Monitoring the Monitor: An Incentive Structure for a Financial Intermediary

Stefan Krasa
Department of Economics

Anne P. Villamil
Department of Economics

Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign
Monitoring the Monitor:
An Incentive Structure for a Financial Intermediary

Stefan Krasa
Anne P. Villamil
Department of Economics

We wish to thank Martin Hellwig, Jeff Lacker, and Steve Williamson for useful comments. We also gratefully acknowledge the financial support of the National Science Foundation (SES 89-09242).

Address of the authors: Department of Economics, University of Illinois, 1206 South Sixth Street, Champaign, IL 61820.
Monitoring the Monitor: An Incentive Structure for a Financial Intermediary

Stefan Krasa       Anne P. Villamil*

First Draft: April 1990  This Draft: April 1991

Abstract

This paper studies financial intermediation (i.e., delegated monitoring) in a costly state verification model. There is a finite number of agents, thus the intermediary cannot fully diversify its portfolio and is subject to default risk. The role of the intermediary is to satisfy simultaneously the different portfolio preferences of borrowers and lenders. Two questions arise when a delegated monitor is subject to non-trivial default risk: (a) What arrangement solves the problem of monitoring the monitor? (b) What intermediary portfolio accomplishes optimal asset transformation between borrowers and lenders? Unlike previous delegated monitoring studies, the law of large numbers is not sufficient to obtain our results. Instead, we appeal to a stronger result, the large deviation principle, which establishes that convergence in the law of large numbers is exponential.

*We wish to thank Martin Hellwig, Jeff Lacker, and Steve Williamson for useful comments. We also gratefully acknowledge the financial support of the National Science Foundation (SES 89-09242).

Address of the authors: Department of Economics, University of Illinois, 1206 South Sixth Street, Champaign IL 61820
1 Introduction

Financial intermediaries provide asset transformation services to agents that are important in many ways. From a theoretical point of view, an economy with financial intermediaries is presumed to be "better" than an economy without intermediaries in the sense that the consumption allocations in the former economy Pareto dominate those in the latter. More practically, inter-mediation appears to have been "revealed preferred" to direct investment as intermediaries account for a non-trivial fraction of GNP in most developed economies (e.g., about 8 percent in the US). Tobin (1963, p. 186) argues that according to the "new view" of monetary economics, the essential function of financial intermediaries is to satisfy simultaneously the portfolio preferences of both borrowers and lenders. He notes:

The reason that the intermediation of financial institutions...can accomplish these [asset] transformations between the nature of the obligations of the borrower and the nature of the asset of the ultimate lender are these: (1) administrative economy and expertise in negotiating, accounting, appraising, and collecting; (2) reduction of risk per dollar of lending by the pooling of independent risks, with respect to both loan default and to deposit withdrawal; (3) governmental guarantees of the liabilities of the institutions and other provisions (bank examination, investment regulations, supervision of insurance companies, last-resort lending) designed to assure the solvency and liquidity of the institutions. For these reasons, intermediation permits borrowers who wish to expand their investments in real assets to be accommodated at lower rates and easier terms than if they had to borrow directly from lenders.

In this paper we study the dominance of intermediated lending over direct lending in a model that operationalizes the simultaneous or "two-sided" (i.e., intermediary-borrower and intermediary-lender) nature of banking emphasized by Tobin and others. The intermediaries that we consider resemble those of Diamond (1984) and Williamson (1986) in the sense that they are "delegated monitors" where one agent monitors all investment projects.¹

¹We use a costly state verification model (like Williamson) with default risk but no
However, our model differs fundamentally from this previous (limit economy) literature because we consider an economy with a finite number of agents and obtain our results for a delegated monitor with a finite sized portfolio. In our model, the intermediary’s default risk is not necessarily perfectly diversified away. The central questions which arise in such an economy are—“who monitors the monitor?” and “what is the structure of optimal contracts that solve the two-sided incentive problem inherent in this environment?”

These questions are not relevant in a limit economy because the probability of bank insolvency and the influence of a bank’s portfolio choice on its solvency disappears. More specifically, it is only in a finite (and in this case more general) economy that both sides of the intermediation problem must be considered. That is, in the finite case the optimal contract must be designed to ensure simultaneously that entrepreneurs report truthfully to the intermediary and that the intermediary reports truthfully to investors. In contrast, in the limit the return on the intermediary’s portfolio converges to the mean of the distribution of returns from loans granted to entrepreneurs (denoted by $\bar{R}$ in our paper). Thus, the intermediary can always guarantee the investors a certain payoff and investors need never worry about monitoring the monitor in the limit. However, when the intermediary contracts with only finitely many entrepreneurs, its portfolio is not perfectly diversified. In this case the intermediary’s asset transformation problem involves not only the choice of a rate of return on deposits ($\bar{R}$), but also effectively involves the non-trivial choice of a risky portfolio (i.e., a distribution of project returns)—which implies directly a particular bankruptcy probability. Further, this bankruptcy probability depends on both the bank’s contracts with entrepreneurs and its contracts with investors.

Deposit withdrawal risk. Villamil (1991) shows that both types of risk can be accommodated in a direct investment problem in a costly state verification model, but our main concern is the monitor’s two-sided incentive problem so we abstract from withdrawal risk. In contrast, Diamond uses a model with non-pecuniary penalties.
We formally specify the model and agents’ optimization problems in Section 2. The paper contains two main results. Theorem 1 in Section 3.1 shows that two-sided simple debt contracts with delegated monitoring dominate direct investment in an economy with non-trivial default risk. Unlike the law of large numbers argument used to obtain results in limit economies, we use the large deviation principle to obtain our results. The large deviation principle shows that convergence in the law of large numbers is exponential. Strictly speaking, this technique is not essential for proving the dominance of delegated monitoring over direct investment. However, in Section 3.2 we show that the large deviation principle is uniquely useful in finite economies because it gives us additional economic insight into the intermediary’s asset transformation problem: It allows us to characterize precisely (for a particular portfolio distribution) how “large” a finite sized intermediary must be to achieve sufficient default risk diversification. This result shows that the increasing returns to scale property that underlies delegated monitoring models (i.e., contracting with an additional firm always lowers costs) need not lead to counterfactual market size predictions—a single intermediary.

Theorem 2 in Section 4.1 shows that simple debt is the optimal contract for both the intermediary-borrower and the intermediary-lender sides of the contracting problem. The intuition behind this result is as follows. The intermediary-lender side of the contract resembles the Gale and Hellwig (1985) and Williamson (1986) problem where simple debt is optimal because it minimizes monitoring costs. However, standard arguments do not apply to the intermediary-borrower side of the contract because the inte-

---

2Simple debt contracts promise holders fixed rates of return when the issuer is solvent and state contingent returns when insolvent. See Section 2.1 for a formal definition and discussion.

3Diamond (1984) and Williamson (1986) use law of large numbers arguments to show the optimality of delegated monitoring when costs are bounded. In Section 3.1 we show that this approach breaks down (so a large deviation argument is essential) when monitoring costs are unbounded (i.e., depend on the size of the intermediary’s portfolio).
termediary (unlike risk neutral investors) must consider the distribution of payments it receives from entrepreneurs. The distribution of payments is important to the intermediary because it affects the delegated monitoring cost borne by investors. Thus, both sides of the problem are interconnected and the structure of the optimal intermediary-borrower contract is not obvious. The main insight in Theorem 2 is that for a sufficiently large intermediary, the marginal change in delegation costs (i.e., monitoring the monitor) is small and is dominated by the marginal change in direct monitoring costs borne by the intermediary. Hence, simple debt is optimal for the intermediary-borrower side of the contract because minimizing the expected costs of monitoring the entrepreneurs remains the main concern—despite the fact that the intermediary must effectively choose both a face value for the debt and a distribution. In Section 4.2 we discuss the relationship between Theorem 2 and banking history.

2 The Model

Consider an economy with finite numbers of two types of risk neutral agents, investors and entrepreneurs. Each entrepreneur $i = 1, \ldots, I$ is endowed with a risky investment project which transforms $y$ units of a single input at time 0 into $\theta_i$ units of output at time 1, but has no endowment of the input. In contrast, each investor $j = 1, \ldots, J$ is endowed with 1 unit of the homogeneous input, but has no direct access to a productive technology. We assume that $y$ is an integer with $y > 1$. Hence, the project of an entrepreneur cannot be financed by a single investor.$^4$ The total available supply of investment is larger than the input required by all entrepreneurs, so $J$ investors can be accommodated by the $I$ entrepreneurs (i.e., $J = yI$). All entrepreneurs and

$^4$This assumption implies that there is duplicative monitoring in the absence of intermediation, because more than one investor must monitor a single entrepreneur. Delegated monitoring economizes on these costs (cf., Diamond (1984)).
investors are symmetrically informed about the distribution of $\theta_i$ at time 0, but asymmetric information exists about the state of the project’s actual realization ex-post: Only entrepreneur $i$ freely observes the realization of $\theta_i$ at time 1, where this realization is denoted by $w$. Let $F(w)$ denote the known distribution of a particular entrepreneur’s project. Finally, suppose that a technology exists which can be used to verify $w$ to non-owners at time 1 with the following characteristics:

(T1) Use of the state verification technology is costly; and
(T2) $w$ is privately revealed only to the individual who requests (deterministic) costly state verification (CSV).\(^5\)

Assumption (T1) is similar with the specification of the CSV technology in Townsend (1979), except that as in Gale and Hellwig (1985, p. 651) the verification cost is comprised of both a pecuniary component and an indirect “pecuniary equivalent” of a non-pecuniary cost. The non-pecuniary costs permit negative utility but rule out negative consumption, while pecuniary equivalents of non-pecuniary costs ensure that the costs can be shared by the contracting parties.\(^6\) In contrast, assumption (T2) differs fundamentally from the specification in Townsend. In his model $w$ is *publicly* announced after CSV occurs, while in our model $w$ is *privately* revealed only to the agent who requests CSV. Assumption (T2) is essential for our analysis since if all information could be made public ex-post, there would be no need to monitor the monitor. However, it also appears to accurately describe the privacy and institutional features which characterize most lending arrangements.\(^7\)

The following assumptions summarize technical aspects of the economy.

