

IOWA STATE UNIVERSITY

Equilibrium Lending Mechanism and Aggregate Activity

Cheng Wang, Ruilin Zhou

September 2004

Working Paper # 04019

Department of Economics Working Papers Series

Ames, Iowa 50011

Iowa State University does not discriminate on the basis of race, color, age, national origin, sexual orientation, sex, marital status, disability or status as a U.S. Vietnam Era Veteran. Any persons having inquiries concerning this may contact the Director of Equal Opportunity and Diversity, 3680 Beardshear Hall, 515-294-7612.

Equilibrium Lending Mechanism and Aggregate Activity*

Cheng Wang
Iowa State University

and

Ruilin Zhou
Penn State University

September, 2004

Abstract

What determines the firm's choice of its mechanism of investment financing? How is the choice of the firm's financing mechanism at the micro level related to the economy's business cycle movements at the aggregate level? This paper develops a model of the credit market where the equilibrium lending mechanism, as well as the economy's aggregate investment and output, are endogenously determined. Among other things, our model predicts that a negative productivity shock can cause an economic downturn that is accompanied not only by a contraction in total outstanding loans, but also by a decline in the ratio of bank loans to non-bank lending, as observed in the 1990-91 U.S. recession.

J.E.L Classification: E44.

*Cheng Wang: Department of Economics, Iowa State University, Ames, IA 50011. chewang@iastate.edu. Ruilin Zhou: Department of Economics, Penn State University, University Park, PA 16802. rzhou@psu.edu. We thank Ed Green, Jeff Lacker, Rafael Repullo for helpful comments.

1. Introduction

What determines the firm's choice of its mechanism of investment financing? How is the choice of the firm's financing mechanism at the micro level related to investment and output at the aggregate level? These questions are often at the center of the discussions with respect to the relationship between investment financing and business cycle movements, including the recent discussions on the nature and role of the so-called "credit crunch" which occurred during the 1990-91 U.S. recession. During this recession, the economy's total outstanding loans fell dramatically, and more importantly, the fraction of intermediated loans fell dramatically relative to unintermediated loans, including public bonds and commercial paper (Friedman and Kuttner 1993). Wright (1995) shows that in the U.S. data, bank loans are highly pro-cyclical, while bond and commercial paper issues are quite counter-cyclical. Kashyap, Stein and Wilcox (1993) find that following a tightening of monetary policy, while there usually is a sharp increase in the amount of commercial paper outstanding, bank loans fall. They argue that monetary contraction tightens the supply of bank credit and hence forces borrowers to switch to commercial paper. They further view this as evidence of the existence of a loan supply channel of monetary policy transmission. Bernanke, Gertler and Gilchrist (1996) advocate a flight to quality view on the same subject. They postulate both that the demand for short-term credit is counter-cyclical, and that firms differ in their degree of access to credit markets. Thus, during a recession, high-grade firms borrow relatively easily by way of issuing commercial paper, while low-grade firms, who can only borrow from banks, are constrained.

With respect to the 1990-91 U.S. recession and credit crunch, several additional hypotheses have been proposed on what caused the crunch. Some argue that the 1988 Basle Accord, which mandates higher capital backing for risky loans, forced banks to switch from commercial loans to government securities. But evidence on this hypothesis is mixed (e.g., Berger and Udell 1994, Haubrich and Wachtel 1993). For example, Berger and Udell (1994) conclude that there is not a significant empirical link between the new regulation and the crunch. They point out that similar credit declines occurred during the 1990-91 U.S. recession for other financial institutions that were not subject to the new rules of the Basle accord. Wagster (1999) supports the conclusion of Berger and Udell with evidence from an international data set. Finally, even if the regulatory explanation does prove to have bearings on the particular 1990-91 crunch, it clearly cannot explain the more general observation that bank loans are pro-cyclical and commercial paper issues counter-cyclical.

The purpose of this paper is to develop a model of the credit market where the equilibrium lending mechanism, as well as the economy's aggregate investment and output, are endogenously determined. We then use the model to show, among other things, that a negative technology shock can cause an economic downturn that is accompanied not only by a contraction in total outstanding loans, but also by a decline in the ratio of bank loans to non-bank lending, as observed

in the 1990-91 U.S. recession, and consistent with the observed pattern that bank loans are procyclical, while bond and commercial paper issues are counter-cyclical. That is, we argue that the observed relationship between financing and recession may just be part of the equilibrium responses of the credit market to a real technology shock.

Our model is built on a two-stage lender-borrower contracting problem that features adverse selection, costly state verification (monitoring) and moral hazard. In the model, after an investment project is funded initially, the entrepreneur (borrower) observes a random signal $\theta \in [0, 1]$ that indicates the project's potential success rate. This signal is private to the entrepreneur unless the investor (lender) pays a fixed cost to monitor. The project can then be terminated or fully implemented. In the latter case, the entrepreneur must make an unobservable effort to carry out the rest of the investment process. The final output of the project then depends stochastically on θ and the entrepreneur's effort. In the model, if the optimal contract involves a positive probability of monitoring the entrepreneur, we brand the optimal contract as a form of intermediated bank lending; and, if the optimal contract involves no monitoring at all, we classify the optimal contract as unintermediated market financing. This interpretation of the model is essential for our purposes. In practice, some business enterprises seek financing from financial intermediaries, while others borrow directly from the credit market (e.g., commercial paper, corporate bond). A key distinction between the two financing mechanisms is that financial intermediaries often engage in extensive monitoring during the process of financing, whereas typical individual lenders do not monitor, or do so much less. A theoretical explanation for this distinction is that monitoring of private information is more efficient when it is delegated to a financial intermediary, rather than when done repetitively by individual lenders (Diamond 1984).¹

We then embed the optimal lending contract in a competitive credit market environment to study, through comparative statics, how the relationship between the equilibrium financing mechanism and the aggregate output responds to disturbances to the model's exogenous variables. We show that in our model, a sufficiently strong negative technology shock can always cause a recession that is accompanied by a switch of the equilibrium lending mechanism from bank loans to direct lending. A negative technology shock is interpreted in our model as a decrease in the potential returns of capital investment (the output of a successful project), H .

In the model, the investor's optimal monitoring policy depends critically on the trade-off between termination (i.e., the choice of the set of realizations of θ with which the project is terminated) and monitoring that the investor uses as incentive devices. The optimal trade-off between termination and monitoring in turn depends on how cost-effective each device is relative

¹The idea that banks are delegated monitors is central to the models of financial intermediation based on costly state verification (e.g., Williamson 1986, 1987). Recent studies on the choice of the optimal financing mechanism by Diamond (1991) and Holmstrom and Tirole (1997) have also taken seriously the notion that bank financing is closely related to monitoring. In both papers, financial intermediaries are modeled as monitors who can detect the borrowers choosing a bad project.

to the other. As H falls, termination becomes less costly, as the projects are worth less if they are continued. On the other hand, as H falls, the costs to the investor of using monitoring as an incentive device are not directly affected. Therefore, as H falls, termination becomes a more efficient incentive device relative to monitoring. In particular, we show that a sufficiently low H will drive the demand for monitoring to zero: monitoring will not be used at all. This is true even if random monitoring is permitted. In other words, a negative technology shock weakens the case for bank loans, leading to a shift of the equilibrium financing mechanism from intermediated bank loans to unintermediated market lending. Meanwhile, as more projects are terminated, total outstanding loans fall, the success rates of the fully implemented projects will be higher, and the economy's aggregate output will fall.

This paper builds on the large literature in contract theory that follows Townsend (1979) in modeling the role of costly state verification (monitoring) in optimal financial arrangements, including Gale and Hellwig (1985), Williamson (1986, 1987), and Boyd and Smith (1997). Monitoring in a slightly different sense is also an essential ingredient in recent models of direct market financing versus intermediated lending. Holmstrom and Tirole (1997) and Repullo and Suarez (1995) focus on the importance of the firm's net worth in determining its financing choice. In both papers, bank monitoring of the firm's choice of investment project is a partial substitute for collateral, mitigating the effects of moral hazard. Diamond (1991) studies the role of the interaction between monitoring and reputation for determining the firm's optimal financing mechanism. We focus instead on the trade-off between monitoring and termination in determining the optimal choice between bank loans and market lending.

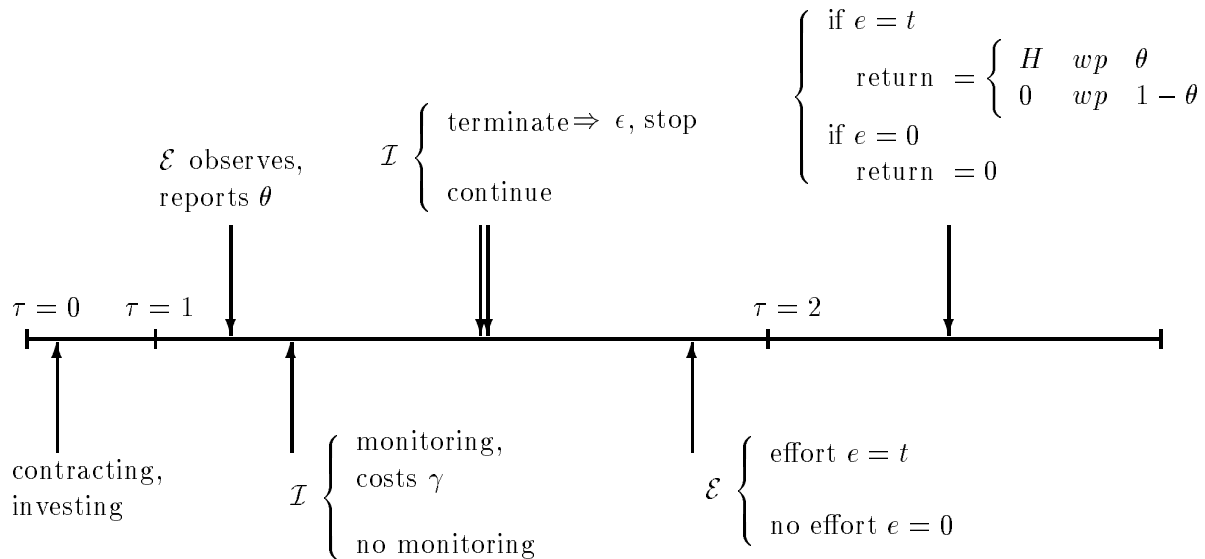
Section 2 presents the model. In Section 3, we study the problem of optimal contracting between an investor and an entrepreneur. Section 4 embeds the optimal contract in a competitive credit market framework. Section 5 concludes the paper.

2. The Model

There are three periods, $\tau = 0, 1, 2$. There are two types of agents, investors and entrepreneurs. The measure of the investors is λ , and the measure of the entrepreneurs is δ . Investors and entrepreneurs all maximize the expected value of $u(c, e) = c - e$, where c is consumption and e is effort. We assume that there is no discounting across periods.

Figure 1 depicts the timeline of events, where \mathcal{I} denotes the investor and \mathcal{E} denotes the entrepreneur. In period 0, each investor has one indivisible unit of the investment good of the model, which can either go to an entrepreneur in exchange for a financial contract to be described later, or earn a constant return equal to one unit of consumption in period 2 through storage. Each investor also has access to $\xi (> 0)$ units of the consumption good in period 2. We will assume that ξ is large enough in a sense that will become clear after we lay out the financial contract.

Figure 1. The timeline of events



At $\tau = 0$, each entrepreneur owns a risky investment project. The entrepreneur has no initial wealth, and hence must rely on external financing for the project. The project requires an investment of one unit of the investment good in period 0, or it simply perishes. If the project is funded, then at the beginning of period 1 the entrepreneur observes a signal θ . Here $\theta \in [0, 1] \equiv \Theta$ is a random variable that represents the state of the project. The precise meaning of θ will be given shortly, but we assume that θ is a continuous random variable on Θ with a distribution function $G(\theta)$ and a density $g(\theta)$. Assume for all $\theta \in \Theta$, $g(\theta) > 0$. The signal θ is directly observable only to the entrepreneur. The investor can observe the realization of θ through a costly monitoring process, which requires $\gamma \geq 0$ units of the investor's effort.

The investor may or may not monitor, but a decision must be made in period 1 whether to continue the project. If the investor monitors the entrepreneur's report of θ , then the termination/continuation decision can be based on the true value of θ . Otherwise, the termination/continuation decision may take into account only the entrepreneur's report of the realization of θ . The termination value of the project is $\epsilon \geq 0$. Suppose the investment is continued. Then the entrepreneur must make an effort $t > 0$ to finally complete the project. The entrepreneur's effort is not observable to other parties. In other words, there is moral hazard.

If the entrepreneur makes the required effort and the project is completed, then with probability θ the project succeeds, in which case the return is $H > 0$, and with probability $1 - \theta$ it fails

and the return is 0. If the entrepreneur does not make the required effort, then the project yields a return of 0 with probability one. In the following, we call H the potential return of the project, and θ the project's success rate.

We make some further assumptions. First, all payments to the entrepreneur must be non-negative (limited liability). Second, renegotiations are not allowed. In other words, we assume that once the contract is signed, both parties can fully commit, at all stages of the investment process, to the terms of the initial contract. Third, we assume that, after the initial investment and before the interim termination/continuation decision, the potential return of the project H is higher than the opportunity cost had the project been terminated; that is,

Assumption (1) $t + \epsilon + \gamma \leq H$.

