

Financial Intermediation as Delegated Monitoring: A Simple Example

Douglas W. Diamond

Banks and other financial intermediaries are the main source of external funds to firms. Intermediaries provided more than 50 percent of external funds from 1970 to 1985 in the United States, Japan, the United Kingdom, Germany, and France (Mayer 1990). Why do investors first lend to banks who then lend to borrowers, instead of lending directly? What is the financial technology that gives the banks the ability to serve as middleman? To answer these questions, this article presents a simplified version of the model in *Financial Intermediation and Delegated Monitoring* (Diamond 1984).¹ The results explain the key role of debt contracts in bank finance and the importance of diversification within financial intermediaries. The framework can be used to understand the organizational form of intermediaries, the role of banks in capital formation, and the effects of policies that limit bank diversification.²

Financial intermediaries are agents, or groups of agents, who are delegated the authority to invest in financial assets. In particular, they issue securities in order to buy other securities. A first step in understanding intermediaries is to describe the features of the financial markets where they play an important role and highlight what allows them to provide beneficial services. It is important to understand the financial contracts written by intermediaries, how the contracts differ from those that do not involve an intermediary, and why these are optimal financial contracts. Debt contracts are central to the understanding

■ The author, the Theodore O. Yntema Professor of Finance at the University of Chicago, Graduate School of Business, is grateful to Tony Kuprianov, Jeff Lacker, and John Weinberg for helpful comments. The views expressed are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

¹ This model is adapted from and extends class notes I have used at the University of Chicago since 1985.

² Some models that extend the role of diversification in Diamond (1984) to other interesting issues in financial intermediation are in Ramakrishnan and Thakor (1984), Boyd and Prescott (1986), and Williamson (1987).

of intermediaries. The cost of monitoring and enforcing debt contracts issued directly to investors (widely held debt) is a reason that raising funds through an intermediary can be superior. Debt contracts include contracts issued to intermediaries by the borrowers that they fund (these are bank loans) and the contracts issued by intermediaries when they borrow from investors (these are bank deposits). Portfolio diversification within financial intermediaries is the financial-engineering technology that facilitates a bank's transformation of loans that need costly monitoring and enforcement into bank deposits that do not.

This article both simplifies and extends the analysis in Diamond (1984). Adding an assumption about the probability distribution of the returns of borrowers' projects makes the analysis simpler. There are a few new results that extend the analysis because this article drops the assumption that nonpecuniary penalties can be imposed on borrowers. The change of assumption implies that there is a minimum level of bank profitability required to provide incentives for bankers to properly monitor loans. This article is not a general survey of the financial intermediation literature. Two recent surveys are Hellwig (1991) and Bhattacharya and Thakor (1993). For a survey of the role of debt in corporate finance, see Lacker (1991).

Intermediaries provide services: this is clear because intermediaries issue "secondary" financial assets to buy "primary" financial assets. If an intermediary provided no services, investors who buy the secondary securities issued by the intermediary might as well purchase the primary securities directly and save the intermediary's costs. To explain the sorts of services that intermediaries offer, it is useful to categorize them in terms of a simplified balance sheet. Asset services are those provided to the issuers of the assets held by an intermediary, e.g., to bank borrowers. An intermediary that provides asset services is distinguished by its atypical asset portfolio. Relative to an intermediary that provides no asset services, it will concentrate its portfolio in assets that it has a comparative advantage in holding. The model presented below provides a foundation for understanding this aspect of intermediation, showing that reduced monitoring costs are a source of this comparative advantage.³ There are other important aspects of intermediation that we do not discuss here: liability services and transformation services. Liability services are those provided to the holder of intermediary liabilities in addition to the services provided by most other securities. Examples include the ability to use bank demand deposits as a means of payment and the personalization of contingent contracts available from life insurance companies. Some liability services, such

³ Fama (1985) notes that banks issue large certificates of deposit which pay market rates of interest for their risk but are also subject to reserve requirements, implying that the reserve requirements are passed along to borrowers. This is evidence in favor of the idea that banks provide asset services.

as check clearing, are well understood, while others relate to difficult issues in microeconomic theory regarding the role of money. Transformation services involve the conversion of illiquid assets into liquid liabilities, offering improved risk sharing and better liquidity compared with investment in the assets held by intermediaries (see Diamond and Dybvig [1983] and Diamond [1995]). Although there may be interactions between these types of service, this article focuses only on asset services.

