FINANCIAL INTERMEDIATAION: DELEGATED MONITORING AND LONG-TERM RELATIONSHIPS

by

Joseph G. Haubrich

9-86

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104

FINANCIAL INTERMEDIATION: DELEGATED MONITORING AND LONG-TERM RELATIONSHIPS

by

Joseph G. Haubrich
Department of Finance
Wharton School
University of Pennsylvania

June, 1985
revised March, 1986

This paper was written while the author was a Visiting Scholar at the Board of Governors. Ed Green provided many helpful comments, but bears no responsibility for the remaining errors. The views expressed here do not necessarily represent the position of the Federal Reserve System.

I. INTRODUCTION

Practical men have always considered long-term relationships an important factor in banking, but explicit recognition of this fact among economists has been rare. In particular, the recent information-based general equilibrium models of banking have focused on essentially single-period schemes and generally ignored multi-period frameworks.[1] This paper represents an initial attempt to consider the consequences of enduring relationships for bank structure and policy in an information based banking model, specifically, that of Diamond's delegated monitoring rubric (Diamond, 1984). Adding a richer time dimension allows repeated lending between the intermediary and borrower. This contract lets a bank produce information and enforce compliance more easily than direct monitoring of the borrower's organization. That is, rather than directly monitoring compliance with costly, complicated, covenants that prevent a debtor firm from cheating or shirking, the bank simply monitors the net outcome over time, holding the threat of no credit over the firm.

The bank itself arises for reasons of delegated monitoring. As in Diamond's model, monitoring of a firm (even via long-term contracts of enduring relationships) by many lenders is extremely costly. The lenders have the incentive to trust the monitoring to a third party who will avoid this duplication. While this intermediary must be monitored itself, Diamond has insightfully shown that the cost of monitoring the intermediary (called the "delegation cost") get arbitrarily small as the intermediary diversifies across borrowers. This occurs because the optimal contract between lenders and the bank is a debt (or deposit) contract on which default is unlikely

---

[1]See, for example, Diamond (1984), Diamond and Dybvig (1983) or Haubrich and King (1983).

because the intermediary is diversified across many borrowers. Furthermore, as the number of borrowers increases, the fixed size of the bankruptcy costs gets spread across more borrowers, making the per borrower cost of monitoring the intermediary very small.

All this has some important ramifications. First, an asymmetry results in banking structure. The bank will enter a long term relationship with projects it funds, because that is cheaper than direct monitoring, but depositors will not enter a long term relationship with the bank, because the debt contract will be cheaper for a well-diversified bank. Secondly, this sheds some light on questions of loan participation and security underwriting by commercial banks. Under certain plausible circumstances, there will be no "lemons" problem associated with selling loans to other intermediaries and thus constraints on selling loans and underwriting some securities are simply misplaced regulations. Finally, this paper provides some explanation for the nature and real effects of banking panics: when an intermediary appears to be about to go under, borrowers will not face the threat of punishment in the future. Thus they cheat, hasten the demise of the bank, and ultimately drive up the cost of borrowing and credit formation.

In the remainder of the paper, section II describes and proves the fundamental results about long term relationships. Section III incorporates those results into a delegated monitoring model of banking based on Diamond (1984) and section IV provides a summary and concluding comments.

II. Enduring Relationships

The ability to write contracts contingent on past behavior has become a popular topic in recent economics. Radner (1981), Rubinstein and Yaari (1983), and Allen (1985) have all examined the uses of these contracts by principles and agents. Townsend (1982), in closely related work, has concentrated on optimal borrowing and lending contracts of the sort important for macroeconomics. Similar applications of game theory to other situations abound: for example, Friedman (1977), who studies multi-period oligopolies, or earlier still Luce and Raiffa (1957), who consider prisoner's dilemma supergames.