Finally, we make a technical assumption to guarantee the uniqueness of solution to the optimal contracting problem in the sections to follow.

Assumption (2)

$$\frac{-H}{H - (t + \epsilon)} \leq \frac{g'(\theta)}{g(\theta)} \leq \frac{H}{t + \epsilon + \gamma}.$$

Clearly, there is a wide range of distribution functions with support $[0, 1]$ that satisfy this condition, including the uniform distribution function.

We now describe the credit market. In period 0, there is a credit market in which investors offer lending contracts to entrepreneurs who exchange investment opportunities for credit and compensation. Given that all projects are identical ex ante, in the credit market equilibrium that we will define later in the paper, each contract promises expected utility equal to u_0^* to the entrepreneur who sells his project in the credit market. This equilibrium expected utility of the entrepreneur will be determined endogenously. But in order to characterize an equilibrium, we must first determine the form of the optimal contract that maximizes the investor's expected utility, given that it pays the entrepreneur a minimum of some u_0 in expected utility.

3. Optimal Contracting

In this section, we determine the optimal contract between a representative investor and a representative entrepreneur, taking as given that, in equilibrium, contracts must promise the entrepreneur expected utility of no less than $u_0 \geq 0$.

3.1. The First-Best Contract

Consider the case where both the realization of θ and the entrepreneur effort t are publicly observable. In this case, a contract must specify a fixed monetary compensation x to the entrepreneur, and, in addition, a termination/continuation policy Φ . Here Φ is a subset of Θ : if

$\theta \in \Phi$, then the project is continued; otherwise, it is terminated. Given the environment, the optimal contract must implement t as the entrepreneur's effort.

Let Φ' denote the complement of the set Φ . The problem of optimal contracting can then be formulated as follows.

$$(P0) \quad \max_{x, \Phi} \int_{\Phi} \theta H dG(\theta) + \int_{\Phi'} \epsilon dG(\theta) - x \quad (1)$$

subject to

$$x - \int_{\Phi} t dG(\theta) \geq u_0. \quad (2)$$

The objective function (1) represents the investor's expected payoff.² Condition (2) is the entrepreneur's participation constraint. Clearly, constraint (2) must be binding, since otherwise the value of x can be reduced to improve the investor's expected payoff without violating the participation constraint. By substituting constraint (2) into the objective function, we can rewrite the optimal contracting problem as

$$\max_{\Phi} \int_{\Phi} (\theta H - t) dG(\theta) + \int_{\Phi'} \epsilon dG(\theta) - u_0. \quad (3)$$

Obviously, the optimal Φ must be an upper interval of Θ .³ Let this interval be $[\theta_{fb}, 1]$, $\theta_{fb} = \operatorname{argmax}_{x \in [0,1]} F(x)$, where

$$F(x) \equiv \int_x^1 (\theta H - t) dG(\theta) + \epsilon G(x).$$

It can be shown that function $F(x)$ is strictly concave under assumption (2),⁴ and therefore the maximization problem (3) has a unique solution:

$$\Phi_{fb} = [\theta_{fb}, 1], \quad \text{where} \quad \theta_{fb} = \frac{t + \epsilon}{H}. \quad (4)$$

Given Φ_{fb} , the entrepreneur's compensation is determined by $x = u_0 + (1 - G(\theta_{fb}))t$, and the investor's expected payoff (expected net return on an investment) is given by

$$V_{fb}(u_0) = H \int_{\theta_{fb}}^1 (\theta - \theta_{fb}) dG(\theta) + \epsilon - u_0 - 1. \quad (5)$$

²For convenience, we omit the constant unit cost of date-0 investment in all of the objective functions.

³If a project with a lower success rate θ is continued, then a project with a higher θ should also be continued.

⁴We have $F'(x) = -[xH - (t + \epsilon)]g(x)$. Obviously, $F'(0) > 0$, $F'(1) < 0$. Also, $F''(x) = -[xH - (t + \epsilon)]g'(x) - Hg(x)$. So for the function $F(x)$ to be concave, we need

$$-\frac{g'(x)}{g(x)} [xH - (t + \epsilon)] < H.$$

Now if $xH - (t + \epsilon) \geq 0$, then the above inequality certainly holds. If $xH - (t + \epsilon) < 0$, then the concavity condition becomes

$$\frac{g'(x)}{g(x)} < \frac{H}{(t + \epsilon) - xH},$$

which holds, by the second inequality of assumption (2).

Note that given that both parties are risk neutral, it is straightforward to show that the following holds. The first-best outcome is attainable when there is only moral hazard concerning the entrepreneur's effort, but no information asymmetry with respect to the project's success rate θ . The first-best outcome is also attainable if there is only asymmetric information concerning the success rate θ , but there is no moral hazard.⁵

3.2. Contracting with Costly Monitoring and Moral Hazard

Now we consider our original problem, where the realization of the project's success rate θ is directly observable only to the entrepreneur. The investor can observe the value of θ at a cost $\gamma > 0$. Moreover, there is moral hazard: the entrepreneur's effort is not observable.

3.2.1. The Definition of contract

With costly monitoring and moral hazard, there are now three components to a loan contract: (i) a monitoring policy M for verifying the state of the success rate θ , (ii) a termination/continuation policy Φ which determines whether the project is terminated after the realization of the state θ , and (iii) a scheme for state contingent compensations to the entrepreneur. Formally, a contract takes the following form:

$$\sigma = \left\{ M; \Phi; x; y(\tilde{\theta}), \tilde{\theta} \notin \Phi; R_0(\tilde{\theta}), R(\tilde{\theta}), \tilde{\theta} \in \Phi \right\}.$$

For now, we abstract from stochastic monitoring. Thus, the monitoring policy M is a subset of Θ in which verification of the reported state will occur. That is, let $\hat{\theta}$ denote the entrepreneur's report of θ , then monitoring takes place if and only if $\hat{\theta} \in M$.

The termination/continuation policy Φ is also a subset of Θ . Unlike in the case of complete information, here the interpretation of Φ must take into account the fact that there is information asymmetry between the investor and the entrepreneur concerning the realization of θ . Let $\tilde{\theta}$ denote the investor's knowledge of the realization of θ on which the termination/continuation decision must be conditioned:

$$\tilde{\theta}(\hat{\theta}, \theta) = \begin{cases} \theta, & \text{if } \hat{\theta} \in M, \\ \hat{\theta}, & \text{otherwise.} \end{cases}$$

Then the project is continued if $\tilde{\theta} \in \Phi$, and it is terminated if $\tilde{\theta} \notin \Phi$.

In the state of termination, the investor seizes the scrap value of the project ϵ , and the entrepreneur receives a payment equal to $y(\tilde{\theta})$. In the states where the project is continued, the

⁵In the case where t is observable to the investor, consider the following contract: $x = u_0$, $\Phi = [\theta_b, 1]$, there is no monitoring, and the entrepreneur's compensation in period 2 is equal to t whenever he reports $\theta \in [\theta_b, 1]$ and makes effort t . This contract is incentive compatible: the entrepreneur is indifferent between reporting truthfully and lying and we assume he reports truthfully. This contract always implements the first-best outcome.

entrepreneur's compensation is $R_0(\tilde{\theta}) \geq 0$ if the project eventually fails; and his compensation is $R_0(\tilde{\theta}) + R(\tilde{\theta}) \geq 0$ if the project succeeds and the return H is realized. Finally, the contract also specifies a fixed cash payment equal to $x \geq 0$ for the entrepreneur. This payment is not contingent on anything that occurs after the investment has taken place.⁶

We assume that if the entrepreneur is indifferent between reporting truthfully and lying, he reports truthfully. This implies that the entrepreneur will not submit a false report of θ for monitoring. In other words, we have:

Lemma 1. *If $\hat{\theta} \in M$, then $\hat{\theta}(\theta) = \theta$.*

At this point, it is useful to define the following subsets of Θ :

$$A \equiv \Phi \cap M, \quad B \equiv \Phi' \cap M, \quad C \equiv \Phi \cap M', \quad D \equiv \Phi' \cap M',$$

where Φ' and M' are the complements of Φ and M , respectively. By Lemma 1, if the entrepreneur's report of the state $\hat{\theta}$ is in $A[B]$, then monitoring will occur and the project will [will not] continue. On the other hand, if $\hat{\theta} \in C[D]$, then the project is not monitored and it will [will not] continue.

Consider set D , the non-monitoring/termination region. Suppose that $\theta_1, \theta_2 \in D$ and $y(\theta_1) > y(\theta_2)$. Then, whenever θ_2 is realized, the entrepreneur could lie and report θ_1 to get the higher payoff $y(\theta_1)$, given that both θ_1 and θ_2 are not monitored. This implies that $y(\theta)$ must be constant on D in order for the contract to be incentive compatible.

Lemma 2. *An incentive compatible contract satisfies $y(\theta) = Y_D$ for all $\theta \in D$.*

Given Lemmas 1 and 2, an incentive compatible contract must satisfy the following three sets of incentive constraints. First, there should be no incentives for the entrepreneur to report untruthfully a $\hat{\theta} \in C$ in order to continue the investment process without being monitored (conditions (6) and (8)). Second, there should be no incentives for the entrepreneur to report untruthfully a $\hat{\theta} \in D$, either to abandon a good project in order to avoid making effort and receive compensation Y_D (condition (7)), or to simply avoid being monitored (condition (9)). Third, there should be no incentives for the entrepreneur to shirk when the project is continued (condition (10)). Formally,

Truth-telling constraints:

$$\forall \theta \in A \cup C, \forall \hat{\theta} \in C \quad \theta R(\theta) + R_0(\theta) - t \geq \max\{\theta R(\hat{\theta}) + R_0(\hat{\theta}) - t, R_0(\hat{\theta})\} \quad (6)$$

$$\theta R(\theta) + R_0(\theta) - t \geq Y_D \quad (7)$$

$$\forall \theta \in B \cup D, \forall \hat{\theta} \in C \quad y(\theta) \geq \max\{\theta R(\hat{\theta}) + R_0(\hat{\theta}) - t, R_0(\hat{\theta})\} \quad (8)$$

$$\forall \theta \in B \quad y(\theta) \geq Y_D \quad (9)$$

Effort constraint:

$$\forall \theta \in A \cup C \quad \theta R(\theta) + R_0(\theta) - t \geq R_0(\theta) \quad (10)$$

⁶Clearly, x is a mathematically redundant component of the contract, and we have introduced x for mathematical convenience.

The entrepreneur's participation constraint is as follows:

$$x + \int_{A \cup C} (\theta R(\theta) + R_0(\theta) - t) dG(\theta) + \int_{B \cup D} y(\theta) dG(\theta) \geq u_0. \quad (11)$$

We are now in a position to define optimality. We call a contract optimal if it maximizes the investor's expected payoff, subject to the incentive constraints, the participation constraint, and the non-negativity constraint (13). That is, an optimal contract solves the following problem:

$$(P1) \quad \max_{\sigma} \int_{A \cup C} [\theta H - \theta R(\theta) - R_0(\theta)] dG(\theta) + \int_{B \cup D} [\epsilon - y(\theta)] dG(\theta) - \mu(A \cup B)\gamma - x \quad (12)$$

subject to (6)—(11)

$$x \geq 0, \quad \forall \theta \in B \cup D \quad y(\theta) \geq 0,$$

$$\forall \theta \in A \cup C \quad R_0(\theta) \geq 0, \quad R_0(\theta) + R(\theta) \geq 0 \quad (13)$$

where μ denotes the probability measure on Θ : for any set $Z \subseteq \Theta$, $\mu(Z) = \int_Z dG(\theta)$.

3.2.2. The optimal contract

We now set out to analyze the properties of the optimal contract. Our first task is to simplify the incentive constraints. The approach we take is to consider a class of optimal contracts that all deliver the same expected utilities to both the entrepreneur and the investor, and then show that each contract in that class is equivalent to a contract whose compensation scheme resembles that of a debt or an equity contract. In what follows, any two contracts are said to be *equivalent* if they satisfy the same set of constraints and promise the same expected payoffs to both the investor and the entrepreneur.

Proposition 1. *For any contract σ that solves (P1), there exists a contract $\hat{\sigma}$ that is equivalent to σ , and $\hat{\sigma}$ has the following properties: for all $\theta \in A \cup C$, $R_0(\theta) = 0$, and for all $\theta \in C$, $R(\theta) = R_C$, where $R_C \geq 0$ is a constant.*

Proposition 1 implies that we can focus on the set of contracts that have a relatively simple compensation structure: conditional on the project being continued, the entrepreneur's compensation is zero if the project fails. Moreover, if the project succeeds, and if there was no monitoring, then the entrepreneur's compensation is independent of his report of θ . The intuition for this is simple. The debt structure is efficient here partly because it imposes the largest possible punishment for a bad outcome. That compensation is constant for θ in C is required by truth-telling. The technical proof of this proposition, however, is somewhat involved because of the tangled truth-telling and effort-making incentive constraints. The proof of Proposition 1 is in the appendix.