If intermediaries provide asset services, they provide services to borrowers who issue assets to them. That is, it matters to the issuer of an asset that the asset is to be held by an intermediary rather than directly by investors. Some costs are lower if the asset is held by an intermediary rather than a large number of individuals. The imperfections that give rise to costs of issuing securities by primary borrowers also give rise to similar costs to an intermediary that issues deposits. I examine how a financial intermediary acting as a middleman can lead to net cost savings, and I develop the implications of this role for the structure of intermediaries. The model yields strong predictions about the contracts used by intermediaries and this provides a setting to analyze important issues in banking policy.

1. AN OVERVIEW OF FINANCIAL INTERMEDIATION: THE COSTS AND BENEFITS OF MONITORING

Theories based on the collection of private information by an intermediary require that there be some benefit to using this additional information in lending. A key result in the agency theory literature is that monitoring by a principal can allow improved contracts. The net demand for this monitoring also depends on the cost of monitoring. This cost depends on the number of lenders who contract with a given borrower.

In contracting situations involving a single lender and a single borrower, one compares the physical cost of monitoring with the resulting savings of contracting costs. Let K be the cost of monitoring and S the savings from monitoring. When there are multiple lenders involved, either each must be able to monitor the additional information directly at a total cost of $m \times K$, where m is the number of lenders per borrower, or the monitoring must be delegated to someone.⁴ Delegating the monitoring gives rise to a new private information problem: the person doing the monitoring as agent now has private information. It is not even verifiable whether the monitoring has been undertaken. Delegated monitoring can lead to delegation costs. Let D denote the delegation cost per

⁴ Another option is nondelegated 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. To avoid complicating the analysis, this option is not considered here.

borrower. A complete financial intermediary theory based on contracting costs of borrowers must model the delegation costs and explain why intermediation leads to an overall improvement in the set of available contracts. That is, delegated monitoring pays when

$$K + D \leq \min [S, m \times K],$$

because $K + D$ is the cost using an intermediary, S is the cost without monitoring, and $m \times K$ is the cost of direct monitoring.

The model in this article illustrates the more general results in Diamond (1984), which analyzes delegation costs by characterizing the organizational structure and contractual form that minimize the costs of delegating monitoring to an intermediary. The first step in studying the benefits of intermediation is to find the best available contracts between borrowers and lenders if there is no intermediary and no monitoring. The optimal unmonitored financial contract between a borrower and lenders is shown to be a debt contract that involves positive expected deadweight liquidation costs which are necessary to provide incentives for repayment.⁵ The gross demand for monitoring arises because one can use lower cost contracts (with reduced liquidation costs), if the project's return can be monitored, with an ex ante cost saving of S .

Monitoring is costly, especially if duplicated. If not duplicated, the act of monitoring must be delegated, and then the information obtained is not publicly observed. As a result of the remaining information disadvantage of those who do not monitor, there may be delegation costs associated with providing incentives for delegated monitoring. The best way to delegate monitoring is for the delegated monitor to issue unmonitored debt, which will be subject to liquidation costs. The delegated monitor is a financial intermediary because it borrows from small investors (depositors), using unmonitored debt (deposits) to lend to borrowers (whose loans it monitors).

2. AN EXAMPLE OF OPTIMAL DEBT WITHOUT DELEGATED MONITORING

Consider a borrower who needs to raise a large quantity of capital. All lenders and borrowers are risk neutral, but borrowers have no capital, and each lender's capital to invest is small relative to the amount needed to fund the borrower's investment. The borrower needs to raise 1 (where the units are millions of dollars, and these units will be mentioned only parenthetically), while each investor has $1/m$ units to invest, implying that a borrower needs to raise capital from m investors if $m > 1$. The example assumes that m is very large: $m = 10,000$,

⁵ This analysis of optimal debt contracts is extended in Gale and Hellwig (1985). On the value of monitoring in this situation, see Townsend (1979).

and each lender has capital or 0.0001 (\$100). Monitoring the borrower costs $K = 0.0002$ (\$200), and duplicated monitoring by each of m investors costs $mK = 2$ and is prohibitively expensive. Because monitoring is expensive, one should examine the best contract available without any monitoring.

Investors do not observe the borrower's operations directly, not even its sales or cash flows. How can the lenders write a contract in which they do not need to monitor this information?