The more recent work is all based on one simple core idea. Rather than directly monitor the agent, the principal can observe the claimed outcome of the agent, and then do a statistical test to see if the agent is cheating or shirking. In a one period principal-agent setting, the agent can claim that he worked hard, but that the productivity variable had a bad draw and so output was low. Several hundred "bad draws" in a row might make the principal suspicious, and he could then punish the agent. This then should make the basic strategy of the principal clear: compare the sample mean of what the agent has declared with the (known) mean that would occur if the agent were honest and not shirking. If the mean of the reported sequence deviates too far from the true mean, punish the agent (for example, refuse to lend to him, charge him higher rates, etc.). Notice that the statistical test employed by the principal must be somewhat delicate. The bound on the permitted deviations from mean must be tight, or the agent will have room to cheat. On the other hand, the bound cannot be too tight or it will punish honest but unlucky agents who really did have a bad draw in that period. It was the great insight of Radner and Rubinstein and Yaari to exploit the statistical work on "tests of power 1" which allow a bound converging to zero, tight enough to prevent cheating, yet converging slow enough to prevent unwarranted retribution.

We now present the basic model of this paper and show the benefits of long term relationships. This is a simple moral hazard model. There are agents, who are entrepreneurs endowed with a risky project and zero wealth. The project takes 1 unit of goods to become active; if activated, it pays off $\tilde{y} = [0, \bar{y}]$. Only agents observe the payout of their project. For now, consider the principal as the bank, (justified in the next section) capable of lending the unit to the project owner. Everyone in the economy is risk neutral, lives forever, has no time preference (i.e., the discount factor is 1) and uses the average utility criterion to compare infinite streams of payments.

This setting is meant to be a simple one in which to establish the usefulness of long term contracts. The unobservability of output proxies for more realistic forms of moral hazard, such as unobservable effort or project risk. The average utility criterion compares infinite streams of income by comparing the average utility of the agent over the stream. While other criteria exist (such as the overtaking criterion), average utility provides an intuitive, convenient, and commonly used way to evaluate undiscounted infinite streams. The remainder of this section shows how longer term contracts enable the principal to replicate the full-information solution. That is, a contract exists that induces truth from the agent.

In showing that a contract can induce the agent to truthfully reveal each period's outcome, three cases must be distinguished. First, the contract must be specified so that should the agent systematically understate his outcome, he will be detected and punished. (Note that since the agent owes a return to the principal, the incentives are for understating outcomes, not overstating.) Secondly, however, the criterion for punishment must not be too strict, else it will induce either improper punishment or cause the agent to over-report output. Finally, if the agent neither systematically under- or over-reports, the agent should choose truthful messages. Were agents risk averse, this would be trivial by Jensen's Inequality.

Now consider the following strategy by the principal, which we will show induces truth-telling in the agent. First, consider the compensation scheme used by the principal. If the agent claims output in period t was y_t, then the bank (principal) takes $g(y_t) < y_t$, and the agent keeps $y_t - g(y_t)$, where $g(\cdot)$ is strictly increasing in y_t. This payment schedule is adhered to as long as the agent stays in the good graces of the principal. If the agent gets caught, the principal stops all lending for a finite "penalty phase," leaving the agent to consume 0 for those periods. With the trigger, and length of these penalty periods chosen correctly, cheating firms find themselves continually without loans, while truthful firms are only occasionally constrained. This can be shown more rigorously along the lines suggested by Rubinstein and Yaari (1983) and Radner (1985), although this particular form of the principal agent problem requires a modified proof.

Let the principal adopt the following strategy. Start with a series $\{p_t\}$ such that $0 < p_t \leq 1$ and $\sum_{t=0}^{\infty} p_t$ converges. If the agent has been punished (t - 1) times already, determine the probability of the agent's record being what it is if he were telling the truth. That is, find the probability that $\frac{1}{N} m_N$, the sample mean of his messages, comes from a process with E(X), where E(X) is true mean of the i.i.d. process generating output. If that probability is less than p_t, punish the agent. Withhold capital so that the agent cannot invest in his project, and so gets only his autarkic utility, $\overline{U}$. Punish the agent until $\frac{1}{t} U_t \leq \overline{U} + \varepsilon_t$, where $\varepsilon_t \searrow 0$.