Given that we can set $R_0(\theta) = 0$ for all $\theta \in A \cup C$, constraint (10) implies $R(\theta) \geq t/\theta > 0$ for all θ in A and C . This in turn implies the non-negativity of $R(\theta)$ for all $\theta \in A$ and the

non-negativity of R_C . We can then rewrite the optimal contracting problem as follows:

$$(P2) \quad \max_{\sigma} \int_A [\theta H - \theta R(\theta)] dG(\theta) + \int_C (\theta H - R_C) dG(\theta) + \int_B (\epsilon - y(\theta)) dG(\theta) + (\epsilon - Y_D) \mu(D) - \mu(A \cup B) \gamma - x \quad (14)$$

subject to

$$x \geq 0; \quad \forall \theta \in B \quad y(\theta) \geq 0; \quad Y_D \geq 0, \quad (15)$$

$$x + \int_A (\theta R(\theta) - t) dG(\theta) + \int_C (\theta R_C - t) dG(\theta) + \int_B y(\theta) dG(\theta) + Y_D \mu(D) \geq u_0 \quad (16)$$

and the following incentive constraints,

$$\forall \theta \in A \quad \theta R(\theta) - t \geq \theta R_C - t, \quad (17)$$

$$\theta R(\theta) - t \geq Y_D, \quad (18)$$

$$\forall \theta \in B \quad y(\theta) \geq \theta R_C - t, \quad (19)$$

$$y(\theta) \geq Y_D, \quad (20)$$

$$\forall \theta \in C \quad \theta R_C - t \geq Y_D, \quad (21)$$

$$\forall \theta \in D \quad Y_D \geq \theta R_C - t, \quad (22)$$

$$\forall \theta \in A \quad \theta R(\theta) \geq t, \quad (23)$$

$$\forall \theta \in C \quad \theta R_C \geq t. \quad (24)$$

Our next proposition shows that the optimal termination/continuation policy is monotonic. That is, if a project with success rate θ is continued, then any project with a higher rate of success is also continued.

Proposition 2. *For any optimal contract σ that solves problem (P2), there exists a contract $\hat{\sigma}$ that is equivalent to σ , and that $\hat{\sigma}$ has the following property: if $\theta_1 \in \Phi$ and $\theta_2 \in \Phi'$, then $\theta_1 > \theta_2$.*

Suppose a project with success rate θ_1 is continued, but a project with success rate $\theta_2 (> \theta_1)$ is terminated. Then by switching the positions of θ_1 and θ_2 , and by re-arranging the compensation schemes properly, one can achieve a Pareto improvement. The proof of Proposition 2 is in the appendix. Given Proposition 2, we can then focus, without loss of generality, on contracts with monotonic termination/continuation policies; that is, contracts in which the set Φ is an upper interval of Θ . Our next lemma shows that for optimality, this upper interval must not be empty.

Lemma 3. *If σ is an optimal contract, then $\Phi' \neq \emptyset$.*

The proof of Lemma 3 is in the appendix. The next proposition specifies the main structure of the optimal contract.

Proposition 3. *The optimal contract has the following characteristics:*

- (i) $B = \emptyset$ and $\Phi' = D$.
- (ii) There exist constants θ_m and θ_n , $0 < \theta_m \leq \theta_n \leq 1$ such that

$$A = [\theta_m, \theta_n), \quad C = [\theta_n, 1], \quad D = [0, \theta_m].$$

- (iii) Moreover, the optimal compensation scheme is:

$$\forall \theta \in A = [\theta_m, \theta_n) \quad R(\theta) = t/\theta, \quad \forall \theta \in C = [\theta_n, 1] \quad R_C = t/\theta_n, \quad Y_D = 0.$$

By Proposition 3, it is never optimal to have the project monitored and then abandoned. Moreover, the optimal monitoring strategy is to monitor the reports of θ which are neither too low, nor too high. Put differently, it is optimal *not* to monitor when the “news” from the entrepreneur is sufficiently good, or sufficiently bad. The entrepreneur’s compensation is zero when the project is abandoned and when the project is continued but fails. When the project is continued and succeeds with return H , the entrepreneur’s compensation is nonlinear in the realization of θ : it is relatively high but decreasing in the success rate θ in the region where monitoring occurs, and it is low (but positive) and constant across the region where monitoring does not occur.

Note that the optimal monitoring strategy is not monotonic over the whole state space Θ , although it is monotonic conditional on the investment being continued. Also, conditional on the continuation of the project, the entrepreneur’s *expected* net compensation is monotonic and piecewise linear in θ : it is zero for all $\theta \in A$ and $(t/\theta_n)\theta > 0$ for all $\theta \in C$.

We explain why Proposition 3 must hold. The intuition for $B = \emptyset$, or that there are no states in which the project is monitored and then abandoned, is simple: if the project is to be abandoned, then monitoring is a waste of resources. In particular, if B is not empty, then the investor can do strictly better simply by moving all the elements in B to D .

Conditional on continuation, it is optimal to monitor lower rather than higher reports of θ , because this minimizes the cost of monitoring. To see this, suppose there is some $\theta' \in A$ that is higher than the lowest state in C . Then move θ' into C (that is, continue the project with report θ' without monitoring), reduce the compensation at θ' from $R(\theta')$ to R_C ; meanwhile, increase the fixed payment x so that the entrepreneur’s expected payoff remains constant. This change does not violate incentive compatibility: (a) The effort-making constraint is satisfied at θ' since it is satisfied at the lowest state in C ; and by assumption, θ' is strictly greater than the lowest state in C ; (b) when the realization is θ' , there are no incentives for the entrepreneur to misreport a θ'' in D ,⁷ and (c) there are no incentives for the entrepreneur to misreport a $\theta'' \neq \theta'$ in C since the payoff is constant in C . However, this change strictly improves the investor’s expected payoff because it reduces monitoring cost in state θ' .

⁷Suppose the entrepreneur reports $\theta'' \in D$. Then the project will be terminated and the payoff for the entrepreneur is Y_D , which is lower than $\theta' R_C - t$, the entrepreneur’s expected payoff if the project continues.

For part (iii) of the proposition, note that it is optimal to set the entrepreneur's compensation, (which is equal to Y_D in the termination region D , and is equal to $R(\theta)$ and R_C in the continuation regions A and C respectively,) to be just high enough that there are sufficient incentives for truth-telling and effort-making. Having the levels of Y_D , $R(\theta)$ and R_C too high is potentially costly: it may cause the entrepreneur's overall expected utility to exceed his reservation utility, given that $x \geq 0$. Now, given (i) and (ii), it is easy to check that the compensation scheme specified in the proposition is the lowest that satisfies all the incentive constraints.

Given Proposition 3, the optimal contract is characterized fully by the variables x , θ_m , and θ_n , with $x \geq 0$ and $\theta_m \leq \theta_n$. This allows us to rewrite the optimal contracting problem (P2) as

$$(P3) \quad \max_{x, \theta_m, \theta_n} \int_{\theta_m}^{\theta_n} (\theta H - t - \gamma) dG(\theta) + \int_{\theta_n}^1 (\theta H - \theta \frac{t}{\theta_n}) dG(\theta) + \epsilon G(\theta_m) - x \quad (25)$$

subject to

$$x + T(\theta_n) \geq u_0, \quad (26)$$

$$x \geq 0, \quad \theta_m, \theta_n \in [0, 1], \quad \theta_m \leq \theta_n, \quad (27)$$

where

$$T(\theta_n) \equiv \int_{\theta_n}^1 (\theta \frac{t}{\theta_n} - t) dG(\theta). \quad (28)$$

Here $T(\theta_n)$ is the lowest expected utility that the contract must promise to the entrepreneur if the investor wishes to continue the project whenever $\theta \geq \theta_n$ and he does not want to monitor. By (28)

$$T'(\theta_n) = -\frac{t}{(\theta_n)^2} \int_{\theta_n}^1 \theta dG(\theta) < 0.$$

Thus, in the absence of monitoring, more must be promised to the entrepreneur the larger the continuation region (i.e., the lower θ_n).

Obviously, problem (P3) has a solution; denote it $\{x^*, \theta_m^*, \theta_n^*\}$. We now solve for this solution. First, if it holds that

$$T(\theta_{fb}) \leq u_0, \quad (29)$$

then clearly

$$\theta_m^* = \theta_n^* = \theta_{fb} = \frac{1}{H}(t + \epsilon) \quad (30)$$

and

$$x^* = u_0 - T(\theta_{fb}). \quad (31)$$

That is, if condition (29) holds, then the optimal contract does not involve monitoring, and the optimal cutoff level for continuation coincides with the first-best continuation level θ_{fb} . In fact, condition (29) is the necessary and sufficient condition for (29)-(30) to be the solution for (P3) and for the first-best continuation/termination policy to be achievable.

Next, suppose that condition (29) does not hold. Then it must hold that $x^* = 0$, for otherwise, given $T'(\theta_n) < 0$, it is possible to reduce the values of x and θ_n simultaneously and to increase the investor's expected payoff without violating the participation constraint (26). With $x^* = 0$, constraint (26) becomes $T(\theta_n) \geq u_0$. Let $\bar{\theta}_n$ be the level of θ at which the constraint binds; that is,

$$T(\bar{\theta}_n) = u_0. \quad (32)$$

Since $u_0 = T(\bar{\theta}_n) < T(\theta_{fb})$, and T is decreasing in θ_n , we have $\theta_{fb} < \bar{\theta}_n \leq 1$. Also because $T(\theta_n)$ is a decreasing function, the participation constraint $T(\theta_n) \geq u_0$ is equivalent to $\theta_n \leq \bar{\theta}_n$.

We can now rewrite problem (P3) for the case of $u_0 < T(\theta_{fb})$ as follows:

$$(P4) \quad \max_{\hat{\theta}_m, \hat{\theta}_n} O(\theta_m, \theta_n) \equiv \int_{\theta_m}^1 (\theta H - t) dG(\theta) + \epsilon G(\theta_m) - T(\theta_n) - \int_{\theta_m}^{\hat{\theta}_n} \gamma dG(\theta) \quad (33)$$

$$\text{subject to} \quad \theta_m, \theta_n \in [0, 1], \quad \theta_n \geq \theta_m, \quad \theta_n \leq \bar{\theta}_n. \quad (34)$$

Problem (P4) is well defined: the objective function $O(\theta_m, \theta_n)$ is strictly concave, and the constraint set defined by (34) is convex. So it has a unique solution, which will be given in our next proposition, Proposition 4. But first, we need to define a set of variables that will be important for characterizing the solution to (P4). Suppose, in problem (P4), constraint (34) is not imposed. Let $\{\hat{\theta}_m, \hat{\theta}_n\}$ be the solution to the resulting unconstrained optimization problem. Then,

$$\hat{\theta}_m = \frac{t + \epsilon + \gamma}{H} \quad (35)$$

$$\frac{t}{(\hat{\theta}_n)^2} \int_{\hat{\theta}_n}^1 \theta dG(\theta) - \gamma g(\hat{\theta}_n) = 0. \quad (36)$$

In addition, let $\tilde{\theta}_n$ be the solution to the following equation:

$$\tilde{\theta}_n = \frac{1}{H}(t + \epsilon) + \frac{t}{(\tilde{\theta}_n)^2 g(\tilde{\theta}_n) H} \int_{\tilde{\theta}_n}^1 \theta dG(\theta). \quad (37)$$

It is easy to show that $\theta_{fb} < \hat{\theta}_m \leq 1$, $0 \leq \hat{\theta}_n < 1$, and $\theta_{fb} < \tilde{\theta}_n < 1$.

Proposition 4. *The optimal contract takes one of the following three forms:*

- (i) *If $u_0 \geq T(\theta_{fb})$, then*
 $\sigma_n^{fb} : \quad \theta_n^* = \theta_{fb}, M^* = \emptyset, \Phi^* = [\theta_{fb}, 1], \text{ and } x^* = u_0 - T(\theta_{fb}).$
- (ii) *If $u_0 < T(\theta_{fb})$, and $\hat{\theta}_m < \min\{\hat{\theta}_n, \bar{\theta}_n\}$, then $\theta_m^* = \hat{\theta}_m, \theta_n^* = \min\{\hat{\theta}_n, \bar{\theta}_n\}$,*
 $\sigma_m^{sb} : \quad \theta_{fb} < \theta_m^* < \theta_n^*, M^* = [\theta_m^*, \theta_n^*], \Phi^* = [\theta_m^*, 1], \text{ and } x^* = 0,$
- (iii) *If $u_0 < T(\theta_{fb})$ and $\hat{\theta}_m \geq \min\{\hat{\theta}_n, \bar{\theta}_n\}$, then $\theta_m^* = \theta_n^* = \min\{\hat{\theta}_n, \bar{\theta}_n\}$,*
 $\sigma_n^{sb} : \quad \theta_{fb} < \theta_n^*, M^* = \emptyset, \Phi^* = [\theta_n^*, 1], \text{ and } x^* = 0.$

The proof of Proposition 4 is in the appendix. The contract σ_n^{fb} in (i) achieves the first-best outcome with no monitoring. When the first-best outcome is not achievable, that is, when $u_0 = T(\bar{\theta}_n) < T(\theta_{\text{fb}})$, either there is monitoring in states of θ that belong to a nonempty region $M^* = [\theta_m^*, \theta_n^*)$ (contract σ_m^{sb}), or there is no monitoring at all (contract σ_n^{sb}). Proposition 4 shows that the exact form of the optimal contract depends critically on u_0 , the promised utility of the entrepreneur. This will be important for the equilibrium analysis of the model that is to come shortly. But first we have:

Corollary. *Whenever first-best is not attainable, there is always over-liquidation at the optimum.*

The above corollary holds because, in the cases where monitoring is optimal, it holds that $\theta_m^* = \hat{\theta}_m > \theta_{\text{fb}}$; in the cases where monitoring is not optimal, it holds that $\theta_n^* = \min\{\tilde{\theta}_n, \bar{\theta}_n\} > \theta_{\text{fb}}$, since $\bar{\theta}_n > \theta_{\text{fb}}$ and $\tilde{\theta}_n > \theta_{\text{fb}}$.