The Best Contract without Monitoring

The firm needs to raise 1 (\$1 million), and each investor requires an expected return of $r = 5\%$. All lenders and the borrower agree that the borrower has a profitable, positive net present value project to fund, but only the borrower will observe how profitable it turns out to be. The borrower can consume any part of the project's return that he does not pay out to the investor. The interpretation is that the borrower can appropriate the return to himself, since no one else observes the project's success. If the project is a retail store, the borrower can take some sales in cash to himself. More generally, the borrower can inflate costs. In practice, the net cash flows to the firm are very unobservable for many firms. In addition, most other conflicts of interest faced by borrowers can be reinterpreted as equivalent to the borrower's ability to retain underreported cash. The ability to retain underreported cash is simply the most extreme example of a conflict of interest.

The project costs 1 to fund, and its realized value is a random variable with realization denoted by V . The distribution of V , the value of the project, known to all borrowers and lenders is

$$H = 1.4 \text{ million, with probability } P = 0.8,$$

$$L = 1 \text{ million, with probability } 1 - P = 0.2.$$

A Simple Candidate for a Contract is Equity

An equity contract in this context is a profit-sharing agreement, where the profit shared depends on the profits reported by the borrower. Let the fraction of reported profits that goes to the outside investor be a , while the borrower retains a fraction $1 - a$, plus any underreported profits. Suppose that the borrower's contract is just to pay a fraction of reported profits, with no other details or penalties specified. The borrower's payoff, given the true value of V and the reported value, denoted by Z , is $V - aZ$. What profit will the borrower report if he is supposed to pay out a fraction of it? The borrower will choose the smallest value of Z . Supposing that the borrower can't make the lender take a share of a reported loss (by reporting $Z < 0$), the borrower will report $Z = 0$. A simple profit-sharing contract works very poorly when profits cannot be verified. It does not even provide incentives to repay $L = 1$, the minimum possible value

of profit. Even adding the requirement that profit reports can never be less than $L = 1$ does nothing to induce higher payments.

No matter what the true value of V , the best response of the borrower to a profit sharing contract is to pay the lowest possible value. If there is no cost to the borrower of understating the amount, the borrower always does. Even if the lender knows the value of V , if the borrower obtains it first and thus controls it, the lender will not be paid unless the borrower suffers some consequence of not paying.

What Can the Lender Do If the Borrower Claims a Low Amount?

The lender would like to impose a penalty for low payments to give incentives for higher payments. There are two interpretations. The lender can liquidate the project if the borrower pays too little, preventing the borrower from absconding with it, or the lender can impose a nonmonetary penalty on the borrower. Bankruptcy in the world today is some combination of these two actions. In ancient history, the nonmonetary penalties were very common, i.e., debtors' prisons and physical penalties. Such sanctions are now illegal, but the loss of reputation of a borrower of a bankrupt firm is similar to a sanction.

Bankruptcy, Liquidation, and the Optimal Liquidation Policy

Suppose that it is not possible to impose a penalty on the borrower or take other assets (outside the business) that are valued by the borrower. See Diamond (1984) for analysis when these nonpecuniary penalties are possible. The only sanction available to give the borrower an incentive to pay is liquidation of the borrower's assets (as in Diamond [1989, 1991]). To focus on the inefficiency of disrupting firm operations, I assume that liquidating the firm's asset gives no proceeds to the lender or to the borrower. The results are similar when liquidation yields a positive amount that is much less than the value of the unliquidated asset. Liquidation and bankruptcy are useful penalties that a borrower can avoid by paying the debt, but regular liquidation is not a good way to run a firm. How does one specify an optimal financial contract between investor and borrower when one can decide to liquidate (to penalize the borrower) or not, contingent on any payment?

Liquidation is best used as a payment-contingent penalty in the following way. If the lender is ever to liquidate for a given payment, he also should liquidate for all lower payments. Suppose instead that the lender does not liquidate if 1 is paid but will liquidate for some higher payment. Then, whenever the borrower has at least 1, he will avoid liquidation by paying 1 and keep the remainder for himself. This makes meaningless the threat to liquidate given higher payments, because the payment will never exceed 1.

The borrower will pay the lowest amount that avoids liquidation, and keep the rest for himself. The only exception is if 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, avoids liquidation. The lender then liquidates 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.

Determination of the Face Value of Unmonitored Debt

This section determines the minimum face value, f , of unmonitored debt which will lead to payments with an expected return of 5 percent on a loan of 1 (\$1 million), or an expected value of 1.05.

