long as the monitor has an incentive to collect verifiable loan payments from all borrowers (rather than colluding to allow diversion). If instead the well diversified bank colludes with borrowers, the well-diversified banks will not be viable.

Proposition 4 If the maximum legal sanction on banks (per loan), Φ_{MCN}, is a non-decreasing function of the number of loans monitored, N, then well-diversified banks will have a cost advantage and will be able to undercut less diversified banks. If the legal sanctions (per loan) are sufficiently decreasing in N, then an undiversified bank has a cost advantage and can undercut a well-diversified bank.

Proof: See appendix.

The idea behind Proposition 4 is as follows. Suppose that the maximum legal sanctions on banks (per loan) are a constant (do not depend on the number of loans per bank). Compare, for example, a viable one loan bank to a well diversified bank . A one loan bank fails with probability 1-P and thus must pay depositors I/P. To be viable, it must prefer repaying deposits to colluding with the borrower when he has cash of H; the one loan bank must set F greater than or equal to I/P plus the unverifiable side payment that the monitor could obtain, net of legal sanctions. Suppose that the well-diversified bank matches this loan face value, F>I/P, on all of its loans and pays a face value of I (per loan) on deposits (because it will never default). This automatically implies that the well diversified bank will not allow diversion, for the following reasons. Both the per-loan side payment for colluding with a borrower with given cash flow and the implied

per-loan legal sanction on the bank are the same for both banks. Suppose that the well diversified bank collects verifiable payments of F from the fraction P of borrowers with cash flow of H. This alone is more than enough to fully repay all of its deposits (PF>I). Even after repaying all deposits, the monitor retains at least as much as the (net) side payment that he could have collected from colluding with the borrowers with cash of H. In addition, the verifiable payments that it collects on its remaining loans (L from the fraction 1-P) accrue entirely to the monitor, because its other loan proceeds (PF) are sufficient to fully repay deposits. Thus, a well-diversified bank that matches the face value of the one loan bank strictly prefers not to divert, and this remains true even if it slightly undercuts the loan face value of the one loan bank. This logic extends to any partly diversified bank competing with a well diversified bank and to cases where maximum legal sanctions per loan are increasing in the size of the bank (implying that well diversified banks have an additional incentive not to collude with borrowers). If per-loan maximum legal sanctions imposed on banks shrink sufficiently for large banks, then a large well diversified bank's advantages can be swamped by the direct effect of their increased incentive to collude with borrowers.

If legal protection is strong, it eliminates the problem of healthy banks colluding with borrowers. Healthy banks can avoid failure by collecting verifiable payments on their loans. When legal protection is weak, the level of legal sanctions imposed on banks determines whether they serve to increase borrower's access to funding by deterring diversion or instead encourage collusion with borrowers, failing to provide discipline. If the per-loan legal sanctions on banks are much smaller for large banks, then the diversification benefits of a larger size cannot overcome the growing attractiveness of collusion.

Large banks face small per-loan sanctions if there is a maximum total legal sanction (not per loan) that can be imposed. This could be due to an actual upper limit on sanctions, such as limited liability, a maximum prison sentence, or a largest possible loss of reputation no matter how large the crime. Another example is corruption in the legal or regulatory systems, where it is cheaper to make a side payment to a judge or regulator rather than suffer a very large sanction. In either case, sufficiently large banks face decreasing sanctions per loan. If the legal system is corrupt in addition to weak, then encouraging small, less diversified banks provides larger increases in borrower's access to funding despite their increased risk of costly failure. These smaller banks do not have access to the pernicious economies of scale in avoiding sanctions.

Another example of small sanctions for large banks is the notion that some banks are "too big to fail." That is, regulators will not close a very large bank because of harm to the economy or to many political elites. As a result, very large banks will collude with borrowers. "Too big to fail" implies too big to monitor borrowers. Since smaller less diversified banks will have an incentive to monitor, they may provide more access to funding, more investment, and higher returns to investors.

If per-loan legal sanctions for large banks are not too small, then the incentive benefits of diversification can be harnessed. Reasonably strong and non-corrupt legal and regulatory systems are essential. In addition, a large and well-diversified economy implies that several competing well-diversified banks can co-exist. This reduces "the too big to fail" problem, because closing one of several large banks causes a smaller amount of temporary pain to an economy than closing the only bank in a nation.