---

\(^5\)Stochastic verification is discussed in the concluding section.

\(^6\)These costs may be thought of as the money paid to an attorney to file a claim (a pecuniary cost) and the monetary value of time lost when visiting the attorney (a pecuniary equivalent). Note that Diamond’s costs were unbounded while these costs are fixed.

\(^7\)Diamond (1984, p. 395) observes: “Financial intermediaries in the world monitor much information about their borrowers in enforcing loan covenants, but typically do not directly announce this information or serve an auditor’s function.”
(A1) The $\theta_i$ are independent, identically distributed random variables on the probability space $(\Omega, \mathcal{A}, P).$  

(A2) The distribution $F$ has a continuously differentiable density $f$ with respect to the Lebesgue measure, and $f(x) > 0$ for every $x \in [0, T]$.

(A3) The ex-post verification cost is a fixed constant.

Entrepreneurs have technologies but no input and investors have input but no technologies. Thus, agents must trade to facilitate production. In the remainder of this section we specify contracts which govern trade among agents and the optimization problems from which these contracts can be derived. Formal proofs that the contracts are indeed optimal are deferred to Sections 3 and 4.

2.1 The Direct Investment Problem

Let all direct, bilateral interactions between investors and entrepreneurs be regulated by a contract whose general form is defined as follows.

Definition 1. A one-sided contract between an investor and an entrepreneur is a pair $(R(\cdot), S)$, where $R(\cdot)$ is an integrable, positive payment function on $\mathbb{R}_+$, such that $R(w) \leq w$ for every $w \in \mathbb{R}_+$ and $S$ is an open subset of $\mathbb{R}_+$ which determines the states where monitoring occurs.

As is standard practice in this literature, we restrict the universe of contracts that we consider to the set of incentive compatible contracts. Let $C = (R(\cdot), S)$ denote this set. The following condition ensures that all contracts under consideration satisfy this restriction: There exists $\tilde{R} \in \mathbb{R}_+$ such that $S = \{w: R(w) < \tilde{R}\}$. It is well known that the imposition of this restriction is without loss of generality because any arbitrary contract can be replaced by an incentive compatible contract with the same actual payoff.

---

8 We will always refer to this probability space without mentioning it explicitly, when writing $P$ for probability or $E$ for expected value.
Therefore, the set of all incentive compatible contracts is fully specified by the tuple \((R(\cdot), \bar{R})\). We study a particular type of contract, called a simple debt contract (SDC), which Gale and Hellwig (1985) and Williamson (1986) have shown is the optimal contract among all one-sided investment schemes. We state this result formally in Section 3 as Theorem GHW. This contract is defined as follows:

**Definition 2.** \((R(\cdot), \bar{R})\) is a simple debt contract if: \(R(w) = w\) for \(w \in S = \{w < \bar{R}\}\) and \(R(w) = \bar{R}\) if \(w \in S^c = \{w \geq \bar{R}\}\).

The payment schedules in Definition 2 resemble simple debt because:

(i) When verification does not occur the payment to the investor is constant (i.e., the entrepreneur pays a fixed amount \(\bar{R}\) for all realizations of the state above some cutoff level), where \(S^c\) is the complement of \(S\).

(ii) When verification does occur the payment to the investor is state contingent (i.e., the entrepreneur pays the entire realization for all outcomes below the cutoff), where \(S\) is viewed as the set of bankruptcy states.

We will often denote SDCs by \(\bar{R}\). These contracts arise as optimal (cost-minimizing) responses to asymmetric information problems in economies with costly deterministic state verification technologies. Agents minimize verification costs by verifying only low realizations of \(\theta_i\) and accepting fixed payments (which do not require monitoring) in all other states.

The investor and the entrepreneur’s direct investment problem can now be stated formally:

\[
\max_{(R(\cdot), S) \in C} \int_0^T [w - R(w)] dF(w)
\]

subject to

\[
\frac{1}{J} \int_0^T R(w) dF(w) - \int_{S^c} c dF(w) \geq r. \tag{1}
\]

This “one-sided problem” describes the nature of trade when investors and entrepreneurs contract directly: The expected utility of a representative
entrepreneur is maximized, subject to a constraint that the expected return of a representative investor, net of monitoring costs \( (c) \), be at least as great as some reservation level \( (r) \). Note that \( R(\cdot) \) is the total payment to all investors, and \( I/J \) is the number of investors required to fund a single project. The problem reflects the assumption that credit markets are competitive.\(^9\)

### 2.2 The Delegated Monitoring Problem

We now construct the intermediary’s “two-sided problem.” In the one-sided problem, agents write direct bilateral contracts: each investor must monitor each entrepreneur with whom he/she contracts in bankruptcy states so duplicative monitoring is inherent. In contrast, if investors elect a monitor to perform the verification task, the two-sided arrangement may eliminate some of the duplicative monitoring even though investors must monitor the monitor sometimes. We begin by considering the election of an intermediary: The lending market is competitive, hence it follows that an investor who wishes to act as an intermediary must offer contracts which maximize the expected utility of the entrepreneurs and assure each investor of at least the reservation level of utility. Otherwise, agents would trade directly or another intermediary would offer an alternative contract (i.e., there is free entry into intermediation) with terms that are preferable to the \( I \) entrepreneurs and/or the remaining \( J - 1 \) investors. Let \( (R(\cdot), S) \) denote the entrepreneur-intermediary side of the contract and \( (R^*(\cdot), S^*) \) denote the intermediary-investor side.\(^{10}\) Thus, the intermediary’s problem embodies optimization by all agents in the economy.

\(^9\)We consider an economy in which there are more agents who wish to invest than investment opportunities. Since the supply of loans is inelastic, the level of return necessary to attract investors is driven down to the reservation level \( r \).

\(^{10}\)Recall that \( (R(\cdot), S) \) is also used to denote the entrepreneur-investor contract in the one-sided model. We avoid introducing additional notation at this point because the incentive problem associated with ensuring that entrepreneurs report truthfully is the same regardless of whether they report to the investors directly or to the intermediary.
We next derive the random variables which describe the income from the intermediary’s portfolio. Recall that \( R(w) \) denotes the payoff by an entrepreneur to the intermediary if output \( w \) is realized and \( \theta_i \) is the random variable which describes the output \( w \) of a particular entrepreneur \( i \) in state \( \omega \). Consequently, the intermediary’s income from this entrepreneur, given transfer \( R(\cdot) \), is \( G_i(R(\cdot); \omega) = R(\theta_i(\omega)) \). The random variables \( G_i \) are independent for each choice of \( R(\cdot) \) because the \( \theta_i \) are independent. Assume that the intermediary contracts with \( i = 1, 2, \ldots, I \) entrepreneurs. Then the average income of the intermediary per entrepreneur under payment schedule \( R(\cdot) \), denoted by \( G^I(R(\cdot); \omega) \), is:

\[
G^I(R(\cdot); \omega) = \frac{1}{I} \sum_{i=1}^{I} G_i(R(\cdot); \omega). \tag{2}
\]

Denote the distribution function of \( G^I(\cdot) \) by \( F^I(\cdot) \).

We now specify the intermediary’s cost structure for monitoring the entrepreneurs and the investors’ cost structure for monitoring the monitor (intermediary). Let \( c \) denote the actual fixed cost incurred by the intermediary when it monitors entrepreneur \( i \), and \( c^*_i \) denote the actual cost incurred by an investor when he/she monitors the intermediary with portfolio of size \( I \). The expected monitoring costs are of primary importance to the intermediary and the investors when they make their decisions at time 0 (i.e., \( \int_S c dF(\cdot) \) and \( \int_{S^*} c^*_i dF^I(\cdot) \) respectively). These expected costs depend on three factors: the actual costs (\( c \) and \( c^*_i \)), the relevant states (\( S \) and \( S^* \)), and (in general) on the size of the intermediary (via \( c^*_i \) and \( F^I(\cdot) \)). We impose the following upper boundary on the investors’ costs of monitoring the monitor:

(CS) The costs, \( c^*_i \), do not increase exponentially in \( I \).

