ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 677 November 2009

The Allocation of Liability: Why Financial Intermediation?

Tianxi Wang

The Allocation of Liability: Why Financial Intermediation?

Tianxi Wang*

Abstract

The paper proposes that the organization of financial markets is decided by the allocation of the liability to repay investors. Based on the liability allocation, the paper examines all possible modes of organizing finance and monitoring in an economy a la Townsend (1979). The equilibrium mode is either Financial Intermediation (FI) where the monitor alone takes the liability, or Conglomeration where it is taken by a Conglomerate composed of entrepreneurs and the monitor. Conglomeration also implements the benefit of diversification, which thus does not drive FI. Moreover, opposed to what Diamond (1984) would predict, monitoring costs advantage FI.

Key words: Allocation of the Liability Organization of Financial Markets Financial Intermediation Conglomeration Mechanism Design

JEL: D86, G00, G30

*University of Essex. Correspondence: Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK. Email: wangt@essex.ac.uk. I am indebted to John Moore for his thorough discussion, and to Gordon Kemp for his help in presentation and language. I also thank Madhav Aney, Jan Bena, Erlend Berg, Sudipto Bhattacharya, Giovanni Ko, Alan Morrison, and Motty Perry for their helpful comments.

1 Introduction

Banks entrap us in a crisis again. But could we get rid of them? The question sounds absurd, given the banks are providing some indispensable services, such as screening and monitoring. However, the services could be alternatively provided in separation from the intermediation of fund flow, with funds flowing directly from investors to entrepreneurs; so the service providers are specialists, not banks. Imagine factor owners, namely investors, the services providers and entrepreneurs, as atoms. Financial Intermediation (FI) is a molecule with a specific arrangement of the atoms. Could we nevertheless arrange them into other molecules where investors and the service providers are both connected to entrepreneurs only? Indeed, for the service of reducing entrepreneurs' private benefits, these two types of modes have been recognized by Holmstrom and Tirole (1997), who, however, fail to distinguish them in economic sense. This paper is the first to suggest that the key is to consider who takes the liability to repay investors. Where it is taken by the service provider alone, the mode is FI and the provider is the bank; where entrepreneurs take the liability by some means, the mode is of direct finance and the service provider is a specialist. The way of fund flow hallmarks the allocation of the liability.

The paper considers the economics of the allocation of the liability in an economy where investors resort to costly auditing to verify entrepreneurs' outputs.[1] However, the outputs can be cheaply observed by an expert, Ms X, through monitoring. There are two entrepreneurs, each with a project. Suppose first they are financed independently, in the state of one only having succeeded, the failed one declares default and is audited, at the auditing costs. Consider then the mode of FI, where Ms X runs a bank that finances both entrepreneurs and is financed by investors. In the above state, if the repayment from the successful entrepreneur suffices for the bank to clear its liability to the investors, the bank is not audited, and it does not audit the

[1] Townsend (1979) studies the contractual problem with this type of friction for the first time. Mookerjee and Png (1989) extend it considering the case with stochastic auditing strategies.

failed entrepreneur, as Ms X knows through monitoring of his actual failure. Compared to the mode of independent finance, therefore, auditing is saved under FI in the above state.[2]

However, besides independent finance and FI, there are other modes of financing. The paper exhausts all such modes. The real race is between FI and a mode called "Conglomeration", under which the entrepreneurs and Ms X form a conglomerate, each entrepreneur running a division, Ms X running the headquarters to monitor the divisions and advise them of how much each should contribute to clear the liability of the whole conglomerate. If funds from one successful division suffice to clear the whole liability, auditing happens only in state 0, when both divisions (entrepreneurs) have failed, as under FI.

As mentioned above, the two modes are differentiated by who is liable to repay the investors. It is Ms X alone under FI, while the entrepreneurs jointly under Conglomeration. The assets of the liability taker are the collateral upon which the investors hold claims. Thus, the collateral is the bank asset under FI, and the pool of all the projects under Conglomeration. In this economy, an asset partaking the collateral has three implications. I, the investors audit it when default is declared; II, they appropriate all its revenues whenever auditing uncovers any fraud in the declaration; III, before the investors are satisfied, not a bit of the asset can be disposed of. These three implications determine the economics of the allocation of the liability.

According to implication I, when default is declared, the investors audit *one* bank asset under FI, but two entrepreneur-projects under Conglomeration. Thus, organizing FI reduces the number of assets to be audited from Two to One. Therefore, FI has "Number Advantage".

To see the disadvantage of FI, assume the investors use stochastic rather than determin-

[2] Diamond (1984) is the first to show that FI dominates the mode of independent finance due to this kind of cost savings. Following him, Williamson (1986), Krasa and Villamil (1992) and Hellwig (2000) among others discuss the optimal contracts of the bank under various circumstances. But none of them (including Diamond) considers alternative financial modes that also accomodate monitoring service and implement the cost savings.

istic auditing strategies. According to implication II, the collateral is appropriated to punish misreporting default. The more the collateral worth, the harsher the punishment, and hence the smaller the probability auditing is needed. The revenues of the bank asset come from the returns of the projects. Hence, the collateral under FI is always worth less than that under Conglomeration. Conglomeration thus has "Collateral Advantage".

Can both advantages be materialized simultaneously by the hybrid mode of auditing Ms X's pocket only but appropriating all the projects as the punishment for cheating? No. In the hybrid mode, the entrepreneurs would always declare default, fill nothing into Ms X's pocket (which is thus to be audited), and promise to share the revenues of the projects with her after the auditing. According to implication III, the revenues are kept untouched during the auditing and are then ready to be shared afterwards. Thus, in the hybrid mode, this promise is credible and the collusion works. On the contrary, this promise is incredible under FI, where the entrepreneurs are free to dispose of what remains as soon as the bank declares the clearance of their liabilities to it. Hence, when the bank is being audited, they will dispose of all the revenues, leaving nothing to Ms X.