Section 8. Conclusion

Weak legal protection encourages misbehavior by borrowers and increases the need for monitored lending. If the legal system provides *very weak legal protection*, then it is impossible to delegate loan monitoring to banks. Banks will always collude with borrowers and will not use any clout they possess over borrowers to collect loans and repay deposits. Small savers anticipate this, and thus banks will not attract deposits. Undelegated direct lending by wealthy family firms will be the main source of finance. If an economy's legal protection is somewhat stronger, delegated monitoring by banks is feasible. The role played by banks depends on the level of protection provided by the economy's legal system, once the level surpasses this minimum.

If there is *strong legal protection* (the sanctions that the legal system can impose are very high), then borrowers and lenders can write debt contracts that sanction debt default by borrowers or bank deposit default by bankers and deter misbehavior by both. However, it is costly to actually impose these legal sanctions if borrowers or banks default frequently. Small investors do not monitor borrowers' ability to repay debt. They can only determine ability to pay by the actual payment they receive, and they will not renegotiate required debt payments if a borrower claims an inability to pay. In this case, the role of banks as delegated monitors is to monitor the ability of borrowers to repay, using the *threat* of legal sanctions to force borrowers to make partial payments when borrowers are unable to pay in full. This avoids the cost of borrower defaults. When many borrowers have bad luck and can make only partial payments, the bank will default and incur costs of bank failure. If banks are well-diversified, the probability that a bank defaults on its deposits due to this bad luck is minimized. Well-diversified banks serve as delegated monitors that allow the strong legal system to keep both borrowers and

bankers from misbehaving without incurring the costs of frequently imposing sanctions due to default. A system with *strong legal protection* that disciplines failing banks will have banks serve as effective delegated monitors if they are well-diversified. This is the simple road to success if legal protection is strong.

When legal protection is *weak* (in between *very weak* and *strong*), ex-ante delegated monitoring can deter borrower misbehavior that legal sanctions cannot, and legal sanctions on banks can provide incentives for banks to use their monitoring clout over borrowers to keep borrowers honest. There is very little wiggle room to make this work, due to the small legal sanctions faced by bankers if they collude with borrowers. With weak legal protection, the details of the structure of the banking system determine if bankers are good (keep borrowers honest) or are evil (encourage borrower misbehavior). It is important to separate banking from commerce lest banks become as evil as a corrupt conglomerate. The benefits of diversification that allow banks to operate efficiently when legal protection is strong become unavailable if larger banks acquire sufficient power to avoid sanctions that would be imposed on small banks. Overall, the strength of legal protection and bank regulation determine the structure and efficiency of the financial system.

Appendix

Proof of Proposition 3.

If a borrower does not divert he must pay F to lenders. If the first borrower pays F but the other does not pay F (and the first borrower does not pay on his behalf), then both get the penalty ϕH, giving the party who does not divert a payoff of $H(1-\phi)- F$. If neither divert, the payoffs are (A,B)= (H-F, H-F). By assumption, under single liability only A will divert.

With reversible diversion, each borrower's outside option if he initially diverts and the other investor can reduce his diversion proceeds by mH and the other can commit himself to divert is $H(1-\phi)-F = \max\{H(1-\phi)-F, H(1-t_A-m-\phi)\}$ for borrower A and $H(1-\phi)-F = \max\{H(1-\phi)-F, H(1-t_B-m-\phi)\}$ for borrower B, by the assumption that monitoring can deter A diversion, $H-F >H(1-t_A-m)\geq H(1-t_B-m)$. Each borrower's outside option is fixed at $H(1-\phi)-F$, so a borrower taking a fixed fraction of the surplus over the outside option will choose the jointly efficient diversion choice: If the bargaining power is such that A gets a μ_A fraction of the surplus over the outside option, the surplus is shared as $(A,B)=(\mu(\text{total surplus}- H(1-\phi)-F)+(1-\mu) H(1-\phi)-F, (1-\mu)(\text{total surplus}- H(1-\phi)-F)+\mu H(1-\phi)-F$. Maximizing total surplus maximizes the payoff of each borrower.