Suppose $f = 1$. When $V = 1$, the borrower pays 1 (paying less would result in liquidation). The borrower gets 0, which is as much as if he paid any lower amount. When $V = 1.4$, the borrower pays 1 (to avoid liquidation), and keeps 0.4 for himself. This implies that with face value of 1, the lender gets 1 for sure, which is less than 1.05 and not acceptable.

Any face value of debt between 1 and 1.4 forces the borrower into liquidation when the project returns 1 but is paid in full when the project returns 1.4. This gives the lender an expected return of $0.8f$, because nothing is received when there is liquidation. Solving for the face value of debt (between 1 and 1.4) that gives lenders a 5 percent expected return solves $0.8f = 1.05$ and yields $f = 1.3125$. Unmonitored debt with that face value works as follows.

Suppose $f = 1.3125$. When $V = 1$, the borrower pays less than 1.3125, and the asset is liquidated. The borrower gets zero for any payment less than or equal to 1. The best interpretation is that the borrower chooses to pay zero when $V = 1$ because it is the best choice when liquidation is generalized to allow the borrower to keep a positive fraction of the retained cash. This leads the lender to liquidate and receive zero, which occurs with probability 0.2. When $V = 1.4$ the borrower pays 1.3125, avoids liquidation, and keeps $1.4 - 1.3125 = 0.0875$ for himself. This is more than he could get from any smaller payment: any smaller payment gives zero. The payment 1.3125 is received with probability 0.8. Liquidation is only avoided when $V = H$ and the face of 1.3125 is paid. The lender receives 1.3125 with probability 0.8 and zero with probability 0.2, which is an expected payment of $0.8(1.3125) = 1.05$. Any lower face value will give the lender an expected rate of return below 5 percent.

When outside investors cannot observe the cash flows and cannot monitor the business, equity contracts do not work. Enforcing them requires excessively costly monitoring. If this monitoring (sitting on the board of directors or keeping close tabs on the business in other ways) is too costly, then simple financial contracts that do not require monitoring are best. These are debt contracts. They induce the borrower to pay investors because default serves as a penalty that the borrower seeks to avoid.

The analysis can be extended to apply not only to defaults on principal and interest covenants of debt contracts but to any other covenant whose violation implies a potential default on a debt contract. Consider a covenant that might be violated for a variety of hard-to-observe reasons. When it is too costly for lenders to determine the reason for the covenant violation, the covenant will “mean what it says,” and involve a default whenever it is violated, rather than being renegotiated based on the reason for the violation.

The Value of Monitoring

Suppose that it is possible for the lender to monitor the value of the borrower’s operations. Then, instead of liquidating when less than the face value of debt is paid, the lender who monitors can instead use the threat of liquidation and offer to refrain from liquidation so long as the borrower repays as much as possible. Instead of always or never offering to accept 1 in lieu of liquidation, the lender can offer to accept it when $V = 1$ but not when $V = 1.4$. This policy leads the borrower to pay f when $V = 1.4$ and 1 when $V = 1$. I assume that the lender has all of the bargaining power and will offer to accept less than f only when $V = 1$.

The value of monitoring is the expected savings in financial distress costs, which are equal to $0.2(1) = 0.2$. This is the savings from monitoring, S , described in Section 1. This savings must be compared with the cost of monitoring. The cost of monitoring the value of the borrower’s project is K . If there were a single lender, then monitoring would cost K . Duplicated monitoring by each of m lenders would cost mK and would be equivalent to a single lender facing a monitoring cost of mK . I assume that the cost of monitoring is incurred ex ante, before a loan is repaid. Ex ante monitoring implies that the lender must learn in advance about the borrower’s business to properly interpret any data about the project’s return. In this case, the lender or lenders must establish a costly relationship in order to monitor the borrower. The results can be reinterpreted as also applying to ex post monitoring, where no relationship is needed and where the costs of monitoring are incurred only when the borrower defaults on the debt. If the lender or lenders can commit in advance to monitor if and only if the borrower pays less than face value, the ex ante monitoring results can be adapted as follows. In place of the fixed cost of ex ante monitoring, K , use the expected cost of ex post monitoring, which is the cost K , multiplied by the probability that the borrower must default. If the borrower knows he will be monitored given a default, he will default only when he has no choice, i.e., when $V = 1$ (see Townsend [1979]). The ability to wait to incur the ex post cost of monitoring yields an expected cost of monitoring equivalent to an ex ante cost of monitoring of $(1 - P)K$ or $0.2K$.