Therefore, the race of FI against Conglomeration, for the two-entrepreneur case, is decided by the strength of Number Advantage relative to Collateral Advantage. If and only if the number advantage is large enough, FI arises in equilibrium.

The comparison between the two modes is extended to the case of a large number of entrepreneurs. The two-entrepreneur case bears two costs: in state 0, auditing still happens since the liability to investors is not repaid, and in state 2, Ms X obtains rent since she is paid by two successful entrepreneurs while one payment suffices to clear the liability to the investors. Both costs are worn away when the number of entrepreneurs goes to infinity, because by the law of large numbers, on average the number of successes is almost fixed, and hence both the probability of the liability to investors not being fully repaid and that of Ms X acquiring rents go to 0. This argument is independent of the organization of monitoring service and is equally applicable to FI and Conglomeration. Therefore, both modes implement this benefit of diversification, which, therefore, does not drive FI, as Diamond (1984) claimed and the literature well accepted.

Two comparative statics results hold true for both cases. First, the larger the size of the projects, the greater the collateral advantage of Conglomeration, hence the less the chance they are financed by the bank.

Second, the chance of FI arising in equilibrium increases with monitoring costs. When the costs increase, entrepreneurs have to raise the repayment to the bank (Ms X) as compensation, which increase the value of the bank asset, and thus reduces the relative collateral advantage of Conglomeration.[3] Monitoring costs are the expenses with which a financial expert acquires soft information on a project, and hence measure the complexity of the project in particular and of the economic system in general. The comparative static result, therefore, predicts that over time, the prevalence of FI is growing and that of Conglomeration shrinking.[4]

This comparative static result is derived only if FI has been compared to Conglomeration (besides others); under both modes monitoring is provided. In fact, monitoring costs must disadvantage FI in Diamond (1984) and the literature following him, which compare FI *only* to independent finance, a mode involving no monitoring.

[3]For two-entrepreneur case, Bond (2004) endogenizes several modes of financing with a similar friction. One difference is in the information technology. In his paper, the disclosure costs are proportional to the number of agents to whom the truth is disclosed, while they are constant here. As a result, the driving force is to raise more investors into senior classes in that paper, while it is the trade-off between the two advantages in this paper and seniority plays no role. And because of that information technology, Bond's paper does not address the allocation of the liability to repay investors. Moreover, his paper does not derive the comparative static result with respect to monitoring costs.

[4]Diaz-Gimenez et al (1992) reports, in their table 3a, that the ratio to GNP of value added by the banking sector is increasing over 1950-1989. The tide of non-financial corporation spin-offs in recent decades is another suggestive evidence.

The rest of the paper is organized as follows. Section 2 sets up the model. Section 3 first exhausts all the possible allocations of the liability and then examines them one by one. Section 4 considers the case of a large number of entrepreneurs. Section 5 concludes. Technical proofs are relegated to the appendices.

2 The Model

Agents and Production

There are two dates, T_0 for investment and T_1 for return, and three classes of risk neutral agents: entrepreneurs, investors, and experts.

There are two entrepreneurs, E_1 and E_2, and each has an independent and identical project. A project needs a unit capital to invest, and returns R with probability q, and nothing with probability $1-q$. All entrepreneurs are penniless at T_0. There are infinite potential investors, each of whom has a small amount of capital, but the aggregate capital is well sufficient to finance all the projects. Experts have neither physical capital nor projects, but have the human capital of monitoring, the meaning of which will be clear soon. There are many experts. Nevertheless an entrepreneur can only accommodate one expert.

All agents are protected by limited liability, and no one discounts across the two dates. Entrepreneurs have all the bargaining power, that is, equilibrium will be driven by maximizing their expected profits.

Information Structure and Technologies

The only friction to finance a project is that only the entrepreneur costlessly observes its outcome, success or failure. For others to find out the outcome, two information technologies are available, with different costs and information strength. The weak technology is monitoring. If an expert has been monitoring a project from T_0 on, then she knows its outcome at T_1 through her personal experience, but she is not able to convince others of what she knows. The strong technology is auditing, which discloses the outcome to the public after it is realized. Accordingly, the monitoring costs per project, denoted by m, are much less than the auditing costs, denoted by C. Only experts know how to monitor, but the investors can access auditing, provided they afford C collectively.[5]

An expert can observe the outcome of a project at minor costs. If she never colludes with the entrepreneur, the investors can simply rely on her word of mouth to know the outcome and the unobservability of the outcome is not a problem any more. To exclude such a trivial solution, this paper is going to allow all possible collusion between the expert and the entrepreneurs. For that purpose, it is assumed that any side transfers between some or all of these non-investors are costlessly observable to none but the parties involved. The problem of collusion plainly precludes Maskin-Moore-Repullo mechanisms from functioning in this economy. Therefore, monitoring adds no value to a single entrepreneur. Together with the restraint of only one expert for one entrepreneur, it follows that if monitoring service is provided at all, it is provided by one expert only who monitors both entrepreneurs. This expert is called Ms X hereinafter.

Denote by $C_2 \geq C$ the costs of auditing Ms X's account. For simplicity, it is assumed that the action of monitoring is contractible, since the according moral hazard problem is not a necessary part of this paper.