This strategy does not hurt agents who tell the truth. If an agent correctly reveals X_t, the probability that the sample mean will be "too far" from the true mean is, by definition, p_t. Thus, by the Borel-Cantelli Lemma

(See Breiman 1968) since $\Sigma p_t < \infty$, the probability of punishment happening infinitely often equals 0. Thus with probability 1 the truthful agent bears only a finite number of punishments, which do not affect his lifetime utility, $\lim_{t\to\infty} \frac{1}{t} U_t$.

This strategy also lets the principal identify and punish cheaters. The agent has some strategy for choosing messages m_t, and benefits only if $\lim_{t\to\infty} \frac{1}{t} \Sigma m_t < E(X)$. The important choice for the agent is the policy or rule, not any particular period's message. If the agent's rule for choosing m_t, given all observed X_t's and the principal's strategy, is measurable, then $\{m_t\}$ is a random process. Since $E|m_t| < \infty$, the ergodic theorem implies $\frac{1}{t} \Sigma m_t \to E(m_1|J)$ a.s., where J denotes the σ-field of invariant events of $\{m_t\}$. The agent profits only if $E(m_1|J) < E(X)$. Now recall $\{p_t\}$. This sequence corresponds to bounds α_t, where punishment occurs if $|\frac{1}{n} \Sigma m_n - E(X)| \leq \alpha_t$ and $\alpha_t \searrow 0$. By the ergodic theorem, for any $\frac{\varepsilon}{3}$ there exists N so that if $n > N$ then $|\frac{1}{n} \Sigma m_n - E(m_1|J)| < \frac{\varepsilon}{3}$. Now choose T* such that $t > T^*$ implies $\alpha_t < \frac{\varepsilon}{3}$. Finally, choose ε so that $E(X) - \varepsilon > E(m_1|J)$. Thus the agent violates the α_t bounds with probability 1. Again using the Borel-Cantelli Lemma, since Σ Prob (violating α_t) $= \infty$ we know the agent is punished infinitely often with certainty. This infinite punishment deters the agent. In each punishment phase $\frac{1}{T} \Sigma U_T < \bar{U} + \varepsilon_t$, so that $\lim_{t\to\infty} \frac{1}{t} \Sigma U_t = \bar{U}$ provided the limit exists. The limit can be shown to exist either by restricting strategies to those supporting convergent utility (as Rubinstein and Yaari do) or by another application of the Borel-Cantelli Lemma.

The final case to consider is if $E(m_t) = E(y_t)$, will the agent claim $m_t = y_t$? Here, since the agent is risk neutral, he doesn't care where or when he gets the consumption, so make the usual assumption that when indifferent, the agent tells the truth.

III. DELEGATED MONITORING AND FINANCIAL INTERMEDIATION

The previous section showed that enduring relationships can induce truth telling in agents. We now incorporate that form of "monitoring" output into a model of financial intermediation based on monitoring. The key point here is that the diversification of an intermediary allows a debt (deposit) contract with a cost of ensuring bank performance that is very small. Thus, while a bank will use a long term agreement with borrowers, it will have only short term arrangements with depositors.

More specifically, this model leads to a financial institution that is highly leveraged, debt producing, and information producing. It uses long term agreements to monitor borrowers' producing information and enforcing contract compliance. An important point is that these long term agreements will replace direct monitoring of the project by the banks, because it is cheaper to check payout history and verify satisfaction of the sample mean bounds that to directly examine the working of the corporation. [For example, in the late 1970's some banks loaning to motion picture corporations had officers reading movie scripts.]

Since the debt of this institution need not circulate as inside money, and the loans it makes may appear closer to equity than debt, it more closely approximates a generic financial intermediary than a bank. Still, to add concreteness, and in the belief that this basic structure undergirds banks as well, this paper uses the more specific term.