Both bounded and most unbounded cost structures satisfy this assumption, hence it is not restrictive.\(^{11}\) Boundedness implies that the marginal

\(^{11}\)Exponentially increasing costs are permissible if they do not increase faster than a bound given by equation (21) in the Appendix, but we regard them as implausible.
cost of the intermediary from contracting with additional firms is decreasing in \( I \), giving it an inherent cost advantage relative to individual investors: The total monitoring costs per depositor are \( \int_S c dF(\cdot) + \int_{S^*} c_i^* dF(\cdot) \). If \( c_i^* \) is fixed (or bounded), then adding an additional firm does not increase costs. However, it decreases the probability of default by the intermediary. This latter effect can be interpreted as the source of the intermediary’s cost advantage in models where costs are bounded from above. Of course, such cost advantages may exist (as Tobin suggests in (1) in the Introduction), but we view unbounded costs as an interesting benchmark case. One unbounded cost example that satisfies assumption CS is \( c_i^* = cI \), where state verification by investors involves verifying the full state (i.e., each of the \( I \) projects in the intermediary’s portfolio). The institutional structure in this example is such that intermediary failures are quite costly. However, if investors can economize further on these costs (e.g., by monitoring only insolvent firms), then delegated monitoring will be even more attractive.

The two-sided contract between the intermediary and each entrepreneur, and the intermediary and the investors, can now be defined.

**Definition 3.** A two-sided contract is a four-tuple \( ((R(\cdot), S), (R^*(\cdot), S^*)) \) having the following properties:

(i) \( R(\cdot) \) is an integrable positive payment function from an entrepreneur to the intermediary such that \( R(w) \leq w \) for every \( w \in \mathbb{R}_+ \), and \( S \) is an open subset of \( \mathbb{R}_+ \) which determines the set of all realizations of an entrepreneur’s project where the intermediary must monitor;

(ii) \( R^*(\cdot) \) is an integrable positive payment function from the intermediary to the investors such that \( R^*(w) \leq w \). For every realization \( w \) of \( G^I(\cdot) \), the payment to an individual investor is given by \( \frac{1}{J - 1} R^*(w) \); and \( S^* \) is

---

\(^{12}\)Recall that \( G^I(\cdot) \) is the average income per entrepreneur defined by equation (2).

\(^{13}\)\( R^*(\cdot) \) is the total payment by the intermediary to investors per entrepreneur. To derive the payment to an individual investor, multiply this amount with \( \frac{1}{J - 1} \).
an open subset of $\mathcal{M}_+$ which determines the set of all realizations of the intermediary’s income from the entrepreneurs the investors must monitor.

We now derive the set of all incentive compatible two-sided contracts. As before, we restrict our analysis to this set without loss of generality. Each entrepreneur will announce an output which minimizes its payment obligations to the intermediary. Let $\bar{w} = \arg \min_{x \in S} R(x)$ be the output that minimizes this payoff over all non-monitoring states, and recall that $w$ is observed directly in the monitoring states $S$. Consequently, the announcement by an entrepreneur is given by $\arg \min_{x \in \{w, \bar{w}\}} R(x)$. A similar condition holds for the intermediary-investor portion of the contract (i.e., $R^*(\cdot), \bar{R}^*$). As in the one-sided problem, the following condition ensures that all contracts are incentive compatible. There exist $\bar{R}, \bar{R}^* \in \mathcal{M}_+$ such that $S = \{w: R(w) < \bar{R}\}$ and $S^* = \{w: R^*(w) < \bar{R}^*\}$. The set of all incentive compatible two-sided contracts is fully specified by the four-tuple $(R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)$.

Finally, a two-sided simple debt contract can now be defined:

**Definition 4.** A contract $(R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)$ is a two-sided simple debt contract (denoted by $(\bar{R}, \bar{R}^*)$) if:

(i) $R(w) = w$ for $w \in S = \{w < \bar{R}\}$ and $R(w) = \bar{R}$ if $w \in S^c = \{w \geq \bar{R}\}$;

(ii) $R^*(w) = w$ for $w \in S^* = \{w < \bar{R}^*\}$ and $R^*(w) = \bar{R}^*$ if $w \in S^{*c} = \{w \geq \bar{R}^*\}$.

We will often denote two-sided simple debt contracts by $(\bar{R}, \bar{R}^*)$.

The intermediary’s two-sided optimization problem can now be stated:

$$\max_{(R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)} \int_0^T [w - R(w)] dF(w)$$

subject to

$$\frac{I}{J-1} \int_0^T R^*(w) dF^I(R(\cdot), \bar{R})(w) - \int_{S^*} c^I dF^I(R(\cdot), \bar{R})(w) \geq r$$

(3)

$$I \left[ \int_0^T [w - R^*(w)] dF^I(R(\cdot), \bar{R})(w) - \int_{S^*} c dF(w) \right] \geq r.$$  

(4)
This problem states that the intermediary maximizes the expected utility of each ex-ante identical entrepreneur subject to two constraints. (3) states that the expected payoff to the \( J - 1 \) remaining investors (i.e., those who did not become intermediaries) must be at least \( r \), the reservation level of utility. (4) states that the profit from intermediation (i.e., net payoffs from the entrepreneurs less the payoff to the investors) must also be at least \( r \).

3 Delegated Monitoring

In this section we first review the direct investment problem (Theorem GHW) solved by Gale and Hellwig (1985) and Williamson (1986). In Section 3.1, the first result of our paper (Theorem 1) establishes that two-sided simple debt with delegated monitoring dominates one-sided direct investment if there are sufficiently many entrepreneurs.\(^{14}\) The key problem is that our finite delegated monitor is not completely diversified (i.e., cannot guarantee investors a riskless payoff). Further, the (constant but default risky) payoff offered by the intermediary depends in a non-trivial way on its investment contracts with entrepreneurs. Thus, the bank not only chooses a payoff \( \bar{R}^* \), but also (implicitly) chooses a distribution \( dF^I \) by choosing \( \bar{R} \). Because the costs of monitoring the monitor may depend on the size of the bank’s portfolio, the proof of Theorem 1 relies on the large deviation principle. This principle, and its relation to the law of large numbers, is discussed. In the proof, we refer to certain mathematical results proved in the Appendix. In Section 3.2 we compute (for two particular portfolio distributions) how “large” a finite sized intermediary must be to achieve sufficient default risk diversification.

Theorem GHW. Simple Debt is the optimal contract among all one-sided investment schemes.

\(^{14}\)We prove the optimality of two-sided simple debt contracts in Section 4.
The strategy of the proof is as follows: Consider two optimal contracts: Let $\bar{R}$ be a simple debt contract and $(A(\cdot), \bar{A})$ be some alternative contract. Since both contracts are optimal, both must yield the same expected payoff to entrepreneurs. With the first contract investors request verification if $w < \bar{R}$. In the alternative contract, verification occurs in all states $w$ such that $w < \bar{A}$. Since $\bar{A} > \bar{R}$ (otherwise the contracts cannot have the same return to entrepreneurs) the expected verification costs must be less for the simple debt contract $\bar{R}$. We will refer to this result in the proof of Theorem 1.

### 3.1 Optimality of Intermediation

We now prove that intermediation is optimal.

**Theorem 1.** Two-sided simple debt contracts with delegated monitoring strictly dominate one-sided direct investment if there are sufficiently many entrepreneurs.

**Proof.** Our arguments depend on the continuity of the constraints, established by Lemma 1 in the Appendix, and can be summarized as follows. Let $\bar{R}$ be the simple debt contract which is optimal among all one-sided schemes described by Theorem GHW. We show that there exists an $\bar{R}^*$ such that (3) is binding for the two-sided contract $(\bar{R}, \bar{R}^*)$, and that by increasing $\bar{R}^*$ the payoff to investors increases and (3) does not bind. We next show that (4) is fulfilled but not binding. Hence increasing $\bar{R}^*$ slightly makes both constraints slack if the number of entrepreneurs is sufficiently large. This proves the Theorem since by lowering $\bar{R}$ we can make the entrepreneurs better off than in the one-sided scheme.

---

15See Gale and Hellwig (1985) or Williamson (1986) for a formal proof.

16In general, the investors' payoff does not increase monotonically with $\bar{R}^*$ because the probability that investors must monitor the intermediary is an increasing function of $\bar{R}^*$. This is also true for one-sided schemes (cf., Gale and Hellwig (1985, p. 662)).

14
We begin by showing that the costs of monitoring the monitor go to zero if $\bar{R}^*$ is less than the intermediary’s expected return from one entrepreneur, $\bar{R}^* < E[G_i(R(\cdot))]$, and if the intermediary is sufficiently large. This means:

$$\lim_{I \to -\infty} \int c_i^* dF^I(w) = 0 \text{ for every } \bar{R}^* < E[G_i(R(\cdot))].$$

(5) follows immediately from (21) of Lemma 4 in the Appendix, which establishes that the probability of a default by the delegated monitor (i.e., the probability that an investor must monitor the monitor), converges to zero exponentially. However, by assumption (CS) the monitoring costs $c_i^*$ do not increase exponentially. The key insight of the proof is that, in expected terms, the costs go to zero.

The Theorem is proved as follows:

(i) The law of large numbers implies that

$$\lim_{I \to -\infty} \int R^*(w) dF^I(R(\cdot))(w) = \bar{R}^* \text{ for every } \bar{R}^* < E[G_i(R(\cdot))].$$

(6) indicates that the probability of a default by the delegated monitor goes to zero. Hence investors get the face value of the SDC with certainty.

(ii) (5) and (6) imply that we can find a two-sided contract $(\bar{R}, \bar{A}^*)$ such that (3) is fulfilled for sufficiently large $I$ but not binding. Because of the continuity of (3) with respect to $\bar{R}$ and $\bar{A}^*$ (Lemma 1), there must exist a face value $\tilde{R}^* < \bar{A}^*$ such that (3) is binding for the two-sided contract $(\bar{R}, \tilde{R}^*)$. By construction, increasing $\tilde{R}^*$ slightly implies that the investors’ payoff increases.