A metaphor may be helpful. Assume each of the entrepreneurs and Ms X has a box (or pocket). An entrepreneur knows what is in his box, but cannot see into other boxes. Ms X knows what is in her box and can also see into an entrepreneur-box at minor costs m. The investors have to spend C to open an entrepreneur's box, and C_2 to open Ms X's box.

Additional assumptions are laid out below.

[5] Notice that here the auditing costs do not vary with the number of the agents to whom the truth is disclosed, as is assumed by Bond (2004).

Assumption 1: At T_0 the investors commit to playing a *stochastic* auditing strategy.

The investors are able to commit, since the action and the costs of auditing are verifiable and thus the investors have no difficulty in overcoming the collective action problem in connection with the commitment. Hereinafter, they will be dealt with as one party. This assumption of commitment facilitates the approach of mechanism design. Without it, I have to analyze a two-stage game, which is technically more complicated, but the main insights of this paper will be passed on. The other point of Assumption 1 is that the investors exercise auditing stochastically, which is the same as in Mookerjee and Png (1989), but different from in Diamond (1984). This assumption of stochastic auditing not only facilitates the mechanism design approach, but also uncovers the indispensable role of monitoring (see the remark following Lemma 2).

Assumption 2: $S \equiv qR - (1-q)C \geq 1$, $(1-q)C \geq \frac{qR}{2}$, and $0 < m < \frac{(1-q)q^2RC}{(q^2R+2(1-q)S)S}$.

Basically S is the "surplus" of a project and $S \geq 1$ ensures it is worthy of being financed. The other parts of the assumption ensure that the auditing costs are significant enough to leverage up various modes of financing, and that monitoring is of costs so low as to be provided in equilibrium.

Assumption 3: Securities issued to investors must bear repayments that weakly increase with the economic fundamental.

That is, the investors are repaid more when more projects succeed and the overall state of the economy is better. This is a feature of realistic securities. This assumption restricts the feasible sets of contracts to investors, and simplifies the problems of finding optimal contracts.

Timing

T_0 Morning: The entrepreneurs *cooperatively* decide how to get financed.[6] They decide

[6] At this time, they act as one and the same designer. I abstract away the game probably played between them at this time, as it would be very complicated to take it into account. For example, one entrepreneur's contracts could be contingent on the others', and vice versa, resulting in a problem of infinite recursiveness.

who takes the liability to repay investors and thus what the collateral is. Then they design accordingly the contractual arrangement (the mechanism) between them and the investors, and Ms X if the monitoring service is used. Only symmetric mechanisms are considered.

$\mathbf{T}_0$ **Afternoon:** The securities are issued to investors. After buying the securities, they commit to a stochastic auditing strategy.

$\mathbf{T}_1$ **Morning**: The outcomes of all the projects are realized. Non-investors could arrange various sorts of collusion.

$\mathbf{T}_1$ **Afternoon:** The liable entity reports the performance of the collateral to the investors and is ready to repay them accordingly. Contingent on the report, they audit the collateral according to the committed strategy, and if the auditing uncovers any fraud in the report, they appropriate the *whole* collateral.[7]

The Liability and the Collateral

The mode of financing is decided by who takes the liability to repay investors in this economy. The investors hold claims upon the revenues of the assets of the liability taker(s); these assets are thus the collateral to secure the claims. In this economy, the fact that an asset partakes the collateral has three implications, as follows:

I, the investors audit it contingent on the report of the performance of the collateral;

II, they appropriate the whole asset whenever auditing uncovers any fraud in the report;

III, any bit of the asset cannot be disposed of before the investors are satisfied, that is, either their claims are fully repaid, or they finish auditing and appropriation.

[7] Any securities to the investors must entail the rights to audit the collateral and furthermore to appropriate a part or the whole of it whenever auditing uncovers a fraud, because this is the only way to incentivize truth-reporting. Ex ante, the harsher the punishment, the lower the incentive to lie, and hence the less the auditing needed. This benefits the entrepreneurs, for the price of the securities has to ex ante compensate the auditing costs the investors expect to incur. In equilibrium, therefore, the securities will entail the maximum punishment, that is, a fraud triggers the appropriation of the *whole* collateral.

Implication III follows from the general principle that the debtor has rights to dispose of the collateral that secures the claims of the creditor only if the liability is declared to have been cleared. These three implications decide the economics of the allocation of the liability to repay investors in this economy.

In the next section, all the modes are listed, examined one by one, and compared to find the equilibrium mode.

3 The Modes of Financing

3.1 All the Possible Modes

Based on the allocation of the liability, we can envisage the following modes.

(1) Independent Finance (IF): each project is financed independently and is the collateral upon which the entrepreneur takes the liability to repay its investors.

(2) Joint Liability without Monitoring: the entrepreneurs, without using the monitoring service, take the joint liability to repay all the investors upon the pool of all the projects. That is, so long as one project has succeeded, all the investors are repaid with its revenues. This mode, as will be shown, is equivalent to IF.

(3) FI: Ms X alone takes the liability to repay all the investors, upon the bank asset that is formed by her investment in the projects.

(4) Conglomeration: the liability is taken upon the pool of all the projects by a conglomerate, where each project is a division managed by the entrepreneur and Ms X runs the headquarters monitoring the divisions. Conglomeration differs from mode (2) in monitoring being provided.

(5) The mix mode: the liable entity consists of Ms X and one entrepreneur, and both runs his project and finances the other projects as the intermediary. The collateral then consists of the directly financed project and the intermediary asset.

We may wonder if the modes listed above exhaust all the possibilities of financial modes. The following lemma answers this question.