The basic model follows Diamond (1984) quite closely. Entrepreneurs in the economy are endowed with 0 wealth and a project requiring 1 unit to produce. The returns to these projects are independent and observable only to the project's owner. There also exist investors, each with 1/m units of the good. They can invest the good in projects, or for a certain return R. These investors have available to them three technologies for checking up on their investments. Obviously something of the sort must exist, or the project owners would claim a bad outcome, and only the certain investment would be chosen. First, investors may directly monitor the project at (disutility) cost K per project. By paying this cost, they directly observe the output of the monitored project. Again following Diamond, this "costly state verification" must take place before y_t is realized. That is, this verification is monitoring from the start of the project: it is having your officers read the scripts before Rocky V comes out, rather than sending in accountants if the profits were too low. The second form of monitoring is the long term relationship examined in section II.[1] Here, with a cost $0 < a < K$, the investor can determine the true output of the project. This contract is cheaper, since we assume that the cost of conducting a statistical test on the messages of the agent is cheaper than direct monitoring, though not costless. Thirdly, there can be contracts with non-pecuniary penalties. Thus, the investors can impose costs on the project-owner without having to expropriate any material--indeed even if there is no material to appropriate. These are meant to include time spent in bankruptcy court, search costs of a fired manager, etc.

If entrepreneurs borrow directly from investors, one of two contract types will emerge. One possibility is for each investor to monitor the project, leading to a total monitoring cost per project of mK or ma. Presumably this is prohibitively expensive. The other possibility involves the non-pecuniary penalties. This will take the form (see below) of debt

---

[1]With many principals and agents, we simplify the problem by assuming agents cannot shift principals. This sidesteps a potentially central issue. Note that the market context raises this issue, and also that a modelling strategy emphasizing reputation (see Klein and Leffler, 1981) may solve it.

contract, with penalties imposed if claimed output falls below some face value. Otherwise, the agent pays the investor the face value of the debt.

If we no longer restrict ourselves to direct investment, though, there is a third possibility. If one firm were to monitor each investment on behalf of all the investors, the per investment monitoring cost would be only a or K. The total cost would be more than that, however, for each depositor must in turn check up on the monitoring firm. This necessity of monitoring the monitor leads to an intermediary rather than a purely information producing firm, such as a bond rating agency. Directly monitoring or even entering an enduring relationship with the bank (monitoring firm) is overall more costly than each investor directly checking the project for himself. However, it may be that a debt contract with the intermediary may provide cheaper means of checking it. We can formalize this notion by introducing the delegation cost D, which is the cost of providing proper incentives to the firm doing the monitoring. That is, whether it reflects the costs of agents directly monitoring the monitor or the implicit expected cost of an incentive contract, D denotes the cost to investors of making the delegated monitoring system work. This delegated monitoring will be profitable if
$K + D < \min[E(\phi(y)), mK]$ where ϕ denotes the cost to the entrepreneur of the non-pecuniary penalty debt contract. Alternately, replace K with a for the comparison involving long term agreements. The point of this section is to show that i) the delegated monitor will use the least costly monitoring technology, and ii) the optimal contract provides incentives to monitor projects, and iii) delegated monitoring will be chosen. This builds on Diamond (1984) by adding the technology to monitor via long-term agreements.

Let there be i = 1, 2, ..., N projects and associated entrepreneurs, which will thus be financed by mN investor-depositors. Denote the return to a particular project by the random variable y_i. Note we have dropped the time subscript for y. Let $g_i(y_i)$ be the function telling what the entrepreneur pays to the intermediary firm as a function of y_i. The output of project i, y_i, is known to the intermediary (often called a bank hereafter) if the intermediary directly monitors (at cost K) or enters an enduring relationship (at per period cost a) with the project. We assume this happens, i.e., that K or a < $E[\phi(y_i)]$ since otherwise there would be no need for a bank: depositors would hold project debt directly. Likewise, depositors will not monitor the bank, for then the per project cost will be a(m + 1)--they could eliminate the middle man and monitor the projects themselves.