Proposition 4 also shows that unless first-best is attainable, it must hold that $x = 0$; that is, the non-state-contingent component of the compensation scheme is zero. Moreover, at the optimum, the entrepreneur earns positive compensation only in states where the project is carried out without being monitored; that is, only when the project is sufficiently good (θ is sufficiently high), and when the project is ultimately successful.

3.2.3. The Investor's Value Function

Let $V(u_0)$ denote the investor's net value as a function of the entrepreneur's reservation utility u_0 . Then,

$$V(u_0) = H \int_{\theta_m^*}^1 (\theta - \theta_{\text{fb}}) dG(\theta) - \int_{\theta_m^*}^{\theta_n^*} \gamma dG(\theta) + \epsilon - (x + T(\theta_n^*)) - 1. \quad (38)$$

Obviously, when the first-best is attainable, $V(u_0)$ coincides with $V_{\text{fb}}(u_0)$ as defined in (5).

By Proposition 4, given the parameters of the model, H , γ , ϵ , and t , there are two possible cases: (1) $\hat{\theta}_m < \hat{\theta}_n$ where for some values of the entrepreneur's reservation utility u_0 , the optimal contract involves monitoring, and (2) $\hat{\theta}_m \geq \hat{\theta}_n$ where there is never monitoring at the optimum, regardless of the value of u_0 . Define

$$\Lambda = \hat{\theta}_n - \hat{\theta}_m. \quad (39)$$

Consider case (1) ($\Lambda > 0$). By Proposition 4, the value function (38) can be divided into four segments, as depicted in the top panel of Figure 2. When $u_0 \geq \hat{u}_0$, where

$$\hat{u}_0 = T(\theta_{\text{fb}}), \quad (40)$$

the optimal contract is the first-best contract σ_n^{fb} . The cutoff level of the entrepreneur's reservation utility that divides σ_m^{sb} and σ_n^{sb} , call it \bar{u}_0 , is determined by $\bar{\theta}_n = \hat{\theta}_m$, or

$$\bar{u}_0 = T(\hat{\theta}_m). \quad (41)$$

Clearly, $0 < \bar{u}_0 \leq \hat{u}_0$.

When the optimal contract is σ_n^{sb} , i.e., when $u_0 \in [\bar{u}_0, \hat{u}_0)$, the cutoff level of θ below which the project is terminated, θ_n^* , is given by $\bar{\theta}_n$. That is, the project is terminated with a higher probability as u_0 decreases, the entrepreneur is paid exactly u_0 , and the function V is strictly decreasing in u_0 .

When the optimal contract is σ_m^{sb} , i.e. when $u_0 \in [0, \bar{u}_0)$, the cutoff level of θ above which the continued project is not monitored, θ_n^* , can be either $\hat{\theta}_n$ or $\bar{\theta}_n$. When $\theta_n^* = \bar{\theta}_n$, the entrepreneur is paid exactly u_0 , and V is strictly decreasing in u_0 . When $\theta_n^* = \hat{\theta}_n$, the entrepreneur is paid $T(\hat{\theta}_n)$, which is independent of u_0 , and as a result V is constant in u_0 . Let u_0^1 be the boundary between the two regions, which is determined by $\hat{\theta}_n = \bar{\theta}_n$; that is,

$$u_0^1 = T(\hat{\theta}_n). \quad (42)$$

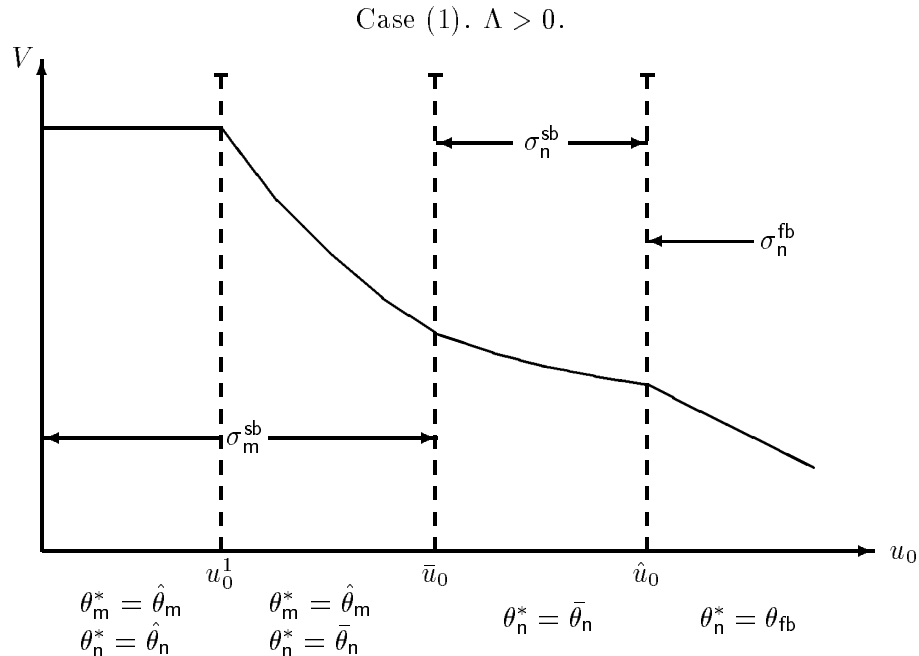
In case (2) ($\Lambda \leq 0$), the value function V is depicted in the bottom panel of Figure 2. Here, monitoring is *not* optimal for all values of u_0 . Moreover, the investor's value function is flat over the range of $u_0 \leq u_0^2$ where

$$u_0^2 = T(\tilde{\theta}_n). \quad (43)$$

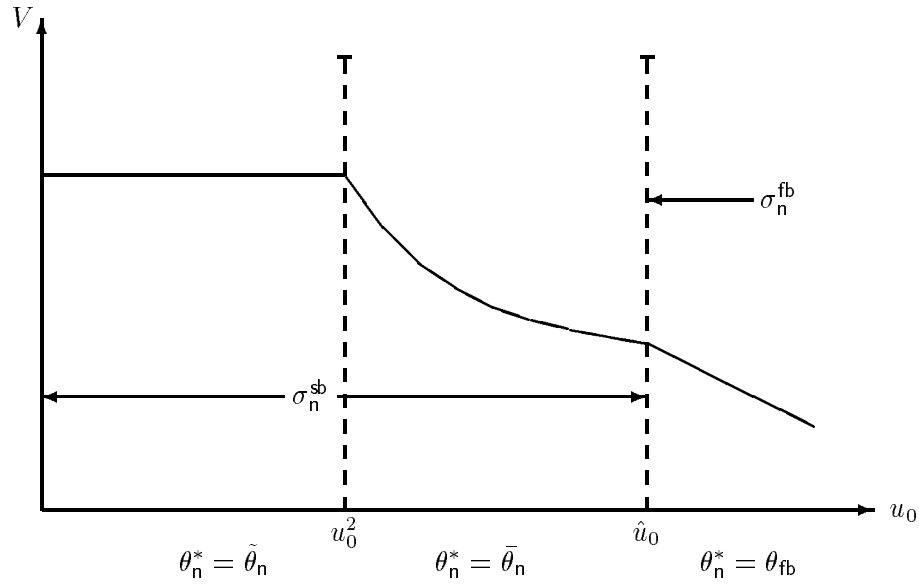
It is easy to verify that $u_0^2 < \hat{u}_0$.

It is important to note that Λ is not a function of u_0 . This means that as long as the model's parameters satisfy $\Lambda \leq 0$, monitoring is always a dominated strategy, regardless of the value of u_0 .

Figure 2. The Investor's Value Function



Case (2). $\Lambda \leq 0$.



3.2.4. Remark: Monitoring and Termination

In this section, we provide some intuitions for the results we have obtained so far. Notice that when $u_0 < T(\theta_{fb})$, costly incentive devices must be used by the investor to implement truth-telling and effort-making. There are three such devices. Device A: The investor can set a higher continuation threshold to terminate projects that otherwise should be continued. Device B: The investor can incur the cost γ to monitor the entrepreneur's report of the success rate θ . Device C: The investor can pay the entrepreneur extra compensation (i.e., in addition to u_0) to relax the incentive constraints.⁸

Notice that it is feasible for the investor to rely solely on any of the three devices to implement incentive compatibility. Specifically, the investor can set $\theta_m = \theta_n$ (this rules out monitoring as part of the optimal contract), where θ_n satisfies $T(\theta_n) = u_0$ (so the use of device C is also ruled out). Alternatively, the investor can monitor in every state of θ and then set $\theta R_m - t = 0$, and then pay the entrepreneur a lump-sum equal to u_0 . Finally, the investor can simply offer the entrepreneur the first-best contract that delivers an expected utility equal to $T(\theta_{fb})$. Clearly, these devices are all costly. The investor must use them efficiently for different combinations of the model's parameters. In most cases (specifically, if u_0 is not too low to make the agency problem too severe), the key to the determination of the investor's optimal choice of his incentive policy is the trade-off between termination and monitoring. Let us now explain how, exactly, this trade-off works.

⁸Essentially, increasing the entrepreneur's expected utility reduces the effects of the constraint of limited liability on incentives.

Figure 2 shows two features of the optimal contract as a function of u_0 . First, monitoring is (weakly) monotonically decreasing in u_0 . Second, for all $\gamma > 0$, either the optimal contract involves no monitoring for all u_0 (case 2) or the optimal contract involves no monitoring for u_0 sufficiently close to \hat{u}_0 ($\bar{u}_0 < u_0 < \hat{u}_0$). Remember \hat{u}_0 is the minimum expected utility that can be attained by a first-best contract.

To understand the logic behind these, suppose we start with $u_0 = \hat{u}_0$ and then lower u_0 . Suppose monitoring is not used (which means θ_n must be set to satisfy $T(\theta_n) = u_0$). Then a lower u_0 implies a higher θ_n . That is, in the absence of monitoring, more projects that have a positive net present value must be terminated when the entrepreneur's expected utility is lowered. This, in turn, implies that there are more benefits that monitoring can potentially offer, and hence the investor should make more use of monitoring. *Note that what monitoring does in the optimal contract is to allow the investor to lower the level of θ above which the project is continued, while not increasing the entrepreneur's expected utility.* (In the optimal contract, $\theta R_m(\theta) - t = 0$, for all $\theta \in \Phi_m$.)

This logic offers an alternative way to derive \bar{u}_0 , the cut-off level of the entrepreneur's utility below which monitoring is optimal and above which monitoring is not optimal. Specifically, let $\bar{u}_0 = T(\theta_*)$. Then at θ_* , the marginal benefit of monitoring must be equal to the marginal cost of monitoring. That is,

$$\theta_* H - t = \epsilon + \gamma,$$

which sets $\theta_* = \hat{\theta}_m$.⁹ This simply confirms equation (42). Note that the above equation also explains why, for any γ , no matter how small, monitoring is *not* optimal as long as u_0 is sufficiently close to θ_{fb} .

The trade-off between monitoring and termination predicts that as the cost of monitoring increases, monitoring is less efficient (while the cost of termination is not changed) and thus should be used less. In particular, $\Lambda = \theta_n - \theta_m$ should decrease as γ increases. Indeed, $\hat{\theta}_m$ increases as γ increases and it is straightforward to show that $\hat{\theta}_n$ decreases as γ increases (see the proof of Proposition 5).

Similarly, as H decreases (increases), that is, as termination becomes less (more) costly, termination should be used more (less) and monitoring should be used less (more). Indeed, for H sufficiently low, $\Lambda \leq 0$ and monitoring will not be used at all (see Proposition 6 in Section 4.2).

Here, it is important to note that the costs to the investor of using monitoring to achieve incentive compatibility remain constant when H changes to make termination more costly. More specifically, this means that for any given termination/continuation policy Φ , the minimum expected monitoring cost that the investor must incur in order to achieve incentive compatibility

⁹To see this more clearly, suppose the investor wants to use some monitoring at θ_* . By the structure of the optimal contract, the investor will choose $[\theta_m, \theta_*]$ to be the monitoring region and the project is continued if $\theta \geq \theta_m$, where $\theta_m < \theta_*$. The investor's benefits from doing this is equal to $\int_{\theta_m}^{\theta_*} (\theta H - t) dG(\theta)$, and the costs $\int_{\theta_m}^{\theta_*} (\epsilon + \gamma) dG(\theta)$.

is constant in H . This can be seen by simply noticing that H does not enter the constraints of problems (P1) and (P2). For any given Φ , the minimum expected cost of monitoring that the investor must incur in order to implement incentive compatibility is equal to the optimal value of the following optimization problem: minimize $\mu(A \cup B)\gamma$ subject to constraints (15)-(24).