Lemma 1 *Modes (1)-(5) exhaust all the possible modes of financing.*

Proof. Consider what the collateral could be. The funds to repay the investors, in the end, come from the revenues of the projects, either directly from them, or indirectly from the asset invested in them, or from the asset mixed with these two types of assets. For the first case, either the two projects become two separate collateral assets, which gives rise to mode (1), under which monitoring is useless; or the two projects are pooled into one collateral asset, which gives rise to modes (4) or (2), depending on whether monitoring is provided or not. The second case, where the investment in the projects forms the collateral asset, leads to FI, where the intermediary must be an expert (Ms X) because of the information advantage. This gives rise to thus mode (3). In the mixed case, the collateral asset consists of one project plus the intermediary asset in the other. Again, only Ms X can sensibly be the intermediary. Thus, the liability is taken jointly by her and the entrepreneur of the directly financed project, which is exactly mode (5). ■

Below, the optimal mechanism of each mode of (1) through (4) is derived. Then these modes are compared in the last subsection, where the mix mode is shown to be dominated by Conglomeration.

3.2 IF and Joint Liability without Monitoring

Under IF, each project is financed independently and is the collateral for the investors who finance the project. A project has two states, success and failure. Only in the state of success are the investors repaid, and the amount of the repayment, denoted by d, defines the security. If the entrepreneur reports a failure, the investors audit the project with probability l. A mechanism is then represented by (d, l).

In the state of success, if the entrepreneur reports the truth, he lays out d to clear his liability. If he lies and claims to have failed, with probability l, the project is audited, which uncovers the fraud and causes the whole revenue, R, to be appropriated; with probability $1-l$, the project is not audited and he escapes the liability. Thus, he expects to lose lR if lying. The incentive compatibility constraint (IC) for the entrepreneur to honestly report the success, therefore, is $d \leq lR$.

With probability q the project succeeds and the investors are repaid with d. With probability $1-q$ it fails, and the investors get repaid with nothing, and incur costs C to audit the project with probability l. Thus, the expected benefit of financing the project is $qd-(1-q)lC$, and the individual rationality constraint (IR) is $1 \leq qd-(1-q)lC$.

Each entrepreneur chooses (d,l) to minimize d subject to the IC and IR above. Both constraints are binding in the optimization. Therefore, the optimal mechanism, denoted by $\{d^I, l^I\}$, is $\{\frac{R}{S}, \frac{1}{S}\}$, where $S \equiv qR-(1-q)C \geq 1$ by Assumption 3.

We turn to mode (2). It arranges joint liability, which alone seems to be able to save auditing costs: if the liability of a failed project is repaid by the other entrepreneur, the project is saved from being audited. This benefit of cross-subsidization is nevertheless not implemented by mode (2), as is shown below.

The collateral under mode (2) is the pool of the two projects. It has 3 states, state $s=0,1$, and 2, defined by the number of successful projects, occurring with probabilities $(1-q)^2$, $2q(1-q)$, and q^2 respectively. A symmetric mechanism of the mode is represented by $\{D_s, l_s\}_{s=1,2}$: in state $s>0$, the investors are repaid with D_s, to which each successful entrepreneur equally contributes $\frac{D_s}{s}$, and a (reportedly) failed project is audited with probability l_s. A successful entrepreneur, if telling truthfully his success, expects to outlay $d^J=(1-q)D_1+q\frac{D_2}{2}$. If lying, his project is audited with probability $l^J=(1-q)l_0+ql_1$. Then, the IC is $l^J R \geq d^J$. The IR is $2q(1-q)D_1+q^2D_2 \geq 2+C((1-q)^2 \cdot 2l_0+2q(1-q)\cdot l_1)$, of which the left hand side (LHS) equals $2qd^J$ and the right hand side (RHS) equals $2+2l^JC$. It follows, together with the IC, that $l^J \geq \frac{1}{qR-(1-q)C} = l^I$. Thus, a failed project is audited with a probability no less than l^I under mode (2), which thus saves no auditing costs. On the other hand, mode (2) is not worse than IF, since joint liability bears as a special case independent liability. Therefore, the two modes are equivalent. To summarize,

Lemma 2 *Under IF a successful entrepreneur outlays $d^I = \frac{R}{S}$. Mode (2) is equivalent to IF, that is, monitoring is indispensable to implements the benefit of cross subsidization.*

The assumption of stochastic auditing is crucial to the indispensability of monitoring, which is not the case, if only deterministic auditing is allowed. Here is an example. Consider the following mechanism of mode (2): $D_2 = 2D_1 = 2D; l_0 = 1, l_1 = l_2 = 0$. The IR is $qD \geq 1 + (1-q)^2C$ and the IC is $(1-q)R \geq D$. If $q(1-q)R \geq 1 + (1-q)^2C \Leftrightarrow (1-q)S \geq 1$, there exists a range of (q, R, C) with which the IR and the IC are satisfied and hence auditing occurs only in state 0 under mode (2), which is this benefit of cross subsidization.

Monitoring helps implement that benefit, for the following reason. Exactly because joint liability makes a failed project less likely audited, it gives a successful entrepreneur higher incentive to hide his success. This incentive compatibility problem dissipates all the benefit of cross subsidization under stochastic auditing. Where Ms X knows the outcome of each project through monitoring, an entrepreneur has to buy her silence to hide the success, which lessens the incentive to lie.

There are various modes of accommodating the monitoring service. First, as presumed by Diamond (1984), there is the mode of FI, which is examined in the next subsection.