As we now show, delegated monitoring can occur if depositors enter a debt with non-pecuniary penalty contract with the bank. That is, the bank will pay off a face value H if it can, and suffer bankruptcy penalties if it cannot. As the number of projects borrowing from the bank, N grows large, the cost of the penalty per project grows small as the odds that all of the project do poorly approach zero. Thus, consider $G_N = \sum_{i=1}^{N} g_i(y_i)$ being the total payment the bank gets from all the projects it monitors. The bank must pay out some of that to its depositors (who after all are the ultimate lenders in this economy). The depositors have the alternative opportunity of investing at a sure return R, so the bank must provide payments Z_N with expected value equal to NR (recall investors are risk neutral). Notice that $Z_N \leq G_N$. Also notice that unless $\Pr(G_N \geq NR) = 1$, (in which case there is no reason for a bank) there is some probability that the bank cannot give the depositors what was promised. Thus some monitoring or incentive structure must exist for the bank.

A debt contract can serve this purpose. When people can impose non-pecuniary costs on the bank, Diamond has shown that the optimal contract takes the following form:

Impose penalty $\Phi(Z_N)$ where

$$\Phi(Z_N) = \max[H_N - Z_N, 0] ,$$

where H_N is the smallest solution to

(4) $$\Pr(G_N < H_N) \cdot E\left[G_N | G_N < H_N\right] + \left[\Pr(G_N \geq H_N)H_N\right] \geq NR .$$

In words, the contract is a debt contract with face value H_N. If the bank chooses not to meet that payment, it endures penalties equal to the difference between the face value and what it pays out. H_N is chosen to be the smallest number such that the expected value of the contract to the lender is R. For a proof, see Diamond (1984).

From this contract, the return to the bank is $E(G_N) - H_N$. This follows from (4). If the bank does not default, it gets G_N and pays out H_N. If it does default, it still gets G_N, and pays $Z_N - (H_N - Z_N)$. The total utility of the bank is then

(5) $$E(G_N) - H_N - N(\text{monitoring cost})$$

where the monitoring cost may be a or K.

The objective of the bank is then to

(6) $$\text{MAX}\left[E(G_N) - H_N - N(\text{mon.cost})\right] .$$

Equation (6) tells us two very important things. First, that the bank wishes to minimize its monitoring costs, so it monitors using the long term relationship, and secondly, that it will monitor projects since that maximizes $G_N(y)$. All that remains to be shown is that this contract involving delegated monitoring is in fact cheaper than environments without a bank. We are thus ready to establish, for this model, the analogue of Diamond's Proposition 2: The cost of delegation per entrepreneur monitored, D_N, approaches zero as $N \to \infty$ if projects have bounded returns distributed independently. As in that paper, the key will be that the delegated monitoring cost is the expected value of bankruptcy costs for the bank.

The proof will follow Diamond (1984) very closely. Choose payment schedules $g_i = g_i(y_i)$ to the bank for projects $i = 1, 2, \ldots, N$ to be such that $E(g_i) = R + a + D_N$. That is, the payments from the project to the bank cover good due depositors, monitoring costs of the project, and monitoring cost of the bank. Then entrepreneurs keep, on average, $E(y) - E(g) = E(y) - R - a - D_N$. Choose the non-pecuniary bankruptcy penalty for the bank to be

$$\Phi(Z_N) = MAX[Z_N - H_N, 0]$$

where $H_N = N(R + \frac{1}{2}D_N)$. As usual, then, the bank sets Z_N to be

$$Z_N \begin{cases} G_N \text{ if } G_N < H_N \text{ (that is, if the bank defaults)} \\ H_N \text{ if } G_N \geq H_N \text{ (that is, if it pays off the face value of the debt)} \end{cases} .$$

The expected utility (net of monitoring costs) to the intermediary is $E(G_N) - H_N - Na = \left[N(R + a + D_N)\right] - \left[N(R + \frac{1}{2}D_N)\right] - aN = \frac{1}{2}N > 0$. Hence the intermediary makes a non-negative profit and agrees to the game. Depositors get, if we define $P_N = \text{Prob}(G_N \leq H_N)$,