(iii) All that remains to show is that (4) is also fulfilled, but not binding. Because of (5), it follows that $\int_S c dF > \int_{S^*} c_i^* dF^I$ for all sufficiently large $I$. This and the fact that (3) is binding implies

$$I \int_0^T R^*(w) dF^I < (J - 1)(r + \int_S c dF).$$

(7) Consequently,

$$I \left[ \int_0^T [w - R^*(w)] dF^I - \int_S c dF \right]$$
\[ > IE[G_i(\bar{R})] - J \int_S c_dF - (J - 1)r \geq Jr - (J - 1)r = r, \]

where the first inequality follows from (7) and the second inequality follows from the fact that \( \bar{R} \) must fulfill constraint (1) of the one-sided problem by assumption. This proves Theorem 1.

In the proof of Theorem 1 we use the large deviation principle. This principle is related to the law of large numbers but is stronger as the large deviation principle shows that convergence in the law of large numbers is exponential (see Lemma 4 in the Appendix). The large deviation principle is important in our analysis for two reasons. First, it is essential in establishing (5) in the proof when monitoring costs depend on the size of the intermediary’s portfolio. Second, regardless of the monitoring cost structure, it allows us to characterize (for a particular distribution) how “large” a finite sized intermediary must be to achieve sufficient default risk diversification. The economic intuition behind the problem addressed by Theorem 1 is that monitoring costs may increase in the number of entrepreneurs (i.e., as the size of the intermediary gets large, the verification costs, \( c^* \), in the bankruptcy state become large as well). However, if the probability of default goes to zero “fast enough,” then the expected value of the costs of monitoring the monitor become insignificant for a well diversified intermediary – even though the intermediary is of finite size and hence is not perfectly diversified. The role of the large deviation principle is to provide a convergence result that is “fast enough” (i.e., faster than that provided by the law of large numbers) to generate gains from intermediation. In Section 3.2 we provide examples which show that sufficient diversification is achieved by remarkably small intermediaries.

Theorem 1 establishes that two-sided arrangements are better than one-sided direct investment. However, the following problem remains: If (3) and (4) are not binding it is not clear who gets the surplus. Thus, the
maximization problem is not well defined. The following Proposition shows that this difficulty does not arise and that optimal contracts exist.

**Proposition 1.** If there are sufficiently many entrepreneurs, then there exist optimal contracts among the set of all two-sided simple debt contracts. The two constraints from the intermediary’s optimization problem bind for all optimal contracts.

**Proof.** The result follows directly from the continuity results of Lemma 1 in the Appendix. The existence of optimal contracts is straightforward since according to Lemma 1 both the constraints and the argument we are maximizing over are continuous functions of $\bar{R}$ and $\bar{R}^*$. To show that both constraints must bind for optimal contracts, consider the following cases.

(i) Suppose by way of contradiction that at an optimum both constraints do not bind. Then $\bar{R}$ can be reduced slightly without violating the constraints. This, however, contradicts the optimality of the contract, so it is not possible that both constraints are slack at an optimum.

(ii) Suppose that only (3) does not bind. Then $\bar{R}^*$ can be reduced slightly without violating the constraint. This reduces the total payment of the intermediary to the investors, but (4) no longer binds and the above argument can be applied to get a contradiction.

(iii) Suppose that only (4) does not bind. We must show that (3) no longer binds if $\bar{R}^*$ is increased by a small amount. This is not straightforward since by increasing $\bar{R}^*$ we increase the expected cost of monitoring the monitor. However, (6) establishes that this expected monitoring cost goes to zero as the size of the intermediary increases. Further, it follows from (4) that $\bar{R}^*$ remains bounded away from $EG_i(\bar{R})$ as $I$ increases.\(^{17}\) Thus, the expected cost of monitoring the monitor remains close to zero (i.e., changes very little) if we increase $\bar{R}^*$ slightly. If only (4) does not bind, both constraints must bind for an optimal contract.

\(^{17}\)Divide both sides of (4) by $I$ and take the limit for $I \to \infty$. 

17
bind the gain from a higher payment to investors exceeds the loss from an increase in monitoring expenditures and (3) does not bind, a contradiction. Hence, both constraints must bind at an optimum.

3.2 Optimality of “Small” Intermediaries

In this Section we show by way of two simple (but not pathological) parametric examples that small intermediaries can support efficient investment arrangements. That is, we compute the rate function given by the “large deviation principle” from Lemma 4 in the Appendix for two alternative distributions of the intermediary’s portfolio. Our results show that intermediation works surprisingly well—even for “small” intermediaries (those with investment portfolios of only 30 ex-ante identical but independent firms).

Both examples have the following characteristics. From Lemma 4 in the Appendix, the rate function $\mathcal{I}(x)$ is given by

$$\mathcal{I}(x) = \sup_{\xi \in \mathcal{R}} x \xi - \log(M(\xi)),$$

where $M(\xi)$ is the moment-generating function, and $x$ is the face value of the intermediary-investor part of the simple debt contract, i.e., $R^*$. Assume that $x = 0.3$ and that the intermediary-firm part of the contract is $\bar{R} = 1$. We consider two different (discrete) firm project return distributions under $\bar{R}$ (i.e., two different distributions of $G_1(\bar{R}; w)$), and hence two different moment-generating functions, with a common expected value of 0.5. In both examples the intermediary contracts with 30 entrepreneurs with independent project returns. Following equations (20) and (21) in the Appendix, the probability of a default by the delegated monitor is given by $e^{-30\mathcal{I}(x)}$. The cost of monitoring an entrepreneur is $c = 0.3$. We now compute (using the intermediary’s optimization problem from Section 2.2) bounds on the cost of monitoring the monitor ($c_f^*$) for which intermediation is optimal for two
different return distributions.\footnote{We compute monitoring cost boundaries for $c$ and $c^*_i$ for which intermediation is optimal when the intermediary has no initial endowment and two investors are needed to finance one project. We omit the intermediary's endowment to simplify computations, but note that all constraints will be less binding if it has an endowment. Constraint (3) is:

\[ \frac{1}{2} \int_0^T R^*(w) dF^I(R(\cdot), \tilde{R})(w) - \int_{S^*} c dF^I(R(\cdot), \tilde{R})(w) \geq r; \]

where $r = \frac{1}{2} \int R(w) dF - \int_S c dF$ because each investor's reservation value is the return available from direct investment. Let $d^*_i$ denote the probability of default by a bank of size $I$, let $d$ denote the probability of default by an entrepreneur, and let $E(R(\cdot))$ denote the expected return from an entrepreneur. Then (3) is fulfilled if

\[ 1/2(1 - d^*_i) \tilde{R}^* - c^*_i d^*_i \geq 1/2 E(R(\cdot)) - dc. \] (F1)

(F1) gives for every $c$ a value of $c^*_i$ under which investors get their reservation utility.

Constraint (4) is:

\[ \int_0^T \left[ w - R^*(w) \right] dF^I(R(\cdot), \tilde{R})(w) - \int_S c dF(w) \geq 0. \]

By substituting (3) into (4), and using the above notation it follows that (4) is fulfilled if

\[ c^*_i d^*_i \leq c. \] (F2)

All computations in Examples 1 and 2 are based on formulae F1 and F2.

\footnote{This is $e^{-0.0822n}$, where $n$ is the number of firms (cf., Lemma 4 and (20) and (21)).}

\textbf{Example 1.} Assume that each firm's discrete project return distribution is as follows: There is a bad state where the return from investment is 0 and a good state where the output is 1, and both states are equally likely. The moment generating function is given by

\[ M(\xi) = \frac{1}{2} + \frac{1}{2} e^\xi. \]

Following Varadhan (1984, p. 9) the rate function $I(x)$ is given by

\[ I(x) = \begin{cases} \log 2 + x \log x + (1 - x) \log(1 - x) & \text{for } 0 \leq x \leq 1; \\ \infty & \text{otherwise.} \end{cases} \]

Thus, $I(0.3) = 0.0822$. When there are 30 firms, the probability of default by a bank using the two-sided simple debt contract $\tilde{R} = 1, R^* = 0.3$ is 0.0847.\footnote{This is $e^{-0.0822n}$, where $n$ is the number of firms (cf., Lemma 4 and (20) and (21)).}
We now compute values for \( c \) and \( c^*_i \) for which intermediation dominates direct investment for this small intermediary that contracts with 30 firms when two investors are needed to finance one project (a worst case scenario assumption). Recall that \( c = 0.3 \). Then as long as \( c^*_i \leq 1.4674c \), investors get their reservation utility (i.e., (3) is satisfied). Further, the intermediary's profit is non-negative if \( c^*_i \leq 11.8046c \) (i.e., (4) is satisfied). Note that the costs of monitoring the monitor, \( c^*_i \), are almost the same as the cost of monitoring an entrepreneur, so the gains from delegated monitoring are not very great for this distribution, given the assumed parametric specifications.