3.3 FI

Under this mode, Ms X becomes the bank and takes the liability to repay all the investors, upon the bank asset. At T_1, she reports how much funds she has collected from the entrepreneurs. Contingent on this report, the investors audit the bank asset with the committed probability at cost $C_2 \geq C$.

Definition 1 *FI has a "Number Advantage" if* $C_2 < 2C$.

The total costs of auditing the two projects are $2C$. $C_2 < 2C$ means banking technically saves auditing costs. This cost saving is called "Number Advantage", as it owes to the fact that the bank sets up a unified book of all its assets and rationalizes the book. the number advantage can be measured by $\frac{2C}{C_2}$.[8]

In this subsection, the marks of some equations are suffixed with "b" to indicate that they are peculiar to the mode of FI, where Ms X is the "b"ank.

The economy has three states, $s = 0, 1$, and 2, as the collateral of mode (2) has, occurring with probabilities $(1-q)^2$, $2q(1-q)$, and q^2 respectively. A general asset contract of the bank is $\{d_s\}_{s=0,1,2}$, and a liability contract is $\{D_s\}_{s=0,1,2}$: in state s, each successful entrepreneur repays Ms X (the bank) with d_s and she then passes D_s to the investors. At T_0, the investors, facing $\{d_s, D_s\}_{s=0,1,2}$, commit to auditing the bank with probability l_s when state s is reported. A mechanism is thus $\{d_s, D_s; l_s\}_{s=0,1,2}$.

By limited liability for Ms X and the entrepreneurs,

$$D_0 = d_0 = 0, \;\; D_1 \leq d_1 \leq R \text{ and } D_2 \leq 2d_2 \leq 2R \tag{LL}$$

Consider the ICs for the bank to truthfully report the revenues of its asset. We only present the binding ICs here, which are those that prevent the bank from misreporting state s to $s-1$

[8] Note that no number advantage is assumed by Diamond (1996), where the costs are the destruction of the low output, L, and thus $C = L$ and $C_2 = 2L$.

for $s = 2, 1$. The non-binding constraints are put in the appendix.

In state 2, the bank repays the investors with D_2 if honoring the contracts. If instead it lies that the state is 1, then, with probability l_1, the bank is audited and the whole asset, worth $2d_2$ in the state, is appropriated, and with probability $1 - l_1$, the reported state ($s = 1$) is accepted as the truth and accordingly the bank outlays D_1 to the investors. The IC for the bank not to misreport state 2 to 1 is, thus,

$$l_1 \cdot 2d_2 + (1 - l_1)D_1 \geq D_2 \tag{G21b}$$

The IC for the bank not to misreport state 2 to 0, (G20b), is not binding.

In state 1, misreporting it to state 2 is never profitable, since by Assumption 3, $D_2 \geq D_1$. Therefore, $l_2 = 0$. If the bank lies that the state is 0, it expects to lose the whole asset, now worth d_1, with probability l_0, and nothing otherwise. Thus, the IC for it not to misreport state 1 to 0 is

$$l_0 d_1 \geq D_1 \tag{G10b}$$

The investors obtain D_1 and D_2 in states 1 and 2 respectively. They audit the bank in states 0 and 1. The IR for the investors is

$$q^2 D_2 + 2q(1 - q)D_1 \geq 2 + [(1 - q)^2 l_0 + 2q(1 - q)l_1]C_2 \tag{IR-Ib}$$

Facing the bank contract $\Psi \equiv \{d_s, D_s\}_{s=0,1,2}$, the investors commit to the strategy $\{l_s\}_{s=0,1,2}$ that minimizes the auditing costs (the second term of the RHS of (IR-Ib)) subject to the ICs and IR above. Denote the optimal strategy by $\{l_s^B(\Psi)\}_{s=0,1,2}$. As mentioned above, $l_2^B(\Psi) = 0$.

The ICs above guarantee the investors not worse off by any fraud, but do not prevent a successful entrepreneur from being exploited by Ms X through partial collusion.[9] In state 2, she may arrange collusion in which she collects $t < d_2$ from E_1 to buy his silence when she declares

[9] A failed entrepreneur has nothing to be exploited. The collusion is qualified with "partial" since it does not involve all the three non-investors.

state 1 to the investors *and* E_2. With the collusion, she collects d_1 from E_2, besides t from E_1, and repays the investors with D_1, but only with probability $1 - l_1$, when this fraud is not uncovered by auditing; without the collusion, she gain $2d_2 - D_2$ in net. The IC for no partial collusion in state 2 is, thus, $2d_2 - D_2 \geq (1 - l_1)(d_1 + t - D_1)$ for any $t < d_2$. Equivalently,

$$2d_2 - D_2 \geq (1 - l_1)(d_1 + d_2 - D_1) \tag{P2b}$$

The IC for no partial collusion in state 1, (P1b), is not binding.

The investors do not care how the bank deals with the entrepreneurs and thus do not take into account (P2b) and (P1b) when choosing the auditing strategy. However, the entrepreneurs can alleviate the problem of partial collusion by inducing more auditing: the higher is l_1, the looser is (P2b).

Lastly, Ms X incurs costs to monitor the two projects. The IR for her is

$$q^2(2d_2 - D_2) + 2q(1 - q)(d_1 - D_1) \geq 2m \tag{IR-X}$$

The entrepreneurs' problem is then as follows.

Problem 1 $\min_{\{d_s;D_s,l_s\}_{s=0,1,2}} 2q(1-q)d_1 + q^2 \cdot 2d_2$, *subject to*

(1): $l_s = l_s^B(\Psi)$ *for* $s = 0, 1, 2$, *where* $\Psi \equiv \{d_s, D_s\}_{s=0,1,2}$.