(7) $$P_N E(G_N | G_N \leq H_N) + (1 - P_N)H_N.$$

Where the solution for H_N implies

$$(1 - P_N)H_N = (1 - P_N)N(R + \tfrac{1}{2}D_N).$$

If P_N is small, that is, if $P_N \in (0, \frac{1}{2}D_N/(R + \frac{1}{2}D_N))$, then (7) > NR. Thus if we can make P_N small, depositors will be satisfied. But P_N is just the probability that G_N will be far below the average value. We rule this out by diversification. By the Weak Law of Large Numbers, for any $\delta > 0$, there exists an N^* s.t. if $N > N^*$, $P_N < \delta$ since $E(G_N) > H_N$. This last inequality follows because $H_N = N(R + \frac{1}{2}D_N)$ which $E(G_N) = N(R + a + D_N)$. Thus, the delegation cost D_N can be made arbitrarily small provided N is large enough. As Diamond puts it, "The intermediary need not be monitored because it takes 'full responsibility' and bears all penalties for any short-fall of payments to the principals. The diversification of its portfolio makes the probability of incurring the penalties very small and allows the information collected by the intermediary to be observed only by the intermediary."

IV. CONCLUSIONS

I think four major points emerge from this model.

First, banks may have an asymmetric structure with respect to long term arrangements. They enter enduring relationships with borrowers, but not with depositors. This appears to concur with reality.

Secondly, although this model explicitly allows a "reputation" type effect for lenders, intermediation is still valuable. Long term contracts or agreements are not sufficient for the entrepreneurs directly to place their commercial paper.

Thirdly, enduring relationships between banks and borrowers need not preclude loan participations. Frequently, it has been asserted that these long term relationships are like marriage, and that the bank acquires a great deal of information that would be impossible to transfer to another lender. In this model, only the past history of outcomes need be transferred. This should be a lot cheaper than getting a feel for the personality of the CEO or detailed knowledge of the firm's structure and group dynamics. If anything, the sort of long term relationship modelled here makes selling loans easier.

Finally, the model proves suggestive for banking panics and runs against a bank. If the bank looks weak, borrowers may feel that it will not be around in the future to provide punishment, and hence may begin to cheat now. This makes the probability of collapse even higher. Inability to use the less costly monitoring mechanism for loans increases the cost of intermediation and thus has an important direct effect on the economy as a whole.

BIBLIOGRAPHY

Allen, Franklin, "Repeated Principle-Agent Relationships with Lending and Borrowing," Economics Letters 17, 1985.

Breiman, Leo, Probability, Addison-Wesley, Menlo Park, 1968.

Diamond, Douglas W., "Financial Intermediation and Delegated Monitoring," Review of Economic Studies LI, 1984.

Diamond, Douglas W., and Philip Dybvig, "Bank Runs, Deposit Insurance and Liquidity," Journal of Political Economy 91, 1983.

Friedman, James W., Oligopoly and the Theory of Games, North-Holland, New York, 1977.

Haubrich, Joseph G., and Robert G. King, "Banking and Insurance," NBER Working Paper No. 1312 (March 1984).

Klein, Benjamin, and Keith Leffler, "The Role of Market Forces in Assuring Contractual Performance," Journal of Political Economy 89, 1981.

Luce, R. Duncan, and Howard Raiffa, Games and Decisions, Wiley, New York, 1957.

Radner, Roy, "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," Econometrica 49.5, Sept. 1981.

__________, "Repeated Principal-Agent Games with Discounting," Econometrica, 53, 1985.

Rubinstein, Ariel, and Menahem Yaari, "Repeated Insurance Contracts and Moral Hazard," Journal of Economic Theory 30, 1983.

Townsend, Robert M., "Optimal Multiperiod Contracts and the Gain from Enduring Relationships under Private Information," Journal of Political Economy 90, 1982.