**Example 2.** Now assume that each firm's (less risky) discrete project return distribution is as follows: The values 0 and 1 occur with probability 1/4 each, and the value 1/2 occurs with probability 1/2. Recall that the rate function \( I(x) \) is given by

\[
I(x) = \sup_{\xi \in \mathbb{R}} x\xi - \log(M(\xi)),
\]

where the moment-generating function in this case is

\[
M(\xi) = 1/4 + \frac{1}{2}e^{\xi/2} + \frac{1}{4}e^\xi.
\]

Recall that \( \bar{R}^* = x = 0.3 \). Thus, \( I(0.3) = 0.16457 \).\(^{20}\) When there are 30 firms, the probability of default by a bank using the two-sided simple debt contract in this parametric example is 0.007176. Recall that \( c = 0.3 \). As long as \( c^*_i \leq 57.5616c \) investors get their reservation utility. The intermediary's profit is non-negative if \( c^*_i \leq 139.3478c \). Thus, intermediation strongly dominates direct investment in this example—even when bankruptcy by the intermediary requires every investor to monitor all projects, i.e., \( c^*_i = cI \) (the case discussed in Section 2.2).

The first example can be viewed as a "worst case scenario" in the sense that this distribution is very risky since all the mass is in the tails. Never-

\(^{20}\)The supremum is reached at \( \xi = -1.6946 \).
theless, with only 30 ex-ante identical firms the probability of insolvency is only 0.0847. In other words, even if the delegated monitor invests in only a relatively small number of firms (30), diversification (and hence delegated monitoring) works well. In the second example where the distribution we start with is less risky, intermediation works even better. With the same number of firms the probability of default is only 0.0007176. These examples show that our results hold not only for some arbitrarily large but finite case, but for surprisingly small intermediaries.

The result that “small” intermediaries can support efficient intermediation arrangements is of particular economic interest in a delegated monitoring context. Inherent in delegated monitoring models of intermediation is an increasing returns to scale phenomenon: When the intermediary contracts with an additional firm, expected monitoring costs are always lowered. This fact has led some to argue that the market structure implied by delegated monitoring theories is counterfactual, i.e., a monopoly market structure. In fact, these examples suggest that this critique has little real content. If diversification works very fast (i.e., goes to the mean exponentially), then the addition of additional firms becomes negligible very quickly as well. Consequently, many “small” but finite intermediaries are likely to attain sufficient risk diversification simultaneously.21

21 Strictly speaking, it is possible to partition an infinite set of entrepreneurs into an infinite number of subsets, where each infinite subset of entrepreneurs contracts with a particular delegated monitor to obtain a model with multiple intermediaries. However, the “sufficient diversification of small portfolios” argument seems more intuitively plausible (to us) as a means of obtaining multiple intermediaries than arguments which depend on partitions of an infinite set of entrepreneurs (and the consequent absence of default risk). The multiple intermediaries that emerge in our model are single agents, rather than multiple agents who form an intermediary coalition as in Boyd and Prescott (1986).
4 Optimality of Two-Sided Simple Debt

In this Section we prove that two-sided simple debt contracts solve the two-sided monitoring problem inherent in an economy with non-trivial default risk. Section 4.1 contains the result (Theorem 2), and Section 4.2 contains a discussion of its economic implications. We begin by stating a Corollary that is essential for establishing the Theorem, and provide an example that illustrates the main economic problem addressed by the proof. The proof that two-sided simple debt is optimal requires a stronger result than Lemma 4 (in the Appendix) used in the proof of Theorem 1. That is, in Theorem 1 we used Lemma 4 to show that the probabilities of default converged exponentially to zero. In Theorem 2 we use Lemma 5 (in the Appendix) to show that the densities also converge exponentially to zero. From this result we get the following Corollary which establishes that the difference in probability between the realization being below \( x_1 \) and \( x_2 \), respectively, is bounded by the absolute value of the difference between \( x_1 \) and \( x_2 \) times a term which converges exponentially to zero.

**Corollary 1.** Let \( \bar{R} > 0 \) and \( z < EG_i(\bar{R}) \). Then there exist \( a > 0 \) and \( \bar{I} > 0 \) such that \( |P(G^I(\bar{R}) \leq x_1) - P(G^I(\bar{R}) \leq x_2)| \leq e^{-aI}|x_1 - x_2| \) for every \( x_1, x_2 \leq z \), for every \( I \geq \bar{I} \), and for every \( \bar{R} \geq \bar{R} \).

The main problem in the proof of Theorem 2 is that two-sided simple debt contracts do not necessarily minimize the expected costs of monitoring the monitor.\(^{22}\) Theorem GHW shows that simple debt contracts are optimal for

\(^{22}\)Consider the following example: To simplify computations we use a discrete distribution, but the example can be extended to a continuous distribution by simple approximation arguments. Assume that there are two entrepreneurs \( i = 1, 2 \), and that the realization of \( \theta_i \) is 0 with probability 0.4; and 1 and 2 each with probability 0.3. Let \((\bar{R}, \bar{R}^*)\) be a simple debt contract with \( \bar{R} = 1 \) and \( \bar{R}^* = 0.7 \) (recall that \( \bar{R}^* \) is the total payment by the intermediary to the investors per firm). Let \((A(\cdot), \bar{R}^*)\) be an alternative contract such that \( A(0) = A(1) = 0 \) and \( A(2) = 2 \). The investor-intermediary part of the contract is the same in both cases, and both contracts yield the same expected return to the entrepreneurs.
the one-sided problem (any other contracts generate higher expected costs for monitoring the entrepreneurs). In the two-sided case we minimize the sum of the expected costs of monitoring the entrepreneurs and of monitoring the intermediary. However, because the second summand need not be minimal for two-sided simple debt, the main idea of the proof is to show that the one-sided and two-sided problems are essentially the same for large intermediaries. By Corollary 1 changes in the intermediary-entrepreneur part of the contract have a very small effect on investors' payoffs. Thus, minimizing the expected costs of monitoring the entrepreneurs is the main concern.

4.1 Optimality of Two-sided Simple Debt Contracts

We now prove that two-sided simple debt is optimal.

**Theorem 2.** If there are sufficiently many entrepreneurs then the optimal contracts are two-sided simple debt contracts. Two-sided simple debt contracts strictly dominate all other types of contracts.

**Proof.** We proceed by contradiction. Assume without loss of generality that there exists some alternative two-sided contract $(A_I(\cdot), A_1^*(\cdot))$, for every $I$, which improves upon the optimal two-sided simple debt contract of Theorem 1. By Theorem 1, we can restrict our analysis to two-sided contracts. We show:

(i) The investor's part of contract $A_1^*(\cdot)$ must be a simple debt contract, $\tilde{A}_1^*$.

Next we choose a two-sided simple debt contract $(\tilde{R}_I, \tilde{R}_1^*)$ such that entrepreneurs have the same expected return and the expected payments from the intermediary to the investors remain constant.\(^{23}\) Furthermore, from Lemma 3

\(^{23}\)First choose $\tilde{R}_I$ such that $EG^I(\tilde{R}_I) = EG^I(A_I(\cdot))$. Because of the continuity results of Lemma 1, a SDC, $\tilde{R}_1^*$, can be chosen such that $\int R_1^*(w) \, dF(R(\cdot)) = \int A_1^*(w) \, dF(A(\cdot))$. However, the probability of default by the intermediary is lower with the second contract. Specifically, it is 0.49 for the alternative contract and 0.64 for the simple debt contract.
in the Appendix, \( \tilde{A}_I^* \geq \tilde{R}_I^* \). The contracts \((\tilde{R}_I, \tilde{R}_I^*)\) must fulfill the conditions of Corollary 1 for all sufficiently large \( I \) (i.e., there exists \( \tilde{R} > 0 \) and \( z < EG^I(\tilde{R}) \) such that \( \tilde{R}_I^* \leq z < EG^I(\tilde{R}) \leq \tilde{R}_I \) for all sufficiently large \( I \)).

We show:

(ii) Using \((\tilde{R}_I, \tilde{R}_I^*)\) instead of \((A_I(\cdot), \tilde{A}_I^*)\), the left hand side of (3) decreases by at most \( c_I^*e^{-a_I} |\tilde{A}_I - \tilde{R}_I| \). The left hand side of (4) increases by at least \( Icm|\tilde{A}_I - \tilde{R}_I| \), where \( m = \min_{x \in [0,T]} f(x) \), because with simple debt contracts the intermediary must monitor entrepreneurs in fewer states of nature. This is essentially the two-sided analog of Theorem GHW.

(iii) Since for large \( I \) the surplus is much greater than the loss, we can show that it is possible to distribute some of the intermediary’s gain to the investors by increasing the face value of \( \tilde{R}_I^* \) such that both constraints are satisfied and \textit{not} binding. Hence, the face value of \( \tilde{R}_I \) can be lowered such that both constraints still hold. The entrepreneurs are better off with a contract with a lower face value. Consequently, the two-sided simple debt contract \((\tilde{R}_I, \tilde{R}_I^*)\) dominates \((A_I(\cdot), \tilde{A}_I^*)\), which provides the contradiction. Therefore all optimal contracts must be simple debt contracts.