3.2.5. Comparative Statics of the Optimal Contract

How the form of the optimal contract depends on the entrepreneur's expected utility u_0 is shown in Figure 2. How does the form of the optimal contract depend on H , γ and ϵ ?

Proposition 5. *Suppose that assumptions (1) and (2) hold. Suppose $V(u_0) \geq 0$. Holding the model's other parameters constant,*

- (i) (a) *There exists $\hat{H} \geq t + \epsilon$ such that the optimal contract is first-best if and only if $H \leq \hat{H}$;*
 (b) *There exists $\bar{H} > \hat{H}$ such that for all $H \in (\hat{H}, \bar{H})$, σ_n^{sb} is optimal, and for all $H > \bar{H}$, σ_m^{sb} is optimal.*
- (ii) *Whether first-best is attainable does not depend on γ . Suppose first-best is not attainable. Then there exists $\bar{\gamma} \in [0, H - t - \epsilon)$ such that for all $\gamma \leq \bar{\gamma}$, σ_m^{sb} is optimal; and for all $\gamma \geq \bar{\gamma}$, σ_n^{sb} is optimal.*
- (iii) *There exist $\hat{\epsilon}$ and $\bar{\epsilon}$, where $0 < \hat{\epsilon} < \bar{\epsilon}$, such that the optimal contract is σ_m^{sb} if $\epsilon \leq \hat{\epsilon}$; the optimal contract is σ_n^{sb} if $\hat{\epsilon} \leq \epsilon \leq \bar{\epsilon}$; the optimal contract is σ_n^{fb} if $\epsilon \geq \bar{\epsilon}$,*

Let's first hold the model's other parameters fixed and consider how the optimal contract responds to a change in H . Let's start with a very low H . A sufficiently low H implies that the ex post success rate of the project must be high enough in order to justify even first-best continuation. In particular, a sufficiently low H implies the first-best continuation/termination cut-off θ_{fb} is sufficiently high and hence $T(\theta_{\text{fb}})$ sufficiently low to make the equation $T(\theta_{\text{fb}}) \leq u_0$ hold and the first-best attained. This sufficiently low H is defined by \hat{H} . Next, let's increase H from \hat{H} . Now that the first-best is no longer attainable, costly incentive devices must then be used for attaining incentive compatibility. As H increases, termination becomes more costly. This makes monitoring relatively more efficient, given that the cost of monitoring remains constant as H changes. Thus as H becomes sufficiently high: higher than \bar{H} , monitoring will become optimal. However as long as H is not too high (lower than \bar{H}) to make monitoring more efficient or too low (lower than \hat{H}) to make first-best attainable, termination will be the dominating incentive device for the investor and the optimal contract is σ_n^{sb} .

Holding the model's other parameters constant, a lower γ strengthens the case for monitoring and a higher γ weakens the case for monitoring. Proposition 6 states that monitoring is optimal if and only if monitoring is sufficiently inexpensive ($\gamma < \bar{\gamma}$).

The effects of a change in ϵ on the form of the optimal contract go in exactly the opposite direction as that of H . When ϵ is higher, the opportunity cost of continuing a project is higher,

termination should occur with a higher probability, θ_{fb} is higher, $T(\theta_{fb})$ is lower. When ϵ is sufficiently high, higher than $\bar{\epsilon}$, $T(\theta_{fb})$ will be lower than u_0 and the first-best is attainable. On the other hand, as ϵ falls, for given realizations of the success rate θ , termination is more costly, the cut-off in θ for termination should be lower, implying that more incentives for truth telling are required, and hence more costly incentive devices are justified.

4. Credit Market Equilibrium

Before we proceed to define a credit market equilibrium, note that we have assumed up to now that the parameters of the model, H , t , ϵ , γ , and the density function $g(\theta)$, satisfy Assumptions (1) and (2). We now make a third assumption: at least for some values of u_0 , it is worthwhile for investors to invest in a project rather than in the storage technology. That is,

Assumption (3) There exists $u_0 > 0$ such that $V(u_0) \geq 0$.

Since the investors' value function $V(u_0)$ is weakly decreasing in u_0 , there exists $\underline{u}_0 > 0$ satisfying $V(\underline{u}_0) = 0$, such that for all $u_0 \leq \underline{u}_0$, $V(u_0) \geq 0$, and for all $u_0 > \underline{u}_0$, $V(u_0) < 0$.

We now describe an equilibrium. We adopt the following competition mechanism: the short side of the market extracts all of the surplus from trade.¹⁰ There are two cases. The first is the case where the economy's total supply of loanable funds exceeds its total demand for funds; that is, $\delta < \lambda$. In this case, competition among lenders will work to maximize the expected payoff for the entrepreneurs. Specifically, it will drive the expected payoff of each entrepreneur u_0 up to \underline{u}_0 at which the investors' expected payoffs are zero. Depending on the parameters of the model and on where \underline{u}_0 is located, the optimal contract can be σ_m^{sb} (if $\Lambda > 0$ and $\underline{u}_0 \in [0, \bar{u}_0]$), σ_n^{sb} (if $\Lambda > 0$ and $\underline{u}_0 \in [\bar{u}_0, \hat{u}_0]$ or if $\Lambda \leq 0$), or σ_n^{fb} (if $\underline{u}_0 \geq \hat{u}_0$).

In the second case, the economy's total supply of funds is less than the total demand for funds; that is, $\delta > \lambda$. In this case, not all projects are funded. Competition for funds will drive the entrepreneur's reservation utility u_0 to zero. In equilibrium, the expected utility of the entrepreneurs whose project is funded is equal to u_0^1 if $\Lambda > 0$, and it is equal to u_0^2 if $\Lambda \leq 0$. The optimal contract takes the form of σ_m^{sb} if $\Lambda > 0$; otherwise, it takes the form of σ_n^{sb} . See Figure 2. Finally, in the special case where $\delta = \lambda$, the two parties can divide the surplus from trade in any arbitrary way. For simplicity, we assume that the entrepreneurs get to extract all of the the surplus.

¹⁰Note that the simple competition mechanism we use may be interpreted as a special case of a bargaining process which in principle can take a more general form. For example, this bargaining process may dictate that the equilibrium fraction of the surplus associated with a project that goes to the investor is a function of δ and λ in the form of, say, $\eta(\frac{\delta}{\lambda})$, with $\eta \in [0, 1]$ and $\eta' < 0$. The special case we use in this paper simply sets

$$\eta\left(\frac{\delta}{\lambda}\right) = \begin{cases} 1 & \text{if } \frac{\delta}{\lambda} > 1 \\ 0 & \text{if } \frac{\delta}{\lambda} < 1. \end{cases} \quad (44)$$

Adopting the more general form of the competition mechanism does not affect our results qualitatively.

4.1. Credit Rationing

Notice that whenever $\delta > \lambda$, in equilibrium there is *always* credit rationing of the type discussed by Stiglitz and Weiss (1981) and Williamson (1987), where among a group of identical borrowers some receive loans and some do not, and those who do are strictly better off than those who do not.

Why is credit rationed? Sometimes it is because of costly monitoring, as in Williamson (1986, 1987).¹¹ Specifically, when σ_m^{sb} is optimal, lowering the entrepreneur's reservation utility u_0 implies a higher $\bar{\theta}_n$, which in turn implies that the expected monitoring cost is higher. Sometimes credit is rationed because of costly termination. In particular, when σ_n^{sb} is optimal, as u_0 decreases, more projects must be terminated in order to make the contract incentive compatible. The notion that credit rationing is a mechanism to avoid excessive termination has not been discussed in the literature. Stiglitz and Weiss (1981) model credit rationing as a mechanism to reduce costly ex post default on loans.

4.2. Equilibrium Comparative Statics

In this section we study the effects of a shock to the model's exogenous variables on the equilibrium financing mechanism and the economy's aggregate output. Throughout this section, we assume Assumptions (1)-(3) are satisfied.

4.2.1. A Negative Technology Shock

Consider the effects of a fall in the level of H while holding the model's other parameters fixed. As H falls, the cutoff level of θ below which termination occurs is higher, and more projects are terminated. There is a flight for quality in the sense that fully implemented projects have higher probabilities to succeed. Moreover, as H falls, the economy's total output falls more than proportionally. This is because, as H decreases, not only does each firm produce less, but there are fewer firms producing. In other words, the effects of the negative shock to H are amplified by the optimal contract.

Proposition 6. *Holding the model's other parameters fixed,*

- (i) $\frac{\partial V(u_0)}{\partial H} > 0$, $\frac{\partial u_0}{\partial H} > 0$, $\frac{\partial \bar{u}_0}{\partial H} > 0$, $\frac{\partial \hat{u}_0}{\partial H} > 0$, $\frac{\partial u_0^1}{\partial H} = 0$, $\frac{\partial u_0^2}{\partial H} > 0$.
- (ii) $\frac{\partial \Lambda}{\partial H} > 0$. *There exists $H^* \geq t + \epsilon + \gamma$, independent of u_0 , such that $\Lambda \leq 0$ if and only if $H \leq H^*$.*

The proof of Proposition 6 is straightforward and is left to the reader. Thus, as H decreases, the investor's value function will shift downward and to the left. Suppose the economy's initial

¹¹Williamson considers a standard costly state verification model.

equilibrium financing mechanism is bank lending. Then, as long as H falls sufficiently, the economy will experience a transition from bank loans to market lending. Market lending is *always* more efficient than bank loans if H is sufficiently low.

To summarize, in the model, a sufficiently strong negative technology shock (that reduces H) causes not only the aggregate output and total outstanding loans to fall, but also the equilibrium lending mechanism to shift away from intermediated bank loans to unintermediated market lending, provided that bank loans are initially the prevailing equilibrium financing mechanism. Clearly, what our model economy experiences here very much resemble what happened during the 1990-91 recession in the U.S. economy.

Note that, here, in order for the initial equilibrium financing mechanism to be bank loans, everything else constant, H must be sufficiently high. That is, in a sense, the economy must be initially in a “boom” in order to experience the kind of transition we have just described. Moreover, as H falls, the transition from bank loans to direct lending is monotonic.¹²

4.2.2. Shocks to δ and λ

Suppose the economy resides initially in an equilibrium where there is an oversupply of funds; i.e., $\delta < \lambda$, all projects are funded, the entrepreneurs extract all of the surplus from trade, and their expected payoff is \underline{u}_0 . The equilibrium lending mechanism is either bank loans or corporate bonds, depending on where \underline{u}_0 is located.

Imagine now the economy receives a shock that increases the number of investment opportunities from $\delta^\circ < \lambda$ to $\delta^n > \lambda$, while the supply of funds remains unchanged. This reversal of power in the credit market implies that now the entrepreneurs’ equilibrium expected utility is reduced from \underline{u}_0 to u_0^1 in the case of $\Lambda > 0$ and to u_0^2 in the case of $\Lambda \leq 0$. Remember that the sign of Λ does not depend on u_0 and, hence, δ and λ . A switch of the equilibrium lending mechanism can only occur in the case of $\Lambda > 0$, where the initial optimal contract prescribes σ_n^{sb} or σ_n^{fb} , and the new equilibrium contract is σ_m^{sb} . In such a case, lending activities will shift from the bond market to bank loans. In the meantime, the induced drop in the entrepreneurs’ reservation utility is likely to trigger more termination, unless \underline{u}_0 is in the flat portion of the value function V . However, the increase of investment opportunities implies that the number of projects that are funded will increase from δ° to λ . Here, the combined effect on total output is ambiguous, depending on whether the downward push of the increased termination can be overturned by the upward lift from the increased initial investment. It is possible that the effect of more termination dominates, and hence the economy’s total output falls, as the economy’s supply of investment projects increases.

¹²In order for monitoring to be initially optimal, the economy must have $\Lambda > 0$ before H falls. It is straightforward to check that in both cases, $\delta > \lambda$ and $\delta < \lambda$, the transition in the equilibrium financing mechanism is monotonic.

Next, consider the implications of a contraction in the supply of total loanable funds. Suppose again that initially $\delta < \lambda$. Suppose that a sequence of policy innovations brings λ to below δ . This shifts the market power away from the entrepreneurs and toward the investors, leading to a decrease in the entrepreneurs' equilibrium expected utility u_0^* (from \underline{u}_0 to u_0^1 or u_0^2), which, in turn, leads to more termination and a potential shift of the equilibrium lending mechanism from direct lending to intermediated lending. As the supply of loanable funds falls, total investment will fall, more termination will occur, and hence the economy's total output will also fall.

In our model, a fall in the supply of credit can create two effects: an interest rate effect and a credit effect. The interest rate effect occurs right when λ falls below δ , where there is a discrete downward jump in the economy's total investment and total output. This effect is caused by a sudden increase in the investor's expected return on a loan (rate of interest). As the expected returns on loans increase, the expected utilities of the entrepreneurs fall, the incentive problems become more severe, more projects are terminated, and aggregate output is lower. After λ has crossed δ , the credit effect takes over. As λ continues to fall, the economy's total investment and total output fall continuously, while the rate of interest remains flat. The same credit effect is discussed in Stiglitz and Weiss (1981), but a somewhat interesting point this paper offers is that a decrease in the total supply of loans may cause aggregate output to fall disproportionately. A fall in loan supply causes fewer projects to be funded initially, and among those that receive initial funding, more will face termination subsequently.