No diversion by either borrower is jointly efficient if $F<\min\{ H(t_A+2\phi), H(\frac{t_A+t_B}{2}+\phi)\}$. A borrower will obtain a fixed proportion of the increase in total surplus between the joint payoff maximizing level and the disagreement point (which is the outside option). The joint surplus if neither divert is $2(H-F) = 2(H-I)$, if both divert it is $H(2-t_A-t_B-2\phi)$, if only A diverts it is $H(2-t_A-F-2\phi) =H(2-t_A-1-2\phi)$and if only B diverts it is $H(2-t_B-F-2\phi)$. It never maximizes joint surplus for only B to divert, because by assumption, $t_B\geq t_A$.

Joint liability leads to no diversion when $2(H-F)\geq H(2-t_A-F-2\phi)$ or when competition implies that F=I, when $I=F\leq H(t_A+2\phi)$ and $2(H-F)\geq H(2-t_A-t_B-2\phi)$ or: $I=F\leq H(\frac{t_A+t_B}{2}+\phi)$, these are jointly true when $I<\min\{ H(t_A+2\phi), H(\frac{t_A+t_B}{2}+\phi)\}$. It leaves diversion identical to individual liability when $H(\frac{t_A+t_B}{2}+\phi)\leq I\leq H(t_A+2\phi)$, which requires $t_B\geq t_A+2\phi$. It leads to joint diversion if $I>\max\{H(t_A+2\phi), H(\frac{t_A+t_B}{2}+\phi)\}$.

Combined with the initial condition that $H(t_B+\phi)\leq I\leq H(t_A+\phi)$, we have the following. If each borrower would divert alone or would not divert given individual liability, there is no effect of joint liability (in this case when diversification effects are absent). If $t_A=t_B$, then there is no

effect of joint liability because joint diversion is profitable if and only if individual diversion is profitable (and asymmetric diversion is dominated). If the borrowers differ in their payoff from diversion, then joint liability can deter diversion by one or extend diversion to the other. Joint liability is most likely to deter diversion if the B borrower has little scope for his own independent diversion (so $H(\frac{t_A + t_B}{2} + \phi)$ is large) and if A's gain from diversion is not too large (so $H(t_A+2\phi)$ is large). If B's incentive to divert is small, but A's large, then joint liability will have no effect on incentives (only A will divert), but the costs of applying sanctions to both will make joint liability undesirable. If B's incentive to divert is not very small, then joint liability with A can induce B to divert under joint liability and incur costly sanctions when he would not under individual liability. This is a particularly undesirable case.

In the case of irreversible diversion, there is another factor, as before. If binding negotiations can occur only after diversion has taken place, and the outside option of a borrower after diverting (e.g., $H(1-t_A-\phi-m)$ for borrower A) is very low, then a borrower with weak bargaining power can be deterred from diversion by joint liability even when diversion maximizes the joint surplus of the borrowers subject to joint liability.

Consider first the case of irreversible diversion where B has all the bargaining power. The payoffs from this case are in table 2 and are explained below.

Table 2: Irreversible diversion, B has all bargaining power.
B has dominant strategy not to divert, and A's best response is not to divert either.

	B not Divert=N	B divert=D
A Not Divert=N	(H-F, H-F)	$(H(1-\phi)-F, H(1-t_B-\phi))$
A Divert=D	$(H(1-t_A-m-\phi), H(1+m-\phi)-F)$	$(H(1-t_A-m-\phi), H(1-t_B+m-\phi))$

The entries in Table 2 are derived as follows (we showed (N,N) above). Consider (D,N) where only A diverts. A's payoff is his outside option of $H(1-t_A-m-\phi)$ because if A rejects his offer and B carries through his threat to stop a crime in progress at zero recovery, B would not pay 2F to avoid penalty ϕH when he could pay F (because by assumption, $F>H(t_A+\phi)\geq H\phi$). To examine B's payoff in this case, note that B can use his threat to collect an unverifiable payment of H(m) which will accrue to B. The payoffs in this case are therefore $(A,B) = (H(1-t_A-m-\phi), H(1+m-\phi)-F)$.