We now prove claim (i). This follows immediately from Theorem 1 be-

\[ \text{This holds since by Lemma 3, } \int A_I^*(w) dF(R(\cdot)) \geq \int A_I^*(w) dF(A(\cdot)). \]  

\[ \text{Thus, to get equality we must choose } \tilde{R}_I \leq \tilde{A}_I^*. \]

By the (indirect) assumption of the proof, \((A_I(\cdot), \tilde{A}_I^*)\) dominates the optimal two-sided SDCs of Theorem 1. This is only possible if the probability that investors must monitor the intermediary goes to zero as \( I \) gets large. Otherwise the expected monitoring costs would go to infinity. Thus, in the limit investors receive the face value \( \tilde{A}_I^* \) with certainty, i.e., \( \lim_{I \to \infty} \int A_I^*(w) dF(A(\cdot)) = \lim_{I \to \infty} \tilde{A}_I^* \). Clearly, the costs of monitoring an individual entrepreneur \( \int s \cdot dF(w) \) remain bounded away from zero as \( I \to \infty \). Dividing both sides of (4) by \( I \) and taking the limit we conclude that \( \lim_{I \to \infty} A_I^* < \lim_{I \to \infty} EG^I(A(\cdot)) \), so there exist \( z, z' \) such that \( \tilde{A}_I^* \leq z < z' \leq EG^I(A(\cdot)) \) for all sufficiently large \( I \). Since \( \tilde{R}_I^* \leq \tilde{A}_I^* \) by Lemma 3 (cf., previous footnote) and \( EG^I(R_I(\cdot)) = EG^I(A_I(\cdot)) \) by definition, it follows that \( \tilde{R}_I^* \leq z < z' \leq EG^I(R_I(\cdot)) \). Now choose \( \tilde{R} \) such that \( z < \tilde{R} \leq R_I \) for all sufficiently large \( I \). This is exactly the condition of Corollary 1.

The two-sided SDC does not necessarily minimize the probability of default by the intermediary. Thus the left hand side of (3) may be smaller under contract \((A_I(\cdot), \tilde{A}_I^*)\).

\[ m > 0 \text{ by assumption (A2).} \]
cause the intermediary is like an entrepreneur whose production is described by the random variable $G^I(\cdot)$.

Next we prove claim (ii). In order to compute the change of the left hand side of (3) we need only compute the change in expected monitoring costs (because the first integral on the left-hand side of (3) does not change by construction of $(\tilde{R}, \tilde{R}^*)$). In the following, let $\tilde{A}_I$ be the simple debt contract with the same face value as $A_I(\cdot)$. Observe that:

\[
\int_{S_{\tilde{R}_I^*}} c^*_i dF(\tilde{R}_I) - \int_{S_{\tilde{A}_I^*}} c^*_i dF(\tilde{A}_I(\cdot)) \leq \int_{S_{\tilde{R}_I^*}} c^*_i dF(\tilde{R}_I) - \int_{S_{\tilde{R}_I^*}} c^*_i dF(\tilde{A}_I). \tag{8}
\]

This inequality follows from two factors: First, the income of the intermediary from contract $\tilde{A}_I$ is higher than from contract $A_I(\cdot)$ in all states, hence less monitoring occurs; and second $\tilde{A}_I^* \geq \tilde{R}_I^*$.

Clearly, the difference in payoff from an individual entrepreneur to the intermediary between the two SDCs with face values $\tilde{A}_I$ and $\tilde{R}_I$ is at most $\tilde{A}_I - \tilde{R}_I$.\(^{28}\) Therefore

\[
G^I(\tilde{R}_I) = \frac{1}{I} \sum_{i=1}^{I} G_i(\tilde{R}_I) \geq \frac{1}{I} \sum_{i=1}^{I} [G_i(\tilde{A}_I) - (\tilde{A}_I - \tilde{R}_I)] = G^I(\tilde{A}) - (\tilde{A}_I - \tilde{R}_I). \tag{9}
\]

Hence (9) implies,

\[
P\{G^I(\tilde{A}_I) \leq \tilde{R}_I^*\} \geq P\{G^I(\tilde{R}_I) \leq \tilde{R}_I^* - (\tilde{A}_I - \tilde{R}_I)\}. \tag{10}
\]

From (8), (10) and Corollary 1 it now follows that

\[
\int_{S_{\tilde{R}_I^*}} c^*_i dF(\tilde{R}_I) - \int_{S_{\tilde{A}_I^*}} c^*_i dF(A_I(\cdot)) \leq c^*_i e^{-aI}(\tilde{A}_I - \tilde{R}_I). \tag{11}
\]

The intermediary’s loss (i.e., the decrease of the left hand side of (4)) can be computed using the main idea of Theorem 1: If agents use contract $A_I(\cdot)$ instead of $\tilde{R}_I$, the intermediary must monitor in additional states $w \in \underline{28} \tilde{A}_I - \tilde{R}_I > 0$ since both contracts have the same expected return but $\tilde{R}_I$ is a SDC (hence it has the lowest face value among all contracts with the same expected return).
[\bar{\bar{R}}_I, \bar{\bar{A}}_I]. \text{ Hence, expected monitoring costs increase by } \int_{\bar{\bar{R}}_I} \bar{A}_I \geq cm(\bar{A}_I - \bar{\bar{R}}_I), \text{ and the total loss is at least } Icm|\bar{A}_I - \bar{\bar{R}}_I|. \text{ This proves (ii).}

Finally we prove (iii). Let } \varepsilon > 0. \text{ From footnote 19 we have } R^*_I < z < EG^I(\bar{R}) \leq \bar{\bar{R}}_I. \text{ Hence, by the law of large numbers there exist } \bar{\bar{h}} > 0 \text{ and } \bar{T} \text{ such that}

\begin{equation}
P(\{G^I(\bar{R}_I) \geq \bar{R}^*_I + \bar{\bar{h}}\}) > 1 - \varepsilon, \text{ for all } I \geq \bar{T}.
\end{equation}

(12) and Corollary 1 imply that by increasing the face value of \(R^*_I\) by \(h < \bar{\bar{h}}\), the payoff to investors (i.e., the left hand side of (3)) increases by at least \(\frac{I}{I-h}(1-\varepsilon) - \bar{\bar{h}}e^{-aI}\). If \(I\) is sufficiently large, this amount is bounded below by \(\frac{I}{2}h(1-2\varepsilon)\). Again because of (12) the intermediary’s profit (i.e., the left hand side of (4)) is decreased by at most \(Ih(1-\varepsilon)\).

By choosing \(h = \frac{I}{I-h}c^Ie^{-aI}(\bar{A}_I - \bar{\bar{R}}_I)\) constraint (3) is fulfilled and not binding (by the computations in the previous paragraph). Given this \(h\), the profit of the intermediary decreases by at most \(I \frac{1-e^{-aI}}{1-2\varepsilon}c^Ie^{-aI}(\bar{A}_I - \bar{\bar{R}}_I)\). Comparing this to the total gain, which is \(Icm\), it is clear that an \(I\) can be chosen independently of \(A_I\), such that the gain is greater than the loss. \(^{29}\) This means that constraint (4) is not binding as well, which proves the Theorem.

### 4.2 Monitoring With Two-sided Debt Contracts

Throughout the paper we have described simple debt contracts by their two-sided payoffs \((\bar{R}, \bar{R}^*)\). However, simple debt contracts also implicitly characterize two-sided monitoring intervals \((S, S^*)\). The key insight in understanding optimal financial intermediation (i.e., delegated monitoring) in environments with non-trivial default risk is that both the intermediary-investor and the intermediary-entrepreneur sides of the problem are interconnected. That is, unlike in the limit case, the entrepreneurs’ contract with the intermediary affects investors’ payoffs and monitoring costs. Thus, the cost, risk

\(^{29}\)Clearly this is the case if \(\frac{I}{I-2\varepsilon}c^Ie^{-aI} < cm\) which holds for all sufficiently large \(I\) and is independent of \(A_I\) and \(\bar{\bar{R}}_I\).
reduction, and monitoring aspects of intermediation described by Tobin in the Introduction (i.e., reasons (1), (2), and (3)) are all related. Theorem 2 shows that if there is sufficient diversification, then two-sided simple debt contracts solve the monitoring problem (i.e., (3) in Tobin’s explanation) because if sufficient risk reduction is achieved then aggregate monitoring costs are minimized (i.e., reasons (1) and (2)). Thus, the “sufficient diversification” in the proofs of Theorems 1 and 2 in essence describes the imperfect asset transformation services provided by banks.

Our results in Sections 3 and 4 have two interesting features: First, our Examples in Section 3.2 show that significant asset transformation can be achieved by surprisingly small intermediaries. Second, Theorem 2 shows that the monitoring problem is solved by private contractual arrangements among agents. These results depend on the assumption that the probability distributions that describe firms’ project returns are independent. This dependence suggests that restrictions imposed on firms by regulators which constrain their ability to contract with firms whose project returns are not correlated severely undermine the efficacy of such private contractual arrangements. Branching restrictions are a prime example of such restrictions, as returns within a restricted geographical area are unlikely to be independent.\textsuperscript{30} Many authors have argued that the reason the Canadian banking system has been more stable than the US banking system is because Canadian banks were able to branch freely while US banks were heavily restricted (e.g., see Myers (1931), (1970)). It is important to remember, however, that the pooling of independent risks is only one (albeit important) aspect of the intermediation problem. As our Examples show, the structure of the underlying portfolio distribution, deposit and loan rates, monitoring costs, and portfolio size also jointly determine the outcome.