(2): (LL), (P2b), (P1b), (IR-X), and $D_1 \leq D_2$.

The solution depends on whether (IR-X) is binding or not. If not, we have the following lemma. Let $B \equiv \frac{2+(1-q)^2 C_2}{2q-q^2}$

Lemma 3 *The optimal mechanism of FI is* $D_1 = D_2 = d_1 = d_2 = B$, *and* $l_0 = 1$ *and* $l_1 = l_2 = 0$, *if* $m < \frac{1}{2}q^2 B$ *when (IR-X) is not binding.*

Proof. See the appendix. ■

$l_1 = 0$ is driven by the trade-off between auditing costs and the rent to Ms X. She obtains the rent because she is the only one who knows of the overall state before auditing. However, auditing discloses this information to the public and thereby reduces her information advantage. Thus, the more auditing is exercised, the less rent she gains. Auditing costs have been assumed large enough so that the entrepreneurs would like gives Ms X a net rent of $V^B = q^2B - 2m$, but triggers auditing only in state 0.

$l_1 = 0$ implies $D_1 = D_2 \equiv D$ by (G21b); thus the liability contract is debt. Together with all these, the binding (P2b) implies that $d_1 = d_2 \equiv d$, which is presumed in Diamond (1984, 1996), but is driven by the partial collusion problem here. To find l_0, substitute $D_1 = D_2 = l_0 d$ ((G10b)) into the binding (IR-Ib),

$$l_0 = \frac{2}{(2q - q^2)d - (1-q)^2 C_2} \tag{1}$$

It is a decreasing function of d. Economically, with d increasing, the value of the collateral under FI (the bank asset) increases. The collateral is appropriated as the punishment for cheating; hence the more the collateral is worth, the less the auditing is needed, that is, the smaller the l_0.

When (IR-X) is not binding, l_0 can be moved freely. As d decreases with l_0 by (1), it is minimized at $l_0 = 1$, which together with (1) implies $d = B$. So we have lemma 3 above.

If $m \geq \frac{1}{2}q^2B$, (IR-X) is binding, so $V^B = 0$. With the concern for the rent to X removed, the driving force is to minimize auditing costs alone. Therefore, $l_1 = 0$. It follows that $D_1 = D_2$ and as above, $d_1 = d_2 = d$, which altogether implies (1). d is now pinned down by (1) and $q^2(2d - dl_0) + 2q(1-q)(d - dl_0) = 2m$, derived from the binding (IR-X). Let $d \equiv d^B(m)$. It strictly increases with m; intuitively, with monitoring costs increasing, the more is needed to pay the bank in order to satisfy its IR. And d is independent of R; indeed R is present neither in the constraints nor in the objective. Then, through (1), $l_0 = l^B(m) \equiv \frac{2}{(2q-q^2)d^B(m)-(1-q)^2C_2}$ is strictly decreasing with m and independent of R.

To summarize the two case, let us extend to define $d^B(m) = B$ for $m < \frac{1}{2}q^2 B$, so $d = d^B(m)$ and $l_0 = l^B(m)$ always. And $C^B \equiv (1-q)^2 C_2 l^B(m)$ always denote the expected auditing costs under FI. Then $V^B = \max(0, q^2 \frac{2+C^B}{2q-q^2} - 2m)$ always.[10] With these notations, the optimal mechanism of FI is summarized as follows.

Proposition 1 *Under FI, a successful entrepreneur outlays $d^B(m)$ as defined above, and the bank is audited only in state 0, with probability $l^B(m) = \frac{2}{(2q-q^2)d^B(m)-(1-q)^2 C_2}$, which is decreasing with m and independent of R, and Ms X obtains net rent of $V^B = \max(0, q^2 \frac{2+C^B}{2q-q^2} - 2m)$.*

3.4 Conglomeration

Under this mode, the liability is taken by a conglomerate, where each project becomes a division run by the entrepreneur, and monitored by Ms X, who becomes the headquarters. Here funds could flow, in a different way from under FI, directly between investors and the entrepreneurs. However, this difference does not matter at all. To highlight this point, we suppose Ms X also intermediates for the flow of funds under Conglomeration, collecting and distributing the investment capital at T_0 and the repayment funds at T_1; she may be regarded as the chief financial officer of the conglomerate. Then, the only difference between Conglomeration and FI, therefore, is in the collateral, and all the other respects are the same.

In this section, the marks of some equations are suffixed with "h" to indicate that they are peculiar to the mode of Conglomeration, where Ms X becomes the "h"eadquarters.

Parallel to the former subsection, a mechanism is $\{d_s, D_s; l_s\}_{s=0,1,2}$: in state s, as under FI, a successful entrepreneur contributes d_s to Ms X, who repays D_s to the investors on behalf of the conglomerate; l_s has a different meaning: l_0 (l_2) is the probability of auditing *each* project if state 0 (2) is reported and l_1 the probability of auditing the reportedly failed project if state

[10] When $m \geq \frac{1}{2}q^2 B$, $q^2 \frac{2+C^B}{2q-q^2} \leq q^2 \frac{2+(1-q)^2 C_2}{2q-q^2} = q^2 B \leq 2m$.

1 is reported.[11] The limited liability constraint is still (LL), as under FI.

Consider the ICs for the liable entity, the conglomerate now, to truthfully report the state to the investors. Again, we only present the binding ICs here, which are those that prevent the conglomerate from misreporting state s to $s-1$ for $s = 2, 1$. The non-binding constraints are put in the appendix.