The payoffs for (D,D) where both divert and B uses his threat to reduce diversion proceeds by Hm to extract a nonverifiable side payment from A of Hm are $(A,B) = (H(1-t_A-m-\phi), H(1-t_B+m-\phi))$.

The final case is (N,D) where B alone diverts; in this case B has no threat to use to extract cash from A, we know that A will not pay more than F to avoid the default caused by B's diversion to avoid a penalty of ϕH (because$(H(1-\phi)-F>H-2F$ due to $2F>\phi H$) , so default and sanctions occur; A's payoff is $H(1-\phi)-F$. B's payoff is $H(1-t_B-\phi)$.

Borrower A has a dominant strategy not to divert if monitoring is powerful and $I=F<H(t_A+m)$. This follows because $(N,N) =H-F > H(1-t_A-m-\phi) = (D,N)$ (which is $I=F<H(t_A+m+\phi)$, assumed to be true) , and $(N,D)=H(1-\phi) -F> H(1-t_A-m-\phi) =(D,D)$, because with powerful monitoring, $I=F<H(t_A+m)$. Given this, B's payoff from diversion is $H(1-t_B-\phi)$ and not diverting pays H-F. B will not divert because $I=F<H(t_B+\phi)$ (as B would not divert under single liability).

If monitoring is powerful ($F<H(t_A+m)$), and A has a dominant strategy to not divert, then borrower B will not divert. B's best response to A not diverting is to not divert because B's payoff from $(N,N) =H-F > H(1-t_B-\phi) = (N,D)$ because $F<H(t_B+\phi)$.

If monitoring is less powerful, then $I=F>H(t_A+m)$, and then A's best response to B diverting is to divert as well; $(N,D)=H(1-\phi) –F< H(1-t_A-m-\phi) =(D,D)$, because $I=F>H(t_A+m)$. The unique Nash equilibrium will remain as (N,N) if $I=F<H(t_B)$, because then borrower B has a dominant strategy not to divert: B's payoff from $(N,N) =H-F > H(1-t_B-\phi) = (N,D)$ because $F<H(t_B+\phi)$, and $(D,N)=H(1-\phi) -F> H(1-t_B-\phi) =(D,D)$, if $F<H(t_B)<H(t_B+\phi)$. If instead, monitoring is less powerful $I=F>H(t_A+m)$, and also $H(t_B) <I=F<H(t_B+\phi)$, there are multiple Nash equilibria. B will divert as a best response to A's diversion and not divert as a best response to B's non diversion. A will divert as a best response to B's diversion and not divert as a best response to A's non diversion. .

If instead A has all the bargaining power (receiving a transfer of mH from B if and only if B diverts and never needing to make a transfer to B), then A has a dominant strategy of diverting, and B, who would not divert even when no transfer was paid to a monitor, has a dominant strategy of not diverting if $H(1-\phi)-F> H(1-t_B-m-\phi)$, or $I=F<H(t_B+m)$, which is true because $I=F<H(t_A+m)< H(t_B+m)$

Table 3: Irreversible diversion, A has all bargaining power.
A has dominant strategy to divert, and B's best response is not to divert.

	B not Divert=N	B divert=D
A Not Divert=N	(H-F, H-F)	$(H(1-\phi+m) –F, H(1-t_B-\phi-m))$
A Divert=D	$(H(1-t_A-\phi), H(1-t_B-\phi)-F)$	$(H(1-t_A+m-\phi),H(1-t_B-m-\phi))$

QED.

Proof of Proposition 4. Proof: I consider the case where a bank has sufficient clout to eliminate all of the proceeds of diversion (m is sufficiently large), but the results do not require this. As a result, the bank can force each borrower who has cash H to make a verifiable payment of $F \leq H$

and each borrower with cash L to make a verifiable payment of L. Alternatively, a bank can use its clout to get an unverifiable payment from the diversion proceeds. The amount of this side payment depends on the cash flow of the borrower, but not the number of loans in the bank's portfolio. To avoid messy notation in this sketch, I just denote these unverifiable payments by U_C (either U_H or U_L).