\textsuperscript{30}See Boyd and Smith (1990) for an environment with different locations and an explicit analysis of the interlocational flow of funds. Their analysis is conducted in a model with no aggregate default risk.
5 Concluding Remarks

In this paper we show that delegated monitoring dominates direct investment and that two-sided simple debt is optimal in a costly state verification model with non-trivial default risk. Our economic environment requires us to introduce new mathematical arguments based on the large deviation principle. This mathematical technique is useful for three reasons: (i) it allows us to compute the actual size of a sufficiently well diversified "small but finite" intermediary; (ii) it is essential for establishing the optimality of delegated monitoring when monitoring costs are unbounded; and (iii) it is essential for establishing the optimality of two-sided debt (regardless of the monitoring cost structure) because uniform convergence of the densities does not follow from the law of large numbers.

Recently, costly state verification studies (cf., Townsend (1988) or Mookherjee and Png (1989)) have shown that the form of the optimal contract may be altered under stochastic monitoring. In contrast, our delegated monitoring result is not affected by stochastic monitoring. This follows from the fact that as in the deterministic case, the probability that a state occurs in the stochastic case which triggers monitoring goes to zero exponentially. Hence the expected cost of monitoring the monitor goes to zero as well, and delegated monitoring continues to dominate direct investment. Our rationale for studying deterministic monitoring is similar to that given for the monitoring cost example given in Section 2.2. We wish to establish gains from delegated monitoring which stem from the intermediary's ability to eliminate duplicative verification costs for a benchmark case. If other factors exist which further reduce monitoring costs (e.g., stochastic monitoring), intermediation will be even more attractive.
6 Appendix

The following notation will be useful in the analysis that follows. Given some two-sided contract, let $\Gamma_1(\cdot)$ denote the aggregate payoff of the $J-1$ investors from the intermediary who interacts with $I$ entrepreneurs, and $\gamma_1$ denote the expected payoff of a single investor from the intermediary. For the two-sided SDC with delegated monitoring, $(R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)$ these payoffs are:

$$\Gamma_I((R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)) = \int_0^T R^*(w) dF^I(R(\cdot), \bar{R})(w); \quad (13)$$

$$\gamma_I((R(\cdot), \bar{R}), (R^*(\cdot), \bar{R}^*)) = \frac{I}{J-1} [\Gamma_I(\cdot)] - \int_S c^* dF^I(\cdot). \quad (14)$$

The following Lemma establishes that the above payoff functions are continuous in the face values $\bar{R}$ and $\bar{R}^*$ for SDCs. This proves continuity of constraint (3). We also prove results which are necessary to get continuity of constraint (4) and of the argument of the two-sided optimization problem.

**Lemma 1.** Let $R(w)$ be one-sided simple debt with face value $\bar{R}$. Then the functions $\bar{R} \mapsto \int R(w) dF(w)$ and $\bar{R} \mapsto \int_{w<\bar{R}} c dF(w)$ are continuous. Furthermore, $(\bar{R}, \bar{R}^* ) \mapsto \Gamma_I(\bar{R}, \bar{R}^*)$ is continuous; and $(\bar{R}, \bar{R}^* ) \mapsto \gamma_I(\bar{R}, \bar{R}^*)$ is continuous at every $(\bar{R}, \bar{R}^*) \in \mathbb{R}^2$, such that $\bar{R}^* < \bar{R}$.

**Proof.** The proof for the first two functions is straightforward. We now prove continuity of $\Gamma_I$. Note that,

$$P(\{G^I(\bar{R}) \geq \bar{R}^*\}) \leq P(\{G^I(\bar{R} + h) \geq \bar{R}^*\}) \leq P(\{G^I(\bar{R}) \geq \bar{R}^* - h\}), \quad (15)$$

for $h > 0$ because of (10). Therefore $\int f(x) dF^I(\bar{R} + h)(x) \leq \int f(x + h) dF(\bar{R})(x)$, for every increasing step-function $f$, and hence for every arbitrary increasing function $f$ by a standard approximation argument. Further, $\|A(\cdot) - R(\cdot)\|_\infty \leq |\bar{A} - \bar{R}|$, for all simple debt contracts $\bar{A}$ and $\bar{R}$. Thus,

$$|\Gamma_I(\bar{R} + h, \bar{R}^* + h^*) - \Gamma_I(\bar{R}, \bar{R}^*)| \leq |h| + |h^*|, \quad (16)$$
for every \( h, h^* > 0 \) and hence for every \( h, h^* \in \mathbb{R} \). This proves continuity of \( \Gamma_I \). It now remains to prove that \((\tilde{R}, \tilde{R}^*) \mapsto \int_{-\infty}^{\tilde{R}^*} c^* dF(\tilde{R})\) is continuous at every \((\tilde{R}, \tilde{R}^*) \in \mathbb{R}\) such that \( \tilde{R}^* < \tilde{R} \). The distribution of \( G^I(\tilde{R}) \) has a density which is bounded by a \( K_n \in \mathbb{R} \) in \((-\infty, \tilde{R})\) (cf., Lemma 4). Therefore (15) implies \( |\int_{-\infty}^{\tilde{R}^*} c^* dF^I(\tilde{R}) - \int_{-\infty}^{\tilde{R}^*} c^* dF^I(\tilde{R} + h)| \leq K_n |h| \). Since this inequality holds for every \( \tilde{R}^* < \tilde{R} \) and since \( \tilde{R}^* \mapsto \int_{-\infty}^{\tilde{R}^*} c^* dF^I(\tilde{R}) \) is clearly continuous, this proves continuity of \( \gamma_I \).

To prove optimality of two-sided simple debt we need an additional technical Lemma which shows that two-sided simple debt contracts maximize total payments to investors (not including monitoring costs). The result is formally stated in Lemma 3. It follows immediately from the next Lemma.

**Lemma 2.** Let \( \mu \) be a probability measure on \([0, M]\). Let \( R(\cdot) \) and \( A(\cdot) \) be two contracts with the same expected value. We assume that \( R(\cdot) \) is a simple debt contract \( \bar{R} \). Then the simple debt contract is less risky than \( A(\cdot) \) in the sense of Rothschild and Stiglitz; i.e., \( \int u(A(w))d\mu(w) \leq \int u(R(w))d\mu(w) \), for all concave functions \( u \).

If we interpret \( u \) as a utility function, then Lemma 2 essentially says that all risk averse consumers prefer simple debt contract to any other contract with the same expected value. This is one of the criteria for comparing the riskiness of two distribution of Rothschild and Stiglitz (1970). By using Theorem 2 of their paper which proves the equivalence of three different concepts, and by using their argument on p. 230 ff, the proof of the Lemma is straightforward. We now give the proof.

**Proof of Lemma 2.** Let \( H_A \) and \( H_R \) be the cumulative density functions of the distributions of \( A(\cdot) \) and \( R(\cdot) \), respectively. Let \( G(t) = H_A(t) - H_R(t) \), and let \( T(y) = \int_0^y G(x)dx \). Then \( A(\cdot) \) is more risky than \( R(\cdot) \) if the following
two conditions are fulfilled:

\[(17) \quad T(M) = 0, \quad \text{and} \quad (18) \quad T(y) \geq 0 \quad \text{for every } y \in [0, M].\]

Let \( g \) be the density of \( G \). Partial integration yields \( T(M) = \int_0^M G(x) \, dx = xG(x)|_0^M - \int_0^M xg(x) \, dx = 0 \), since the integral over \( xg(x) \) is the expected value of \( A(\cdot) - R(\cdot) \). This proves (17). Note that \( G(t) \geq 0 \) for every \( 0 \leq t < \bar{R} \), and \( G(t) \leq 0 \) for every \( \bar{R} \leq t \leq M \). This together with (17) proves (18). Theorem 2 of Rothschild and Stiglitz (1970) implies that \( A(\cdot) \) is more risky than \( R(\cdot) \). This proves the Lemma.

Lemma 3. Let \((\bar{R}, \bar{R}^*)\) be a two-sided simple debt contract and \((A(\cdot), \bar{R}^*)\) be an alternative contract where \( EG^I(A(\cdot)) = EG^I(\bar{R}) \). Then \( \Gamma_I(\bar{R}, \bar{R}^*) \geq \Gamma_I(A(\cdot), \bar{R}^*) \).

Proof. Note that

\[
\Gamma_I(R(\cdot), R^*(\cdot)) = \int \cdots \int R^*(\frac{1}{I} \sum_{i=1}^I R(w_i)) \, dF(w_1)dF(w_2)\cdots dF(w_I). \quad (19)
\]

Since \( R^*(\cdot) \) is a concave function we can apply Lemma 2 and inductively substitute \( R(\cdot) \) by \( A(\cdot) \) in (19). This proves the Lemma.

Next we show that convergence in the law of large number is exponential. This result is used in Section 3 to establish gains from delegated monitoring. The proof follows immediately from the large deviation principle.

Lemma 4. Let \((X_i)_{i \in N}\) be a sequence of independent identically distributed random variables with values in \([0, M]\) and distribution \( \mu \). Let \( \mu_n \) be the distribution of \( \frac{1}{n} \sum_{i=1}^n X_i \). Then \( \mu_n([0, b]) \) converges to zero exponentially as \( n \to \infty \) for every \( b < EX_i \).
Proof. $M(\xi) = \int e^{\xi x} d\mu(x) < \infty$ for every $\xi \in \mathbb{R}$, since $\mu$ has a compact support. Let $\mathcal{I}(x) = \sup_{\xi \in \mathbb{R}} (\xi x - \log M(\xi))$. Then Cramér's theorem (Varadhan (1984), Theorem 3.1) implies that $\mathcal{I}$ is a "rate function", and that $\mu_n$ satisfies the large deviation principle with rate $\mathcal{I}$. Hence,

$$1/n \log \mu_n([0, b]) \leq -\mathcal{I}(b),$$

(20) for every $b < EX$. It therefore remains to prove that $\mathcal{I}(b) > 0$ for every $b < EX$.