3. FINANCIAL INTERMEDIATION

If all m lenders monitor, and m is large, then the cost of monitoring is mK , and monitoring is too expensive. If there were many large investors with personal capital above 1, then monitoring at cost K would be available. With a small supply of large investors who can lend 1 on personal account (fewer such investors than profitable large projects), and no way to delegate monitoring, some projects that would benefit from monitoring will be financed with unmonitored, widely held debt. This section shows how financial intermediaries can be set up to create synthetic large investors. There will be a profit opportunity to set up such intermediaries if none are present. If there are few large investors and no intermediaries, then loans are made at 31.25 percent. Finding a way to make monitored loans at 31.25 percent can allow a banker to make a profit. If intermediation reduces the cost of making monitored loans and there is free entry, bankers will not earn excess profits but instead loan rates will be pushed down.

Suppose that there are no large investors, only small investors each with 0.0001 (\$100) to lend, and 10,000 small lenders are needed to finance 1 (\$1 million). Suppose the cost of monitoring V is $K = 0.0002$ (\$200) for each. If each of 10,000 lenders were to monitor whenever there is a default on the loan, the cost would be 2, which is prohibitive, and no one would monitor. When the monitoring cost is prohibitive, the optimal contract is widely held debt with face value 1.3125 (see the subsection entitled “Determination of the Face Value of Unmonitored Debt”). Delegating monitoring to one agent avoids duplication, but can cause incentive problems for the agent who was delegated the monitoring task. Small lenders will not observe the effort put into monitoring, or the information monitored by the agent. The agent (let’s call him or her “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 to monitor the monitor? The answer is for the banker to face liquidation as a function of the amount paid to the 10,000 small lenders (depositors). This provides incentives to the banker in the same way it does to a borrower: the banker is always better off paying a sufficient amount to avoid liquidation.

Liquidation is a sanction that the banker tries to avoid. For simplicity and for symmetry with the assumption made about liquidation of borrowers’ projects, I assume that liquidation of the bank is only a sanction and yields no cash to the small depositors or to the banker. There are several ways to interpret this high cost of bank liquidation. One interpretation is that when too little is paid to the depositors, the assets of the bank’s borrowers are liquidated to make sure that the banker and the borrowers have not colluded to take funds owed to depositors. Another interpretation is that liquidating the bank’s assets consumes all of the assets. In addition, because the banker gets zero

when there is a default on deposits, a banker who anticipates that the bank is about to fail will reduce any discretionary component of monitoring. The reduced monitoring will decrease the value of bank assets. The assumption that borrowers and lenders get zero serves as a simple shorthand for these more complicated aspects of the cost of bank liquidation.

Delegated Monitoring without Diversification Does Not Succeed

Suppose that the banker monitors a single loan (runs a one-loan bank) on behalf of the small lenders, and does not diversify across loans. When the borrower's project returns 1, the banker can monitor and collect the 1 without actually liquidating. However, the bank itself would need to be liquidated in this case, because the face value of the bank's debt must exceed 1. If the bank's debt contract with the small depositors has a face value of 1 or less, the small depositors never receive more than 1, which delivers less than the 1.05 expected repayment they need to receive the required 5 percent expected return.⁶ If the bank is liquidated when its loan defaults by paying 1, the bank is liquidated whenever the borrower would have been liquidated, had the borrower used widely held debt. Unless the 10,000 lenders each monitor the banker (costing 0.0002 each or a prohibitive total of 2), 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 will not succeed.

Can the Banker Use Diversification to Reduce Delegation Costs?

Suppose the banker monitored not one loan, but 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 distributed but are otherwise just like that of the single borrower (each loan has a 0.8 probability of returning 1.4 and a 0.2 probability of returning 1). The banker attracts 2 (\$2 million) in "deposits" from 20,000 investors and lends it out to two different borrowers. The banker gives each borrower a debt contract with face F (\$ F million) and collects F when the borrower has 1.4 and monitors to collect 1 when the borrower has 1. As a result, the banker does not need to use costly liquidation to enforce his loan contract with either borrower. The banker issues unmonitored debt

⁶ In the text I ignore the \$100 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 1 (million), but only 0.9999 (million). The expected repayment to give a 5 percent expected return is then $(1.05)(0.9999) = 1.04995$. This is equivalent to the case where the banker has none of his own capital but outside investors require a 4.995 percent expected return. The one-loan bank is not viable even when only a 4.995 percent return must be given to outside depositors.

deposits that are widely held, and the bank is liquidated whenever it pays less than face value to any investor. This requires no monitoring by the 20,000 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 0.001 (\$100) deposit has face value $\frac{1}{10,000} B$.