5. Concluding Remarks

We have studied a model of the credit market in which both the economy's total output and the equilibrium source of financing are endogenously determined. We have shown that the model provides a *real* explanation for the observation that, during recessions, bank lending usually falls, relative to corporate bond issuance.

Throughout the paper, we have assumed that monitoring is deterministic. That is, the investor monitors the entrepreneur's report of the type of his project with probability zero or one. This assumption is not essential. In an expanded version of the paper, we show that our results can be extended to environments where stochastic monitoring is permitted.

Another major simplifying assumption of the model is that the economy's total demand for and supply of funds are exogenously fixed. This could be relaxed. For instance, one could imagine that the availability of funds is an increasing function of the expected return on a loan to the investor, or one could assume that a higher expected return on a project to the entrepreneur brings a supply of more projects. But as long as these relationships are not sufficiently elastic, the comparative statics properties of the model will remain valid.

Appendix

Proof of Proposition 1.

Step 1. We show that there exists a contract $\tilde{\sigma} = \{M; \Phi; \tilde{x}; y(\theta), \theta \in B, Y_D; R(\theta), R_0(\theta), \theta \in A; \tilde{R}(\theta), \tilde{R}_0(\theta), \theta \in C\}$ such that $\tilde{\sigma}$ is equivalent to σ , and that $\tilde{R}(\theta)$ and $\tilde{R}_0(\theta)$ are constants on C . Note that $\tilde{\sigma}$ is identical to σ except for $\tilde{R}(\theta)$, $\tilde{R}_0(\theta)$, and \tilde{x} . Without loss of generality, assume that C has a minimum point, and let $\theta_1 \equiv \min_{\theta \in C} \theta$. Let $\tilde{R}(\theta) = R(\theta_1)$, $\tilde{R}_0(\theta) = R_0(\theta_1)$ for all $\theta \in C$, and allow \tilde{x} to be determined later.

(i) We first show $\tilde{\sigma}$ is incentive compatible. We need to show only that the revision on C satisfies conditions (6)—(8), and (10). Note that conditions (6) and (8) are obviously satisfied on C , given that $\theta_1 \in C$, and $\tilde{R}(\theta) = R(\theta_1)$, $\tilde{R}_0(\theta) = R_0(\theta_1)$, for all $\theta \in C$. Since condition (7) holds for $\theta = \theta_1$, we have for any $\theta \in C$, $\theta \geq \theta_1$,

$$\theta \tilde{R}(\theta) + \tilde{R}_0(\theta) \geq \theta_1 R(\theta_1) + R_0(\theta_1) \geq Y_D.$$

Thus $\tilde{\sigma}$ also satisfies (7) for all $\theta \in C$. Similarly, condition (10) holds for $\theta = \theta_1$, which implies $\theta_1 R(\theta_1) \geq t$. Then, for any $\theta \in C$, we have $\theta \tilde{R}(\theta) \geq \theta_1 R(\theta_1) \geq t$. That is, (10) is satisfied with any $\theta \in C$.

(ii) With $\tilde{\sigma}$ instead of σ , the entrepreneur's expected utility is different only in C . Let \tilde{x} be defined as follows:

$$\tilde{x} = x + \int_C (\theta R(\theta) + R_0(\theta)) dG(\theta) - \int_C (\theta R(\theta_1) + R_0(\theta_1)) dG(\theta).$$

We need to show $\tilde{x} \geq 0$. But by (6), $\forall \theta \in C$,

$$\theta R(\theta) + R_0(\theta) \geq \theta R(\theta_1) + R_0(\theta_1) = \theta \tilde{R}(\theta) + \tilde{R}_0(\theta).$$

That is, with $\tilde{\sigma}$, for all $\theta \in C$, entrepreneur \mathcal{E} 's expected payoff is less than or equal to that of the original contract. We therefore have: $\tilde{x} \geq x \geq 0$.

By (ii), investor \mathcal{I} 's payoff is the same with contract $\tilde{\sigma}$ as with contract σ . So, we have shown that $\tilde{\sigma}$ is equivalent to the original contract σ .

Step 2. We further demonstrate that the contract $\tilde{\sigma}$ is equivalent to a third contract $\hat{\sigma} = \{M; \Phi; \hat{x}; y(\theta), \theta \in B, \hat{Y}_D; \hat{R}(\theta), \hat{R}_0(\theta), \theta \in A \cup C\}$, which is otherwise identical to $\tilde{\sigma}$ except

$$\forall \theta \in A \quad \hat{R}_0(\theta) = 0, \quad \hat{R}(\theta) = R(\theta) + \frac{1}{\theta} R_0(\theta) \quad (45)$$

$$\forall \theta \in C \quad \hat{R}_0(\theta) = 0, \quad \hat{R}(\theta) = R_C \equiv \tilde{R}(\theta) = R(\theta_1) \quad (46)$$

$$\hat{Y}_D = Y_D - R_0(\theta_1), \quad \hat{x} = \tilde{x} + \mu(C)R_0(\theta_1) + \mu(D)(Y_D - R_0(\theta_1)). \quad (47)$$

(i) We show that this new contract $\hat{\sigma}$ promises the same expected utilities as does $\tilde{\sigma}$ to both entrepreneur \mathcal{E} and investor \mathcal{I} . The entrepreneur's expected payoff on A under $\hat{\sigma}$ is the same pointwise as under σ since for each $\theta \in A$,

$$\theta R(\theta) + R_0(\theta) = \theta(R(\theta) + \frac{1}{\theta}R_0(\theta)) + 0 = \theta\hat{R}(\theta) + \hat{R}_0(\theta).$$

By (46), under $\hat{\sigma}$, if the project with θ in C succeeds, entrepreneur \mathcal{E} receives the expected payoff R_C that he would receive under $\tilde{\sigma}$. His total expected payment on C when the project fails, $\mu(C)R_0(\theta_1)$, and part of the payment on D , $\mu(D)(Y_D - R_0(\theta_1))$ (which is positive by (8)), are moved from C and D , respectively, into the constant payment \hat{x} (an increase from \tilde{x}). Therefore, the two contracts give the same expected payoffs to both agents.

(ii) We show that the new contract $\hat{\sigma}$ is incentive compatible. First, since the changes on A do not affect the entrepreneur's expected payoff pointwise, the left-hand side of the relevant constraints (6) and (7) are the same as those under $\tilde{\sigma}$. Second, note that for any $\theta \in \Theta$, and any $\theta' \in C$, $\tilde{R}_0(\theta') = R_C \geq 0 = \hat{R}_0(\theta')$, and $\theta\tilde{R}(\theta') + \tilde{R}_0(\theta') - t \geq \theta R_C - t = \theta\hat{R}(\theta') + \hat{R}_0(\theta') - t$. Furthermore, $Y_D \geq \hat{Y}_D$. That is, the right-hand sides of conditions (6)—(9) under $\hat{\sigma}$ are all smaller than that under $\tilde{\sigma}$. Therefore, for any $\theta \in A \cup B$, conditions (6)—(9) are satisfied under $\hat{\sigma}$. Next, given that the definition of \hat{Y}_D by (49), and that conditions (6)—(8) are satisfied for any $\theta \in C \cup D$ under $\tilde{\sigma}$, they are also satisfied under contract $\hat{\sigma}$. Last, since constraint (10) is satisfied under $\tilde{\sigma}$, for any $\theta \in A$, $\theta R(\theta) \geq t$, and $\theta_1 R_C \geq t$. By (47), $\hat{R}(\theta) \geq R(\theta)$, so for any $\theta \in A$, $\theta\hat{R}(\theta) \geq \theta R(\theta) \geq t$. For any $\theta \in C$, $\theta \geq \theta_1$, so $\theta\hat{R}(\theta) = \theta R_C \geq \theta_1 R_C \geq t$. Therefore, condition (10) holds under $\hat{\sigma}$.

We have shown that incentive constraints (6)—(10) hold for $\hat{\sigma}$, and that both agents receive the same expected payoff under contract $\hat{\sigma}$ as under $\tilde{\sigma}$. Therefore, the two contracts are equivalent.

■

Proof of Proposition 2.

We first introduce some notation. Let X and Y be sets of real numbers. We say $X >^* Y$ if and only if for all $x \in X, \forall y \in Y, x > y$. Next, let $\mathcal{P} = \mu \times \mu$ be the product measure on $\Theta \times \Theta$. We say $X \gtrsim Y$ if $X >^* Y$ almost surely, that is, $X \gtrsim Y$ if $\mathcal{P}\{(x, y) \mid x \in X, y \in Y \text{ and } y > x\} = 0$. Given the above, the proposition states that for any given optimal contract σ , there is contract which is equivalent to σ and which satisfies $\hat{\Phi} >^* \hat{\Phi}'$.

We first show that σ satisfies $\Phi \gtrsim \Phi'$, which is equivalent to showing $A \gtrsim B, A \gtrsim D, C \gtrsim B$, and $C \gtrsim D$. Before proceeding, assume each of the four sets A, B, C and D has positive measure. (If one of the sets has measure 0, the corresponding assertion holds automatically.)

(i) We show $A \gtrsim B$. Suppose not, then there exist $\Theta_A \subseteq A$ and $\Theta_B \subseteq B$ such that $\Theta_B >^* \Theta_A$. Without loss of generality, suppose that Θ_A and Θ_B satisfy $\mu(\Theta_A) = \mu(\Theta_B) \neq 0$,¹³ and that $R(\theta)$

¹³Given θ is continuously distributed, the sets Θ_A and Θ_B can be cut arbitrarily small to satisfy this property.

has a minimum on Θ_A . Now consider an alternative contract $\tilde{\sigma}$ which is identical to σ except

(a) $\tilde{B} = B \cup \Theta_A \setminus \Theta_B$, and $\forall \theta_A \in \Theta_A$,

$$\tilde{y}(\theta_A) = \frac{1}{\mu(\Theta_B)} \int_{\Theta_B} y(\theta) dG(\theta).$$

(b) $\tilde{A} = A \cup \Theta_B \setminus \Theta_A$, and $\forall \theta_B \in \Theta_B$,

$$\tilde{R}(\theta_B) = R_{min}(\Theta_A) \equiv \min_{\theta \in \Theta_A} R(\theta),$$

(c) If $\int_{\Theta_B} \theta_B R_{min}(\Theta_A) dG(\theta_B) < \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A)$, then

$$\tilde{x} = x + \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A) - \int_{\Theta_B} \theta_B R_{min}(\Theta_A) dG(\theta_B)$$

otherwise, $\tilde{x} = x$.

We need only verify that the incentive constraints (17)—(20) and (23) hold for $\tilde{\sigma}$. Since $R(\theta_A) \geq R_C$ for all $\theta_A \in \Theta_A$, it holds that $\tilde{R}(\theta_B) = R_{min}(\Theta_A) \geq R_C$, for all $\theta_B \in \Theta_B$, or $\tilde{R}(\theta_B)$ satisfies (17). Now define $\check{\theta}_A \equiv \text{argmin}_{\theta \in \Theta_A} R(\theta_A)$. By (18) and (23), $\check{\theta}_A R(\check{\theta}_A) \geq Y_D t \geq t$. Thus for all $\theta_B \in \Theta_B$,

$$\theta_B \tilde{R}(\theta_B) \geq \check{\theta}_A \tilde{R}(\theta_B) = \check{\theta}_A R(\check{\theta}_A) \geq Y_D t \geq t.$$

That is, $\tilde{R}(\theta_B)$ satisfies (18) and (23). Since σ satisfies constraints (19) and (20), we have for all $\theta \in \Theta_B$, $y(\theta) \geq \max\{\theta R_C - t, Y_D\}$. Therefore, for all $\theta_A \in \Theta_A$,

$$\tilde{y}(\theta_A) = \frac{1}{\mu(\Theta_B)} \int_{\Theta_B} y(\theta) dG(\theta) \geq \max\left\{ \frac{R_C}{\mu(\Theta_B)} \int_{\Theta_B} \theta dG(\theta) - t, Y_D \right\} \geq \max\{R_C \theta_A - t, Y_D\},$$

or $\tilde{y}(\theta_A)$ satisfies (19) and (20). Thus we have shown that $\tilde{\sigma}$ is incentive compatible.

Next, we show that $\tilde{\sigma}$ Pareto dominates σ . By construction, $\int_{\Theta_A} \tilde{y}(\theta_A) dG(\theta_A) = \int_{\Theta_B} y(\theta_B) dG(\theta_B)$. Suppose $\int_{\Theta_B} \theta_B R_{min}(\Theta_A) dG(\theta_B) \geq \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A)$. Then moving from σ to $\tilde{\sigma}$ the entrepreneur's expected payoff is changed by

$$\int_{\Theta_B} \theta_B R_{min}(\Theta_A) dG(\theta_B) - \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A) > 0,$$

and the investor's expected payoff is changed by

$$\int_{\Theta_B} \theta_B (H - R_{min}(\Theta_A)) dG(\theta_B) - \int_{\Theta_A} \theta_A (H - R(\theta_A)) dG(\theta_A) > 0,$$

since $\Theta_B >^* \Theta_A$. Thus both parties are better off under $\tilde{\sigma}$ than under σ .