For fixed $b$ let $H(\xi) = \xi b - \log M(\xi)$. Since $H(0) = 0$, it is sufficient to show that $H'(0) \neq 0$. This, however, can be easily verified:

$$H'(\xi) = b - \frac{\int x e^{\xi x} f(x) dx}{\int e^{\xi x} f(x) dx}$$

consequently $H'(0) = b - EX \neq 0$. This and (20) immediately imply that

$$\mu_n([0, b]) \leq e^{-\mathcal{I}(b)n}, \text{ for every } b < EX.$$  

(21)

This proves the Lemma.

For the optimality of two-sided simple debt contracts proved in Section 4 we need a stronger result than Lemma 4. In particular, we must show that the densities of $\mu_n$ also converge to zero exponentially on every interval $[0, b]$ where $b < EX$.

Lemma 5. Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of independent, identically distributed random variables with values in $[0, M]$ and distribution $\mu$. We assume that $\mu = \bar{\mu} + \lambda \delta_R$ where supp $\bar{\mu} \subset [0, R]$ and $\bar{\mu}$ has a density $f$ with respect to the Lebesgue measure on $\mathbb{R}$. We assume that $f$ is continuously differentiable on $[0, M]$. Let $\mu_n$ be the distribution of $\frac{1}{n} \sum_{i=1}^{n} X_i$, and let $\delta_R$ be the Dirac point measure. Then $\mu_n = \bar{\mu}_n + \lambda_n \delta_R$ where supp $\bar{\mu}_n \subset [0, R]$. $\bar{\mu}_n$ has a density $f_n$ with respect to the Lebesgue measure which converges uniformly to zero as $n \to \infty$ on all compact subsets of $[0, EX_i)$. This convergence is exponential.

31This is the form of the rate function for intervals $[0, b]$, where $b < EX_i$ (see Varadhan (1984, p. 8)).
The idea of the proof of Lemma 5 is as follows. We first show that the derivative of the density functions of $G^I(R(\cdot))$ is bounded by a polynomial term (24). This is straightforward if the densities are continuously differentiable with compact support in $IR$, see for example Floret (1981) Exercise 14.24. In our case the densities are discontinuous at 0 and $R$ which requires us to deal with the derivatives at these two points separately. The idea of the proof, however, is essentially the same. If the Lemma does not hold then this immediately implies that the measure of an interval $[0, b]$ can only converge to zero with polynomial speed (25). However, by Lemma 4 the probability of every such interval with $b < G^I(R(\cdot))$ converges to zero with exponential speed as $I \to \infty$ (21). This contradiction proves Lemma 5.

**Proof:** We first derive an upper boundary for the derivative of $f_n$. The main problem is that $f$ is discontinuous at 0 and $R$ as a function defined on $IR$.

**Claim 1:** Let $u$, $v$ be functions on $IR$ with support in $[0, T]$. Let $u$ be continuous from the right and bounded. Let $v$ be continuously differentiable in $[0, T]$. Then the convolution $u \ast v$ is continuous from the right and \( |(u \ast v)'|_\infty \leq |u|_1 |v'|_\infty + 2 |u|_\infty |v|_\infty. \)

Let $z(t, x, h) = u(t) \frac{1}{h} [v(x + h - t) - v(x - t)]$. Furthermore let $B_{1,h} = \{ t: -h \leq x - t \leq 0 \}$ and let $B_{2,h} = \{ t: T - h \leq t - x \leq T \}$. Then

\[
\int_{B_{1,h}} z(t, x, h) \, dt = h \int_0^1 z(x + th, x, h) \, dt = \int_0^1 u(x + th) v(h - th) \, dt.
\]

Consequently $\lim_{h \to 0} \int_{B_{1,h}} z(t, x, h) \, dt = u(x)v(0)$, since $u$ and $v$ are differentiable from the right. A similar argument yields $\lim_{h \to 0} \int_{B_{2,h}} z(t, x, h) \, dt = u(x)v(T)$. Therefore

\[
(u \ast v)'(x) = \lim_{h \to 0} \int z(t, x, h) \, dt = \int u(t)v'(x - t) \, dt + u(x)v(0) + u(x)v(T),
\]

which proves claim 1.

33
Claim 2: The distribution of $\sum_{i=1}^{n} X_i$ has a density $g_n$ on $[0, nR)$. The point $\{nR\}$ has probability $\lambda^n$. The right-hand derivative $g'_n(x)$ exists for every $n \in \mathbb{N}$ and for every $x \in [0, nR)$. Further, there exists a constant $K \leq 2\|f'\|_\infty + 4\|f\|^2_\infty$ such that $|g'_n(x)| \leq n^2K$ for every $n \in \mathbb{N}$ and for every $x \in [0, nR)$.

For the proof of claim 2 we proceed by induction. For $n = 1$ there is nothing to prove. We now assume that the inequality holds for $n - 1$. Using the formula for the density of the sum of two independent random variables and accounting for the “point mass” $\lambda$ at $R$ and $\lambda^{n-1}$ at $(n - 1)R$ we get

$$g_n(x) = \int g_{n-1}(t)f(x-t)\,dt + \lambda^{n-1}f(x-(n-1)R) + \lambda g_{n-1}(x-R), \quad (22)$$

for every $x \in [0, nR]$. (22) implies

$$\|g_n\|_\infty \leq \|g_{n-1}\|_1\|f\|_\infty + \|f\|_\infty + \|g_{n-1}\|_\infty \leq 2\|f\|_\infty + \|g_{n-1}\|_\infty.$$ 

Consequently

$$\|g_n\|_\infty \leq 2n\|f\|_\infty. \quad (23)$$

Using (22), (23) and claim 1 yields

$$\|g'_n\|_\infty \leq \|g_{n-1}\|_1\|f'\|_\infty + 4n\|f\|^2_\infty + \lambda^{n-1}\|f'\|_\infty + \lambda\|g'_{n-1}\|_\infty.$$ 

By substituting the induction hypothesis for $K = 2\|f'\|_\infty + 4\|f\|^2_\infty$, we conclude the proof of claim 2.

Note that $f_n(x) = \int_0^x g_n(t)\,dt$ is the density of $\mu_n$ with respect to the Lebesgue measure. We get

$$|f'_n(x)| \leq n^4K, \text{ for every } n \in \mathbb{N}, \text{ and for every } x \in [0, R). \quad (24)$$

In the second part of the proof we proceed indirectly. Assume that there exists a $b < EX_i$ such that $f_n$ does not converge uniformly to zero on $[0, b]$ with exponential speed. That means that there exist a sequence $\varepsilon_n$ that does
not converge exponentially to zero and a sequence \((y_n)_{n \in \mathbb{N}}\) in \([0, b]\) such that \(f_n(y_n) \geq \varepsilon\) for every \(n \in \mathbb{N}\). We prove that under these circumstances, (24) implies that \(\mu_n([0, b])\) cannot converge to zero exponentially which contradicts Lemma 4.

Let

\[
h_n(x) = \begin{cases} 
\varepsilon - n^4 K(x - y_n), & \text{if } y_n \leq x \leq y_n + \frac{x}{n^4 K}; \\
0, & \text{otherwise.}
\end{cases}
\]

Let \(b' \in (b, \text{EX}_i)\). Then there exists an integer \(\bar{n}\) such that \(y_n + \frac{x}{n^4 K} \leq b'\) for every \(n \geq \bar{n}\). Consequently \(0 \leq h_n(x) \leq f_n(x)\) for every \(n \geq \bar{n}\), because of (24). Therefore

\[
\mu_n([0, b']) = \int_0^{b'} f_n(x) \, dx \geq \int_0^{b'} h_n(x) \, dx = \frac{\varepsilon^2}{2n^4 K}.
\]  

(25)

By (21) and (25) we get \(\frac{\varepsilon^2}{2n^4 K} \leq e^{-an}\) for every \(n \geq \bar{n}\), a contradiction. This proves the Lemma.

We are now ready to prove Corollary 1 from Section 4.

**Proof of Corollary 1.** The distribution of \(\theta_i\) has by assumption a density \(f\) which is continuously differentiable on \([0, M]\). The densities of \(F^1(\bar{R})\) are then given by \(f_\bar{R} = f|_{[0, \bar{R}]} + \delta_\bar{R}\), so the constant \(K\) of (24) can be chosen independently of \(\bar{R}\). Thus inequality (25) holds uniformly for every \(\bar{R}\). Further, \(P(G^n(\bar{R}) \leq x) \leq P(G^n(\bar{R}) \leq x)\) for every \(x \in \mathbb{R}\). Thus, (21) holds uniformly for every distribution \(\mu_n\) of \(G^n(\bar{R})\), where \(\bar{R} \geq \bar{R}\), so Lemma 5 also holds uniformly for every \(\bar{R} \geq \bar{R}\). This proves the Corollary.
References:


J. Tobin (1963), "Commercial Banks as Creators of Money," Richard D. Irwin, Homewood, IL.