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 $1 + F$. If both default, the bank will receive 1 from each, or 2. The diversification from having two borrowers borrow from the bank will reduce agency costs. The distribution of payments to the bank, if the banker monitors, is as follows:

Payment	Probability	Probability that Payment is \geq this value	Explanation
$2F$	$0.64[P^2]$	0.64	both pay F
$F + 1$	$0.32[2(P)(1 - P)]$	0.96	one pays F , one 1
2	$0.04[(1 - P)^2]$	1.00	both pay 1

Assume that liquidating the bank yields nothing to depositors or to the banker, similar to the liquidation of borrowing firms. The bank has total face value of deposits of $2B$. If the bank must be liquidated when it collects face value of F from one borrower and 1 from the other, it will be liquidated whenever at least one loan defaults, and there will be no possible savings in costs of financial distress. Alternatively, if the bank can and will pay its deposits when one loan defaults, it defaults only when both loans default, and it can reduce the probability of liquidation to $0.04 = (1 - P)^2$. To examine when payment of all deposits is possible when just one loan defaults, the total payment received by all depositors will be $2B$ with probability 0.96 and 0 with probability 0.04. The expected payment is $0.96(2)B$. The initial capital needed to make two loans is 2 (\$2 million), and it requires a 5 percent expected rate of return, implying that $0.96(2)B = 2(1.05)$, or $2B = 2.1875$, is the promised payment to 2 (\$2 million) in deposits. Equivalently, let the promised interest rate on bank deposits be r_B . Then, because $2B = 2(1 + r_B)$, the promised interest rate on the bank deposit is $r_B = 9.375\%$.

If the bank is to be able to pay 2.1875 when one loan defaults (paying 1) and the other does not default (paying F), then $1 + F$ must be at least 2.1875, and the face value of each loan must satisfy $F \geq 1.1875$. If the bank made loans with this face value, it could avoid liquidation with probability 0.96. In summary, if the bank monitors its loans, it will have the cash and the incentives to pay bank deposits in full with probability 0.96 so long as $F \geq 1.1875$ or the interest rate on bank loans is at least 18.75 percent.

Will the Bank Monitor?

A remaining question is whether the banker will choose to monitor the loans. Without monitoring, the bank would not be able to offer to take 1 when only 1 is available and would instead liquidate the borrower's asset. 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 and is liquidated). The entire increase in the banker's return comes from increasing the return when just one loan defaults.

If the banker who monitors obtains nothing whenever at least one loan defaults, there will be no incentive to monitor. An incentive to monitor requires that monitoring increases the bank's expected payment by at least 0.0002 (\$200) per loan. If the banker monitors neither loan, then the bank will fail when just one loan defaults, and the banker will get zero. If a loan that is monitored defaults, and the other loan does not, the banker's return will be $1 + F - 2B = 1 + F - 2.1875$. This is the ex post increase in the banker's return due to monitoring. Monitoring one of the loans gives this increased return with the probability that it alone defaults, or with probability 0.16. Monitoring of one of the loans will be in the banker's interest if $0.16(1 + F - 2.1875)$ exceeds the cost of monitoring or 0.0002. Monitoring one loan will pay if $F \geq 1.18875$. Monitoring both loans gives the same increased return with the probability that one of the two loans is the only default, or with probability 0.32. Monitoring both loans is in the banker's interest so long as $0.32(1 + F - 2.1875)$ exceeds 0.0004, which also implies $F \geq 1.18875$. So long as the interest rate on bank loans exceeds 18.875 percent, the banker is willing to invest \$400 worth of time to monitor all loans because it increases the value of his residual claim on the bank.