Suppose $\int_{\Theta_B} \theta_B R_{min}(\Theta_A) dG(\theta_B) < \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A)$. Then under $\tilde{\sigma}$ the entrepreneur's expected payoff decreases on Θ_B compared to what she receives under σ on Θ_A , but the decrease is made up exactly by the increase of x to \tilde{x} , so her total expected payoff remains the same.

Now the investor's expected payment to the entrepreneur is the same, but the investor's expected payoff is increased by

$$H\left(\int_{\Theta_B} \theta_B dG(\theta_B) - \int_{\Theta_A} \theta_A dG(\theta_A)\right) > 0.$$

This is because projects with higher success rates are continued. Again, $\tilde{\sigma}$ Pareto dominates σ .

(ii) We show $A \succsim D$. Suppose not, then there exist $\Theta_A \subseteq A$ and $\Theta_D \subseteq D$ such that $\Theta_D >^* \Theta_A$. Without loss of generality, suppose Θ_A and Θ_D satisfy $\mu(\Theta_A) = \mu(\Theta_D) \neq 0$, and $R(\theta)$ has a minimum on Θ_A . Now consider an alternative contract $\tilde{\sigma}$ which is identical to σ except

(a) $\tilde{D} = D \cup \Theta_A \setminus \Theta_D$, and $\forall \theta_A \in \Theta_A$, $\tilde{y}(\theta_A) = Y_D$.

(b) $\tilde{A} = A \cup \Theta_D \setminus \Theta_A$, and $\forall \theta_D \in \Theta_D$, $\tilde{R}(\theta_D) = R_{\min}(\Theta_A) \equiv \min_{\theta \in \Theta_A} R(\theta)$, $\forall \theta_D \in \Theta_D$.

(c) If $\int_{\Theta_D} \theta_D R_{\min}(\Theta_D) dG(\theta_D) < \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A)$,

$$\tilde{x} = x + \int_{\Theta_A} \theta_A R(\theta_A) dG(\theta_A) - \int_{\Theta_D} \theta_D R_{\min}(\Theta_A) dG(\theta_D),$$

otherwise, $\tilde{x} = x$.

Now since $\Theta_D >^* \Theta_A$ and (22) holds for σ , we have $Y_D \geq \theta_D R_C - t \geq \theta_A R_C - t$ holds for all θ_A and θ_D , and hence constraint (22) is satisfied by contract $\tilde{\sigma}$.

As in the proof for $A \succsim B$, we can show that $\tilde{R}(\theta_D)$ satisfies constraints (17), (18), and (23), and thus $\tilde{\sigma}$ is incentive compatible. As in the proof for $A \succsim B$, we can show that $\tilde{\sigma}$ Pareto dominates σ , a contradiction.

(iii) We show $C \succsim B$. Suppose not. Without loss of generality, assume that there exists $\Theta_B \subseteq B$ and $\Theta_C \subseteq C$ such that $\Theta_B >^* \Theta_C$, and $\mu(\Theta_B) = \mu(\Theta_C) \neq 0$. Now consider an alternative contract $\tilde{\sigma}$ which is identical to σ except

(a) $\tilde{B} = B \cup \Theta_B \setminus \Theta_C$, and $\forall \theta_C \in \Theta_C$,

$$\tilde{y}(\theta_C) = \frac{1}{\mu(\Theta_B)} \int_{\Theta_B} y(\theta) dG(\theta).$$

(b) $\tilde{C} = C \cup \Theta_C \setminus \Theta_B$, and $\forall \theta_B \in \Theta_B$, $\tilde{R}(\theta_B) = R_C$.

Since every $\theta_B \in \Theta_B$ satisfies (19) and (20), $y(\theta_B) \geq \max\{\theta_B R_C - t, Y_D\}$. Then, for any $\theta_C \in \Theta_C$,

$$\tilde{y}(\theta_C) = \frac{1}{\mu(\Theta_B)} \int_{\Theta_B} y(\theta) dG(\theta) \geq \max\left\{\frac{R_C}{\mu(\Theta_B)} \int_{\Theta_B} \theta dG(\theta) - t, Y_D\right\} \geq \max\{R_C \theta_C - t, Y_D\}.$$

That is, $\tilde{y}(\theta_C)$ satisfies (19) and (20). Also, take an arbitrary $\theta_C \in \Theta_C$, $\theta_C R_C - t \geq Y_D$ and $\theta_C R_C \geq t$. Since for any $\theta_B \in \Theta_B$, $\theta_B > \theta_C$, we have $\theta_B R_C - t \geq Y_D$ and $\theta_B R_C \geq t$, or,

constraints (21) and (24) are satisfied on Θ_B . So the modified contract satisfies all the relevant incentive constraints.

By construction, $\int_{\Theta_C} \hat{y}(\theta_C) dG(\theta_C) = \int_{\Theta_B} y(\theta_B) dG(\theta_B)$. But since $\Theta_B \succ^* \Theta_C$, the entrepreneur's expected payoff is increased by

$$\int_{\Theta_B} \theta_B R_C dG(\theta_B) - \int_{\Theta_C} \theta_C R_C dG(\theta_C) > 0,$$

and the investor's expected payoff is increased by

$$\int_{\Theta_B} \theta_B (H - R_C) dG(\theta_B) - \int_{\Theta_C} \theta_C (H - R_C) dG(\theta_C) > 0.$$

That is, both agents' expected payoffs are strictly higher under $\tilde{\sigma}$ than under σ .

(iv) Last, we show $C \succ D$. Constraints (21) and (22) directly imply that $C \succ^* D$, which further implies $C \succ D$.

To summarize, we have shown that $\Phi \succ \Phi'$. Given that contract σ satisfies $\Phi \succ \Phi'$, it is trivial to show that there is an equivalent contract $\hat{\sigma}$ that satisfies $\hat{\Phi} \succ^* \hat{\Phi}'$. Since $\Phi \succ^* \Phi'$ can only be violated on a measure zero set, we can rearrange monitoring and continuation/termination policies on this measure-zero set to eliminate the violations without affecting the payoffs. ■

Proof of Lemma 3.

Consider an optimal contract σ . By Proposition 2, we assume $\Phi = [\theta_1, 1]$ and $\Phi' = [0, \theta_1]$. Suppose $\Phi' = \emptyset$. Then consider contract $\hat{\sigma}$ which is otherwise identical to σ except

(a) $\hat{D} = [0, t/H)$, $\hat{A} = A \cap [t/H, 1]$, $\hat{C} = C \cap [t/H, 1]$.

(b) $\hat{Y}_D = 0$.

(c) If $R_C > H$, then $\hat{R}_C = H$; if $R_C \leq H$ (including the case $C = \emptyset$), then $\hat{R}_C = R_C$.

(d) $\hat{x} = x + \int_{A \cap \hat{D}} (\theta R(\theta) - t) dG(\theta) + \int_{C \cap \hat{D}} (\theta R_C - t) dG(\theta) + \int_{\hat{C}} \theta (R_C - \hat{R}_C) dG(\theta)$.

Notice that since $t > 0$, $\mu(\hat{D}) \neq 0$.

By construction, for all $\theta \in \hat{D}$, $\theta \hat{R}_C - t \leq \frac{t}{H} \hat{R}_C - t \leq 0 = \hat{Y}_D$, and hence constraint (22) is satisfied on \hat{D} . The contract $\hat{\sigma}$ satisfies constraint (17) since $R_C \geq \hat{R}_C$. $\hat{\sigma}$ also satisfies (18) since (23) holds under σ . If $\hat{R}_C = R_C$, then clearly constraints (21) and (24) are both satisfied. If $H = \hat{R}_C < R_C$, then for all $\theta \in \hat{C}$, $\theta \geq t/H$, $\theta \hat{R}_C = \theta H \geq t$, hence constraints (21) and (24) are also satisfied. Therefore, $\hat{\sigma}$ is incentive compatible. Finally, the expected payment to the entrepreneur under contract $\hat{\sigma}$ is the same as under contract σ , but the investor's expected payoff is increased by $\epsilon \mu(\hat{D}) - \int_{\hat{D}} (\theta H - t) dG(\theta) > 0$. This contradicts the fact that the σ is optimal. ■

Proof of Proposition 3.

We first show that given the optimal continuation/termination policy Φ , the optimal monitoring region A is a lower interval of Φ and the non-monitoring region C is the complement upper interval of Φ . Suppose this is not true. That is, suppose there is an optimal contract σ such that a subset Θ_A of A is embedded in C , that is, for all $\theta_A \in \Theta_A$, $\theta_A > \inf_C \theta$. Without loss of generality assume $\mu(\Theta_A) \neq 0$.

Consider contract $\hat{\sigma}$ which is otherwise identical to σ except

(a) $\hat{A} = A \setminus \Theta_A$, $\hat{C} = C \cup \Theta_A$, and $\forall \theta_A \in \Theta_A, \hat{R}(\theta_A) = R_C$.

(b) $\hat{x} = x + \int_{\Theta_A} \theta(R(\theta) - R_C)dG(\theta)$.

To show that $\hat{\sigma}$ is incentive compatible, we need only check that constraints (21) and (24) are satisfied for all $\theta_A \in \Theta_A$. Since for all $\theta \in C$, $\theta R_C - t \geq Y_D$, we have $(\inf_C \theta)R_C - t \geq Y_D$, which in turn implies for all $\theta_A \in \Theta_A$, $\theta_A R_C - t \geq Y_D$ given that $\theta_A > \inf_C \theta$. That is, (21) is satisfied. Constraint (24) is implied by (21) since $Y_D \geq 0$.

By construction of \hat{x} , the entrepreneur's expected payoff remains the same under $\hat{\sigma}$. But the investor gains by the savings of the monitoring cost $\gamma\mu(\Theta_A) > 0$. This contradicts the fact that the contract σ is optimal.

Next, we show that $B = \emptyset$. Let σ be an optimal contract which has $B \neq \emptyset$. By Proposition 2, $\Phi >^* \Phi'$, and from the above proof, $C >^* A$. Hence we can let $A = [\theta_m, \theta_n)$ and $C = [\theta_n, 1]$, where $\theta_m \leq \theta_n$. Consequently, $\Phi = A \cup C = [\theta_m, 1]$ and $\Phi' = B \cup D = [0, \theta_m)$. Notice Lemma 3 implies $\theta_m > 0$.

Consider an alternative contract $\hat{\sigma}$ which is otherwise identical to σ except

(a) $\hat{R}_C = t/\theta_n$.

(b) $\hat{D} = D \cup B$, and $\hat{Y}_D = 0$.

(c) $\hat{x} = x + \int_C \theta(R_C - \hat{R}_C)dG(\theta) + \int_B y(\theta)dG(\theta) + \int_D Y_D dG(\theta)$.

Using the equations $R_C \geq t/\theta_n = \hat{R}_C$ and $\hat{Y}_D = 0$, it is easy to check that the contract $\hat{\sigma}$ satisfies all the incentive constraints including (17), (18), (21), (22), (24), as well as the non-negative constraints (13). Moreover, the construction of \hat{x} implies that the entrepreneur's expected compensation under $\hat{\sigma}$ is the same as under σ . However, under $\hat{\sigma}$ the investor's expected payoff is increased by the savings of the monitoring cost $\gamma\mu(A \cup B) > 0$. This contradicts the assumption that σ is optimal.

Finally, we show (iii) holds. Suppose σ is optimal and it has a compensation scheme that differs from what is given by the proposition. We need only show that σ is equivalent to a contract $\hat{\sigma}$ whose compensation scheme takes the form that is given by the proposition. Let

the compensation scheme of $\hat{\sigma}$ be given by $\hat{R}(\theta) = t/\theta$ for all $\theta \in A$, $\hat{R}_C = t/\theta_n$, $\hat{Y}_D = 0$, and $\hat{x} = x + \int_A \theta(R(\theta) - t/\theta)dG(\theta) + \int_C \theta(R_C - \hat{R}_C)dG(\theta) + \mu(D)Y_D$. It is easy to check that $\hat{\sigma}$ satisfies incentive constraints (17)–(24). Since the compensation schedule of the contract σ also satisfies these constraints, in particular, for all $\theta \in A$, $R(\theta) \geq t/\theta = \hat{R}(\theta)$, for all $\theta \in C$, $R_C \geq t/\theta_n = \hat{R}_C$, and $Y_D \geq 0 = \hat{Y}_D$, we have $\hat{x} \geq x$. Clearly, the compensation scheme of $\hat{\sigma}$ conforms with the proposition, and $\hat{\sigma}$ is equivalent to σ . ■

Proof of Proposition 4.

We first show that by assumptions (1) and (2), the objective function $O(\theta_m, \theta_n)$ is strictly concave in both θ_m and θ_n .