In state 2, the conglomerate outlays D_2 if honoring the contracts. If it lies that the state is 1, then with probability l_1, the reportedly failed project is audited, the fraud uncovered, and the conglomerate loses $2R$; otherwise it outlays D_1, according to the reported state. The IC for the conglomerate not to misreport state 2 as state 1 is, thus,

$$l_1 \cdot 2R + (1 - l_1) \cdot D_1 \geq D_2 \tag{G21h}$$

The IC preventing misreporting state 2 as state 0, (G20h), is not binding.

Similarly, in state 1, the IC for the conglomerate not to misreport it as state 0 is

$$l_0 R \geq D_1 \tag{G10h}$$

Misreporting state 1 as state 2 is never profitable, as under FI, and thus $l_2 = 0$.

The investors audit two projects in state 0 and one in state 1. The IR for them is

$$q^2 D_2 + 2q(1-q)D_1 \geq 2 + [(1-q)^2 \cdot 2l_0 + 2q(1-q)l_1]C. \tag{IR-Ih}$$

The investors, facing the security $\Psi \equiv \{D_s\}_{s=0,1,2}$, commit to the auditing strategy that minimizes the expected auditing costs subject to the the ICs and the IR above. Let the optimal strategy be $\{l_s^H(\Psi)\}_{s=0,1,2}$. We saw $l_2^H(\Psi) = 0$.

[11] If the report is $\{s, f\}$ (project 1 succeeds and 2 fails), the truth is either $\{s, f\}$ or $\{s, s\}$. In other words, if the truth is $\{f, f\}$ or $\{f, s\}$, there is no incentive to misreport it to be $\{s, f\}$. Thus, only the reportedly failed project needs to be audited in state 1.

Move on to consider partial collusion, which the investors do not take into account when deciding the auditing strategy. In state 2, Ms X obtains $2d_2 - D_2$ if honoring the contracts. She could arrange partial collusion in which she collects $t < d_2$ from E_1 to buy his silence when she declares state 1 to the investors and E_2. With the collusion, she obtains $(t + d_1 - D_1)(1 - l_1)$. Previously, under FI, E_1 gained in net $d_2 - t$ from the collusion, even if her lying had been uncovered, since his pocket was not subject to appropriation by the bank's investors. On the contrary, it is subject now under Conglomeration; thus he obtains $(1 - l_1)(R - t)$ with the collusion, and $R - d_2$ without. The IC for no partial collusion in state 2 is, therefore, $2d_2 - D_2 \geq (1 - l_1)(t + d_1 - D_1)$ for any t such that $(1 - l_1)(R - t) \geq R - d_2$. Equivalently,

$$d_2 + R - D_2 \geq (1 - l_1)(d_1 + R - D_1) \tag{P2h}$$

The IC for no partial collusion in state 1, (P1h), is not binding.

Lastly, the IR to Ms X is (IR-X), the same as under FI, since she obtains the difference between the paid-in and the paid-out in the same way. The entrepreneurs' problem is then:

Problem 2 $\min_{\{d_s;D_s,l_s\}_{s=0,1,2}} 2q(1-q)d_1 + q^2 \cdot 2d_2$, *subject to*

(1): $l_s = l_s^H(\Psi)$ *for* $s = 0, 1, 2$, *where* $\Psi = \{D_s\}_{s=0,1,2}$.

(2): (LL), (P2h), (P1h), (IR-X), and $D_1 \leq D_2$ *(Assumption 3).*

Again, the solution depends on whether (IR-X) is satisfied or not. If not, we have the following. Let $H \equiv \frac{2R}{q^2R+2(1-q)S}$.

Lemma 4 *The optimal mechanism of Conglomeration is:* $D_1 = D_2 = d_1 = d_2 = H; l_0 = \frac{2}{q^2R+2(1-q)S}, l_1 = l_2 = 0,$ *if* $m < \frac{1}{2}q^2H$.

Proof. See the appendix. ■

Again, the mechanism is driven by the trade-off between the auditing costs and the rent to Ms X, and as the former has been assumed large enough, the optimal mechanism triggers auditing only in state 0 ($l_1 = 0$), but gives Ms X net rent of $V^H = q^2H - 2m$.

The difference is that, here $l_0 = \frac{2}{q^2R+2(1-q)S}$ decreases with R, but the counterpart under FI $l_0 = 1$ is independent of R. This is due to the difference in the collateral, which is the pool of the projects under Conglomeration and the bank asset under FI. The higher is R, the more is the collateral under Conglomeration worth and hence the less is the auditing needed. Under FI, if the entrepreneurs want to similarly lower l_0 by increasing the value of the collateral, they have to fill more into the bank asset, which they definitely dislike; rather they pick $l_0 = 1$.

The same line of argument applies for the case where the IR for X is binding. In this case, under FI, the entrepreneurs have to pay more to the bank, which, we showed, raises the bank asset's value and thus pushes down l_0. On the contrary, under Conglomeration, the collateral's value is independent of m, and so is l_0: thus, $l_0 = l^H \equiv \frac{2}{q^2R+2(1-q)S}$, no matter if (IR-X) is binding.

As under FI, $d(= d_1 = d_2)$ is pinned down by the binding (IR-X) for this case: $d = H + (\frac{m}{q} - \frac{qH}{2})$ for $m \geq \frac{1}{2}q^2H$. The increment, $\frac{m}{q} - \frac{qH}{2}$, pays for the part of the monitoring costs that is not offset by the gross rent, q^2H.