The two-loan banker must earn a small profit in excess of the cost of monitoring. The need to provide the bank an incentive to monitor and to avoid bank failure when just one loan defaults (by cross-subsidizing the losses from the defaulting loan with the profit from the nondefaulting loan) leads to profits for the banker who was delegated the monitoring of the loan. The banker will monitor only if it yields a profit, and due to limited liability and limited wealth, the banker never makes deposit payments in excess of loan repayments. The need to provide incentives puts a floor on the banker's expected profit, which is sometimes called a control rent, because the banker's control of decisions requires that the rent (profit) go to him. 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 2.1875, or

$2.3775 - 2.1875 = 0.19$, with probability 0.64, when neither loan defaults;
 $2.18875 - 2.1875 = 0.00125$, with probability 0.32, when one loan defaults;
 and 0, with probability 0.04, when both loans default.

This works out to a total expected payment of 0.122 (\$122,000) or $(0.19)0.64 + (0.00125)0.32 = 0.122$. This is a return to the banker of 0.061 per loan, which is in excess of 0.0002 the cost per borrower of monitoring, and the banker earns a control rent of $0.061 - 0.0002 = 0.0608$.⁷

The delegation cost per borrower, D , equals the cost of financial distress of the bank or $0.04(2) = 0.08$, plus control rent to the banker of 0.0608 or a total of 0.1408. All parties are better off with the banker as delegated monitor. The borrower prefers to borrow at 18.875 percent from the bank, versus at 31.25 percent direct. The investors get a 5 percent expected return in either situation. The banker is happy with any claim with an expected payment above \$400 and ends up with an expected payment of \$122,000.

Summary of Financial Intermediation and Diversification

I consider three types of contracting arrangements: (1) no monitoring: a widely held traded debt contract with face = 1.3125 for each borrower; (2) direct monitoring by investors, which saves distress costs of $S = 0.2$ but costs $mK = 2$; and (3) delegated monitoring by an intermediary, which saves distress costs $S = 0.2$ at cost monitoring plus delegation cost, $K + D = 0.1408$.

Diversification within the intermediary works to make option (3) work by reducing the liquidation cost of providing the bank an incentive to repay small investors. To simplify, I use an example where the diversification from a bank making only two loans was sufficient to give the bank reduced delegation costs. However, it is more generally true that diversification allows financial intermediation to provide low-cost delegated monitoring. 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 will default 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 liquidation costs.⁸ In addition, the control rent needed to

⁷ One can do a bit better, as in footnote 6. The banker has capital of his or her own to invest. The bank need not raise 2 (million), but only 1.9999 (million). The face value of bank debt owed to depositors is then $1.9999 \frac{1.05}{0.96} = 2.18739$. This allows the face value of bank loans to be reduced slightly. The binding constraint is the banker's incentive to monitor, or $0.0004 \leq 0.32(1 + F - 2.18739)$, implying that the face value of bank debt is $F = 1.188614$ (instead of 1.18875). This leads to a payoff to the banker of 0.118721. Because the banker spends 0.0004 of his time on monitoring and is owed 0.0001(1.05) for his 0.0001 capital, there is a total control rent of 0.118216, or 0.059108 per loan.

⁸ For a formal limiting argument about well-diversified intermediaries, see Diamond (1984), and for a generalization see Krash and Villamil (1992).

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 capital down to 5.02 percent. In the limit of perfect diversification, the face value of bank debt approaches $F = 1.06275$, which is the solution to $0.8F + 0.2(1) = 1.0502$; it gives the bank a 5 percent expected return after covering the 0.0002 (\$200) cost of monitoring. This is too strong because in practice the default risk of borrowers is not independent, it is positively correlated. In addition, the number of loans in the bank's portfolio is limited.

The general message is that diversification allows banks to transform monitored debt into unmonitored debt, delegating the monitoring to bankers. The banks' organizational form minimizes the sum of monitoring and financial distress costs.

Policy Implications

There are important implications of this view of intermediaries. Because there are costs of bank failure, and there are incentive benefits from the bank receiving the profit derived from its monitoring, banks can increase their value by hedging or avoiding risks that they cannot control or reduce via monitoring. For example, monitoring can do nothing to influence the level of riskless interest rates. Thus, there is no incentive reason for the bank to bear general interest rate risk. The bank's high leverage means that a small loss might force a costly default. Hedging of interest rate risk is desirable, through futures markets or interest rate swaps, because it can remove risks that have no incentive value to bank managers. Banks rely on diversification to eliminate the risks of being very highly levered. Unless a risk is intimately related to their monitoring task, banks should avoid risks that are not diversifiable unless the bank can remove the risk from its balance sheet through another (swap or futures) transaction.