The function $O(\theta_m, \theta_n)$ is strictly concave if its Hessian matrix is negative definite. By the definition of function $O(\theta_m, \theta_n)$ in equation (33), $\partial O(\theta_m, \theta_n)^2 / \partial \theta_n \partial \theta_m = 0$. So, we only need to show that the second derivatives with respect to θ_m and θ_n are strictly negative.

$$\frac{\partial^2 O(\theta_m, \theta_n)}{\partial \theta_m^2} = -[\theta_m H - (t + \epsilon + \gamma)]g'(\theta_m) - Hg(\theta_m).$$

If $\theta_m H - (t + \epsilon + \gamma) \geq 0$, then $\partial O(\theta_m, \theta_n)^2 / \partial \theta_m^2 < 0$ since by the first inequality of assumption (2),

$$\frac{-H}{H - (t + \epsilon + \gamma)} < \frac{-t}{H - (t + \epsilon)} \leq \frac{g'(\theta_m)}{g(\theta_m)}.$$

If $\theta_m H - (t + \epsilon + \gamma) < 0$, then $\partial O(\theta_m, \theta_n)^2 / \partial \theta_m^2 < 0$ is equivalent to

$$\frac{g'(\theta_m)}{g(\theta_m)} < \frac{H}{(t + \epsilon + \gamma) - \theta_m H}$$

which is implied by the second inequality of assumption (2). With respect to θ_n ,

$$\frac{\partial^2 O(\theta_m, \theta_n)}{\partial \theta_n^2} = -\frac{t}{\theta_n}g(\theta_n) - \frac{2t}{\theta_n^3} \int_{\theta_n}^1 \theta dG(\theta) - \gamma g'(\theta_n) < -\frac{t}{\theta_n}g(\theta_n) - \gamma g'(\theta_n) \equiv J(\theta_n).$$

By assumptions (1) and (2),

$$-\frac{t}{\gamma \theta_n} < -\frac{t}{\gamma} \leq \frac{-t}{H - (I_1 + t + \epsilon)} \leq \frac{g'(\theta_m)}{g(\theta_m)}$$

hence, $J(\theta_n) < 0$, or equivalently, $\partial O(\theta_m, \theta_n)^2 / \partial \theta_n^2 < 0$. So, $O(\theta_m, \theta_n)$ is strictly concave in both θ_m and θ_n .

Given that the objective function $O(\theta_m, \theta_n)$ is strictly concave, and that the constraint set defined by (34) is convex, problem (P4) has a unique solution. We can solve (P4) with *Lagrange's* method. Let ρ_1 be the multiplier for constraint $\theta_m \leq \theta_n$, and ρ_2 be the multiplier for constraint $\theta_n \leq \bar{\theta}_n$. Then the Lagrange is given by

$$L(\theta_m, \theta_n, \rho_1, \rho_2) = \int_{\theta_m}^1 (\theta H - t) dG(\theta) + \epsilon G(\theta_m) - T(\theta_n) - \int_{\theta_m}^{\theta_n} \gamma dG(\theta) + \rho_1(\theta_n - \theta_m) + \rho_2(\bar{\theta}_n - \theta_n).$$

The first-order conditions are

$$\frac{\partial L}{\partial \theta_m} = (-H\theta_m + t + \epsilon + \gamma)g(\theta_m) - \rho_1 = 0 \quad (48)$$

$$\frac{\partial L}{\partial \theta_n} = \frac{t}{(\theta_n)^2} \int_{\theta_n}^1 \theta dG(\theta) - \gamma g(\theta_n) + \rho_1 - \rho_2 = 0 \quad (49)$$

$$\frac{\partial L}{\partial \rho_1} = \theta_n - \theta_m \geq 0, \quad \rho_1 \geq 0 \quad \text{with complementary slackness} \quad (50)$$

$$\frac{\partial L}{\partial \rho_2} = \bar{\theta}_n - \theta_n \geq 0, \quad \rho_2 \geq 0 \quad \text{with complementary slackness.} \quad (51)$$

Depending on which of the constraints binds, there are four possible solutions for $\{\theta_m^*, \theta_n^*\}$.

- (a) $\rho_1 = 0, \rho_2 = 0$. Then neither constraint binds. By (48) and (49), $\theta_m^* = \hat{\theta}_m$ and $\theta_n^* = \hat{\theta}_n$.
- (b) $\rho_1 = 0, \rho_2 > 0$. Then by (48), $\theta_m^* = \hat{\theta}_m$, and θ_n^* is given by the binding constraint: $\theta_n^* = \bar{\theta}_n$.
- (c) $\rho_1 > 0, \rho_2 = 0$. Then $\theta_m = \theta_n$ and $\theta_n < \bar{\theta}_n$. Substituting $\rho_2 = 0$ and θ_m by θ_n , we get equation (37) from (48),(49) and (50). Given that $\tilde{\theta}_n$ is its solution, $\theta_m^* = \theta_n^* = \tilde{\theta}_n$.
- (d) $\rho_1 > 0, \rho_2 > 0$. Then both constraints binds: $\theta_m^* = \theta_n^* = \bar{\theta}_n$.

To summarize, the solution to (P4) can be one of the two classes, depending on whether $\theta_m^* = \theta_n^*$:

- (1) When $\hat{\theta}_m < \min\{\hat{\theta}_n, \bar{\theta}_n\}$, which includes cases (a) and (b), $\theta_m^* < \theta_n^*$. Then the monitoring region M^* is not empty, $M^* = [\theta_m^*, \theta_n^*)$, and the project-continuation region is $\Phi^* = [\theta_m^*, 1]$, where $\theta_m^* = \hat{\theta}_m, \theta_n^* = \min\{\hat{\theta}_n, \bar{\theta}_n\}$, and $\theta_m^* > \theta_{fb}$.
- (2) When $\hat{\theta}_m \geq \min\{\hat{\theta}_n, \bar{\theta}_n\}$, which includes cases (c) and (d), $\theta_m^* = \theta_n^*$. Hence, the monitoring region is empty, $M^* = \emptyset$, and the project-continuation region is given by $\Phi^* = [\theta_n^*, 1]$, where $\theta_n^* = \min\{\tilde{\theta}_n, \bar{\theta}_n\}$, and $\theta_n^* > \theta_{fb}$.

Proof of Proposition 5.

The effects of parameters H and ϵ on the attainability of the first-best are strightforward. In what follows, we consider only the case when the first-best is not achievable.

(i) Both $\hat{\theta}_n$ and $\bar{\theta}_n$ do not depend on H . When $H = \hat{H}$, $\bar{\theta}_n = \theta_{fb} \leq (t + \epsilon + \gamma)/H = \hat{\theta}_m$, since $u_0 = T(\theta_{fb}) = T(\bar{\theta}_n)$. Then, regardless of $\hat{\theta}_n$, $\hat{\theta}_m > \bar{\theta}_n \geq \min\{\hat{\theta}_n, \bar{\theta}_n\}$. When $H \rightarrow \infty$, $\hat{\theta}_m \rightarrow 0 < \min\{\hat{\theta}_n, \bar{\theta}_n\}$. Since $\hat{\theta}_m$ is a continuous function of H , and $\min\{\hat{\theta}_n, \bar{\theta}_n\}$ does not depend on H , there exists an $\bar{H} > \hat{H}$, such that for all $H \in (\hat{H}, \bar{H}]$, $\hat{\theta}_m \geq \min\{\hat{\theta}_n, \bar{\theta}_n\}$, which by Proposition 4, implies that the optimal contract is the one without monitoring σ_n^{sb} , and for all $H > \bar{H}$, $\hat{\theta}_m < \min\{\hat{\theta}_n, \bar{\theta}_n\}$, which, by the same proposition, implies that the optimal contract is the one with monitoring σ_m^{sb} .

(ii) The following facts are relevant to the proof of this statement.

(a) Given condition (29) does not hold, $\theta_{fb} < \bar{\theta}_n$, and both θ_{fb} and $\bar{\theta}_n$ are not functions of γ .

(b) When $\gamma \rightarrow 0$, $\hat{\theta}_m \rightarrow \theta_{fb} < \bar{\theta}_n$ and $\hat{\theta}_n = 1$, hence, $\hat{\theta}_m < \min\{\hat{\theta}_n, \bar{\theta}_n\} = \bar{\theta}_n$. When $\gamma = H - t - \epsilon$ (where $H - t - \epsilon$ is the maximum γ that is allowed by assumption (1)), $\hat{\theta}_m = 1 > \min\{\hat{\theta}_n, \bar{\theta}_n\}$.

(c) It is obvious that $\hat{\theta}_m$ is an increasing function in γ . Also, $\hat{\theta}_n$ as a solution to equation (36) is a decreasing function of γ , since totally differentiate (36) with respect to θ_n and γ at $\hat{\theta}_n$, we have

$$d\hat{\theta}_n/d\gamma = g(\hat{\theta}_n) / \left. \frac{\partial O(\theta_m, \theta_n)^2}{\partial \theta_n^2} \right|_{\theta_n=\hat{\theta}_n} < 0$$

given that function O is strictly concave. Since $\bar{\theta}_n$ does not depend on γ , $\min\{\hat{\theta}_n, \bar{\theta}_n\}$ is also a decreasing function of γ .

Both $\hat{\theta}_m$ and $\min\{\hat{\theta}_n, \bar{\theta}_n\}$ are continuous functions of γ . By (c), $\hat{\theta}_m$ is increasing in γ and $\min\{\hat{\theta}_n, \bar{\theta}_n\}$ is decreasing in γ . By (b), as $\gamma \rightarrow 0$, $\hat{\theta}_m < \min\{\hat{\theta}_n, \bar{\theta}_n\}$, but at $\gamma = H - t - \epsilon$, $\hat{\theta}_m > \min\{\hat{\theta}_n, \bar{\theta}_n\}$. Therefore, there exists a $\bar{\gamma} \in (0, H - t - \epsilon)$ such that $\hat{\theta}_m = \min\{\hat{\theta}_n, \bar{\theta}_n\}$, for all $\gamma < \bar{\gamma}$, $\hat{\theta}_m < \min\{\hat{\theta}_n, \bar{\theta}_n\}$, and for $\gamma \in [\bar{\gamma}, H - t - \epsilon]$, $\hat{\theta}_m \geq \min\{\hat{\theta}_n, \bar{\theta}_n\}$. Hence by Proposition 4, the optimal contract is the one with monitoring σ_m^{sb} for $\gamma < \bar{\gamma}$, and it is the one without monitoring σ_n^{sb} for $\gamma \in [\bar{\gamma}, H - t - \epsilon]$.

(iii) Both $\hat{\theta}_n$ and $\bar{\theta}_n$ do not depend on ϵ . Let $\bar{\epsilon} \geq 0$ be such that

$$\hat{\theta}_m(\bar{\epsilon}) = \frac{t + \bar{\epsilon} + \gamma}{H} = \min\{\hat{\theta}_n, \bar{\theta}_n\}.$$

If no such $\bar{\epsilon}$ exists, set $\bar{\epsilon} = 0$. Then by Proposition 4, for all $\epsilon < \bar{\epsilon}$, $\hat{\theta}_m(\epsilon) < \min\{\hat{\theta}_n, \bar{\theta}_n\}$, hence the optimal contract is σ_m^{sb} . For all $\epsilon \in [\bar{\epsilon}, \hat{\epsilon})$, $\hat{\theta}_m(\epsilon) \geq \min\{\hat{\theta}_n, \bar{\theta}_n\}$, hence the optimal contract is σ_n^{sb} .

■

References

- Bernanke, B., M. Gertler, and S. Gilchrist (1996). "The Financial Accelerator and the Flight to Quality," *The Review of Economics and Statistics* Vol. LXXVIII, No. 1, 1-15.
- Bolton, P. and X. Freixas (2000). "Equity, Bonds, and Bank Debt: Capital Structure and Financial Market Equilibrium under Asymmetric Information," *Journal of Political Economy* 108, 324-351.
- Boyd, J. H. and B. D. Smith (1997). "Capital Market Imperfections, International Credit Market, and Nonconvergence," *Journal of Economic Theory* 73, 335-364.
- Diamond, D. W. (1984). "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies* 51, 393-414.
- Diamond, D. W. (1991). "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt," *Journal of Political Economy* 99, 689-721.
- Friedman, B., and K. Kuttner (1993). "Economic Activity and the Short-Term Credit Markets: An Analysis of Prices and Quantities," *Brookings Papers on Economic Activity*, 193-266.
- Gale, D. and M. Hellwig (1985). "Incentive Compatible Debt Contracts: the One Period Problem," *Review of Economic Studies* 52, 647-663.
- Holmstrom, B. and J. Tirole (1997). "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics* Vol. CXII, 663-92.
- Kashyap, K., J. C. Stein and D. W. Wilcox (1993). "Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance," *American Economic Review* 83, 78-98.
- Repullo, R. and J. Suarez (1995). "Credit Market and Real Economic Activity: A Model of Financial Intermediation," Centre for Economic Policy Research Discussion Paper NO. 203.
- Stiglitz, J. and A. Weiss (1981). "Credit Rationing in Markets with Imperfect Information," *American Economic Review* Vol. 71 No.3, 393-409.
- Townsend, R. (1979). "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* 22, 265-293.
- Williamson, S. (1986). "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing," *Journal of Monetary Economics* 18, 159-79.
- Williamson, S. (1987). "Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing," *Quarterly Journal of Economics* 102, 135-146.