A one loan bank (N=1) must default on its deposits with face value F≥I when its only borrower has cash of L<I, and therefore it fails with probability 1-P. When it fails and legal sanctions are imposed, the depositors recover zero. Let B_N denote the per loan face value of deposits and F_N the face value of loans for an N loan bank. To offer depositors an expected return of I requires that the face value of the one loan bank's deposits are B_1=I/P. To be able to pay these deposits when the borrower has cash of H requires that the one loan bank set the face value, F_1, of its loan such that $F_1 \geq B_1 \geq I/P$. For the one loan bank to prefer to collect a verifiable payment when its borrower has cash of H requires its payoff from this, F_1-B_1, be at least as attractive as taking the diversion payment that it could collect, U_H, and suffering the legal sanction, Φ_{MH1}, obtaining U_H-Φ_{MH1}. Combining this with $B_1 \geq I/P$ yields:

$$F_1 \geq I/P + U_H - \Phi_{MH1} \qquad \text{(One loan bank will not divert).}$$

This condition that the bank not collude and divert must be true for the one loan bank to be viable.

Consider a well-diversified bank (with $N \to \infty$) which charges face F_∞ on loans and sets the face value of deposits per loan of B_∞. If the bank chooses to collect verifiable payments, it will collect F_∞ from the fraction P of its borrowers with cash of H, and it will collect L from the fraction 1-P with cash of L and pay B_∞=I to depositors. Collecting verifiable payments gives the bank a per loan payoff of:

$$P(F_\infty) + (1-P)L - I.$$

If the bank instead shares diversion proceeds with the borrowers, its per loan payoff is: $P(U_H - \Phi_{MH\infty}) + (1-P)(U_L - \Phi_{ML\infty})$.

The bank will collect verifiable payments if and only if:

$$P(F_\infty - U_H + \Phi_{MH\infty}) + (1-P)(L - U_L + \Phi_{ML\infty}) - I \geq 0 \qquad \text{(Well-diversified bank will not divert)}$$

To determine the circumstances where the well diversified bank can undercut the loan interest rates of the one loan bank, I use a loan face of $F_\infty = F_1 - \varepsilon$. From the condition that the one loan bank will not divert, the well-diversified bank can undercut if $F_\infty = I/P + U_H - \Phi_{MH1} - \varepsilon$. Substituting this into the condition that the well-diversified bank will not divert implies that it will not divert if the following condition, (*) is non-negative:

$P(I/P+ U_H-\Phi_{MH1} - \varepsilon -U_H +\Phi_{MH\infty}) + (1-P)(L- U_L+ \Phi_{ML\infty}) - I \;=$

$P(\Phi_{MH\infty}-\Phi_{MH1} - \varepsilon) + (1-P)(L- U_L+ \Phi_{ML\infty}).$ (*)

If the maximum monitor legal sanctions per loan are nondecreasing in N, then $\Phi_{HM\infty} -\Phi_{HM1}\geq 0$, therefore there exists an $\varepsilon>0$ such that condition (*) is positive whenever

$P(\Phi_{HM\infty} -\Phi_{HM1} - \varepsilon) + (1-P)(L- U_L+ \Phi_{ML\infty}) \geq P(-\varepsilon) + (1-P)(L- U_L+ \Phi_{ML\infty})$, or

$$\frac{1-P}{P}(L-U_L+\Phi_{ML\infty})\geq \varepsilon > 0$$

This is positive because the maximum diversion proceeds U_L are always weakly less than L (at very least a fraction $t\geq 0$ is destroyed by diversion), the legal sanction for diversion $\Phi_{ML\infty}$, is positive and $P<1$.

If the maximum legal sanctions per loan are decreasing functions of N, and they become sufficiently small as N gets large, then a well-diversified bank cannot undercut the loan rates of an undiversified bank. Condition (*) is decreasing in ε, thus if it is negative when $\varepsilon=0$, the well-diversified bank cannot undercut; it is negative when $\varepsilon=0$ if and only if:

$P(\Phi_{MH\infty}) + (1-P)(\Phi_{ML\infty}) <P(\Phi_{MH1}) + (1-P)(L- U_L)$.