Diversification makes bank deposits much safer than bank loans, and in the limit of fully diversified banks with independently distributed loans, bank deposits become riskless. This suggests that even without deposit insurance, deposits ought to be very low risk. Laws that limit bank diversification remove much of the technological advantage of the banking contract. The prohibition on interstate banking in the United States, only recently eliminated, made delegation costs much larger and banks much riskier than they would be without the prohibition. The delegation cost from excessively limited diversification has two components. One is the increased probability of bank failure, which may also have contributed to the historical political pressure for deposit insurance. The other component is excessively high control rents: small undiversified banks require higher levels of future profits to remove their manager's otherwise poor incentives. This suggests that in the United States, where the economy

is large enough to have several competing, well-diversified intermediaries, the increased diversification from geographical deregulation may reduce managerial moral hazard and help eliminate the need for high future bank profits (high charter value) to provide good incentives to bankers. If this is correct, banks and similar financial intermediaries will be more stable in the future than in recent experience in the United States.

4. CONCLUSIONS

The purpose of this article is to clarify the roles of debt and diversification in the financial engineering that is banking. Debt has several roles related to financial intermediation. The right to liquidate on default provides any outside lender with power over the borrower, inducing the borrower to repay the debt. This power is limited by the borrower's right to repay the debt in full and remove the lender's liquidation rights. However, liquidation is potentially inefficient. If the lender cannot monitor the borrower's business, then the lender should liquidate whenever there is a default, no matter what the cause. If the lender can monitor the situation, then the ability to selectively remove the threat to liquidate in return for a concession from the borrower can provide power over the borrower without using inefficient liquidation. Financial intermediaries such as banks can centralize costly monitoring and avoid the duplication of effort of the monitoring of borrowers by small investors. Banks monitor debt (loan) contracts, and issue unmonitored debt (deposit) contracts. Diversification is the financial-engineering technology that makes monitoring of deposit contracts unnecessary when monitoring of loan contracts is necessary. This allows banks to deliver delegated monitoring. Debt, monitoring, and diversification are the keys to understanding the link between financial intermediation and delegated monitoring.

REFERENCES

- Bhattacharya, Sudipto, and Anjan V. Thakor. "Contemporary Banking Theory," *Journal of Financial Intermediation*, vol. 3 (October 1993), pp. 2–50.
- Boyd, John H., and Edward C. Prescott. "Financial Intermediary-Coalitions," *Journal of Economic Theory*, vol. 38 (April 1986), pp. 211–32.
- Diamond, Douglas W. "Liquidity, Banks, and Markets," University of Chicago CRSP Working Paper. December 1995.
- . "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt," *Journal of Political Economy*, vol. 99 (August 1991), pp. 689–721.

- _____. "Reputation Acquisition in Debt Markets," *Journal of Political Economy*, vol. 97 (August 1989), pp. 828–62.
- _____. "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, vol. 51 (July 1984), pp. 393–414.
- _____, and Philip H. Dybvig. "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, vol. 91 (June 1983), pp. 401–19.
- Fama, Eugene F. "What's Different about Banks?" *Journal of Monetary Economics*, vol. 15 (January 1985), pp. 29–39.
- Gale, Douglas, and Martin Hellwig. "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, vol. 52 (October 1985), pp. 647–64.
- Hellwig, Martin. "Banking, Intermediation and Corporate Finance," in Alberto Giovannini and Colin Mayer, eds., *European Financial Integration*. Cambridge: Cambridge University Press, 1991.
- Krasa, Stefan, and Anne P. Villamil. "Monitoring the Monitor: An Incentive Structure for a Financial Intermediary," *Journal of Economic Theory*, vol. 57 (June 1992), pp. 197–221.
- Lacker, Jeffrey M. "Why Is There Debt?" Federal Reserve Bank of Richmond *Economic Review*, vol. 77 (July/August 1991), pp. 3–19.
- Ramakrishnan, Ram T. S., and Anjan V. Thakor. "Information Reliability and a Theory of Financial Intermediation," *Review of Economic Studies*, vol. 51 (July 1984), pp. 415–32.
- Townsend, Robert M. "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, vol. 21 (October 1979), pp. 265–93.
- Williamson, Stephen D. "Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing," *Quarterly Journal of Economics*, vol. 102 (February 1987), pp. 135–45.
- Winton, Andrew. "Costly State Verification and Multiple Investors: The Role of Seniority," *Review of Financial Studies*, vol. 8 (Spring 1995), pp. 91–123.