The right hand side is strictly positive. Therefore, if the total legal sanctions on the bank decrease sufficiently in N, such that this condition holds, the undiversified bank cannot be undercut by the well-diversified bank , in fact it can undercut the well-diversified bank.

This result carries over to comparing a bank making a finite number of loans, $N>1$, to a well-diversified bank. Consider an N loan bank that makes N independent loans and defaults on its deposits with probability $1-P_N$ and thus must set $B_N\geq I/P_N$. A well diversified bank's portfolio is a well-diversified portfolio of the smaller banks' portfolios of loans. If an N loan bank is viable, then the well-diversified bank's incentive to not divert is then automatically satisfied when its sets $B_\infty=I$. There exists a fraction of loans f' such that the N loan bank fails if and only if the fraction f of its borrowers who have a cash flow of H satisfies $f<f'$. A viable N loan bank has an incentive not to divert for all fractions $f\geq f'$. Using this condition for the N loan bank to not divert if $f=f'$ and setting the per-loan face value of its deposits $B_N\geq I/P_N$, condition (*) becomes:
$P_N(\Phi_{MH\infty}-\Phi_{MHN} - \varepsilon) + (1-P_N)(L- U_L+ \Phi_{ML\infty})$, and the results of Proposition 4 hold.
QED

References

Baltensperger, Ernst, "Alternative approaches to the theory of the banking firm," *Journal of Monetary Economics*, Volume 6, Issue 1, January 1980, Pages 1-37.

Benston, George and CW Smith Jr "A Transactions Cost Approach to the Theory of Financial Intermediation," *The Journal of Finance*, 1976, Vol. 31, No. 2, pp. 215-231.

Bolton, Patrick and Xavier Freixas, "Equity Bonds and Bank Debt: Capital Structure and Financial Market Equilibrium under Asymmetric Information", *Journal of Political Economy*, April 2000.

Bond, Philip, "Joint liability among bank borrowers," *Economic Theory* (January 2004), volume 23, issue 2: 383-394.

Calomiris, Charles W. and Charles M. Kahn "The Role of Demandable Debt in Structuring Optimal Banking Arrangements" *American Economic Review* 81:3 (Jun., 1991): 497-513.

Diamond, Douglas W., 1984, Financial Intermediation and Delegated Monitoring, *Review of Economic Studies*, 51, 393-414.

Diamond, Douglas W., "Monitoring and Reputation: The Choice Between Bank Loans and Directly Placed Debt," *Journal of Political Economy* 9: 4 (Aug., 1991), pp. 689-721.

Douglas W. Diamond, 1996. "Financial intermediation as delegated monitoring: a simple example," *Economic Quarterly*, Federal Reserve Bank of Richmond, issue Sum, pages 51-66.

Dubey, Pradeep, John Geanakoplos and Martin Shubik (2005), "Default and Punishment in General Equilibrium," *Econometrica* 73 (1), 1-3.

Gale, Douglas and Martin Hellwig, "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, Vol. 52, No. 4 (Oct., 1985), pp. 647-663

Guinnane, Timothy W. and Maitreesh Ghatak "The Economics of Lending with Joint Liability: Theory and Practice." *Journal of Development Economics* 60:195-228, 1999.

Holmström, B. and J. Tirole [1997], "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 52 (August): 663-692.

Krasa, Stefan and Ann P. Villamil, "Monitoirng the Monitor: An Incentive Structure for a Financial Intermediary," *Journal of Economic Theory*, 57, 197-221, 1992.

Lacker, Jeffrey M., 1991. "Why is there debt?," *Economic Review,* Federal Reserve Bank of Richmond, Jul/Aug 1991, pages 3-19.

Lacker, Jeffrey M. and John A. Weinberg, "Optimal Contracts under Costly State Falsification," *The Journal of Political Economy*, Vol. 97, No. 6. (Dec., 1989), pp. 1345-1363.

Shelifer, A., R. La Porta, F. Lopez-de-Silanes, and R. Vishny "Legal Determinants of External Finance," *Journal of Finance*, July, 1997.

Townsend, Robert, 1979, "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory* 21, 265-293.

Winton, Andrew, 1995, "Costly State Verification and Multiple Investors: The Role of Seniority." *Review of Financial Studies* 8:91-123