To summarize, let $C^H \equiv 2(1-q)^2Cl^H$ denote the expected auditing costs under Conglomeration. From the binding (IR-Ih), $H = \frac{2+C^H}{2q-q^2}$. It follows that $V^H = q^2\frac{2+C^H}{2q-q^2} - 2m$ if $m < \frac{1}{2}q^2H$; otherwise, $V^H = 0$. So $V^H = \max(0, q^2\frac{2+C^H}{2q-q^2} - 2m)$ always. With these notations, the optimal mechanism of Conglomeration is summarized as follows.

Proposition 2 *Under Conglomeration, a successful entrepreneur outlays* $d^H = H + \max(0, \frac{m}{q} - \frac{qH}{2})$*, and the conglomerate is audited only in state 0, each project audited with probability* $l^H = \frac{2}{q^2R+2(1-q)S}$*, and Ms X obtains the net rent of* $V^H = \max(0, q^2\frac{2+C^H}{2q-q^2} - 2m)$.

3.5 Comparisons and the Equilibrium Mode

Let us put aside the mixed mode for a while and compare IF, which is equivalent to mode (2), FI, and Conglomeration. First, by Assumption 3, $m < \frac{(1-q)q^2RC}{(q^2R+2(1-q)S)S}$, which implies $H +$

$\max(0, \frac{m}{q} - \frac{qH}{2}) < \frac{R}{S}$, that is, $d^H < d^I$. Therefore, IF is dominated by Conglomeration. We are left to compare FI to Conglomeration. FI dominates Conglomeration if and only if $d^B \leq d^H$. For this inequality, we have the following lemma.

Lemma 5 *$d^B \leq d^H$ if and only if $C^B \leq C^H$*

Proof. $2qd^X = 2 + V^X + 2m + C^X$ for $X = H$ and B: the accounting equation states that the total outlays of the entrepreneurs exactly cover the investment costs (2) and the rent to Ms X and the monitoring costs and the auditing costs. It follows that $d^B - d^H = V^B - V^H + C^B - C^H$. By Propositions 1 and 2, $V^X = \max(0, q^2 \frac{2+C^X}{2q-q^2} - 2m)$, which implies that $V^B - V^H$ has the same sign as $C^B - C^H$. Therefore, $d^B - d^H$ has the same sign as $C^B - C^H$. ■

As $C^B = (1-q)^2 C_2 l^B(m)$ and $C^H = 2(1-q)^2 C l^H$, $C^B \leq C^H$ if and only if $\frac{l^B(m)}{l^H} \leq \frac{2C}{C_2}$. As $l^H = \frac{2}{q^2 R + 2(1-q)S}$, the inequality is equivalent to

$$\frac{2C}{C_2} \geq \frac{l^B(m)(q^2 R + 2(1-q)S)}{2} \tag{FI-C}$$

Then, by Lemma 5, FI dominates Conglomeration if and only if (FI-C) holds true.

Comparing (G10) and (G21) between the two modes shows the advantage of Conglomeration: (G10b) is tighter than (G10h) and (G21b) is tighter than (G21h). The former has the same RHS as the latter, but a smaller LHS, because $d_1 \leq R$ (for (G10)) or $2d_2 \leq 2R$ (for G21)). The two inequalities reflect the fact that collateral under FI (the bank asset) is always worth less than that under Conglomeration (the pool of the projects), because the bank asset is always a part of the pool of the projects. The higher value of the collateral leads to the less auditing needed. Therefore, Conglomeration has the "Collateral Advantage". On the other hand, FI has the number advantage, measured by $\frac{2C}{C_2}$. (FI-C) shows that FI dominates Conglomeration if and only if so measured number advantage is larger enough.

We turn to the mixed mode now. It is dominated by Conglomeration, by the following intuitive argument. The collateral of the mixed mode, composed of one project plus the intermediary asset within the other project, is worth less than the pool of the two projects. Moreover, in case of default, the investors still have to audit two assets. Therefore, the mixed mode, compared to Conglomeration, has neither Collateral Advantage nor Number Advantage and is dominated by it.

With the last mode dominated, the real race is between FI and Conglomeration. The equilibrium mode is FI if and only if (FI-C) holds true, that is, the number advantage is larger than the threshold $T(m, R) = \frac{l^B(m)(q^2R+2(1-q)S)}{2}$, which measures the collateral advantage of Conglomeration. Obviously $\frac{\partial T}{\partial R} > 0$; intuitively, the larger is R, the higher is the value of the projects and the bigger is the collateral advantage of Conglomeration. $\frac{\partial T}{\partial m}$ has the same sign as $\frac{dl^B(m)}{dm}$, which, by Proposition 1, is negative. Intuitively, the larger is m, the more is needed to pay the bank as compensation, the higher is the value of the bank asset, and the smaller is the relative collateral advantage of Conglomeration to FI. R captures the gross size of the projects and m captures the degree of complexity, as it measures the expenses with which a financial expert acquires soft information on a project.

To summarize,

Proposition 3 *The equilibrium mode is either FI or Conglomeration, depends on whether (FI-C) holds true or not. The chance of FI being the equilibrium mode decreases with the size of the projects and increases with the degree of their complexity.*

We move on to compare FI with Conglomeration for the case of a large number of entrepreneurs. This is for two purposes, to check the robustness of the results of Proposition 3, and to examine the claim of Diamond (1984) that FI is driven by the benefit of diversification, which is the benefit of cross subsidization magnified by the Law of Large Numbers (LLN). The benefit of cross subsidization, we have seen, does not drive FI for the two-entrepreneur case, since it is also implemented by Conglomeration. Will large numbers make any difference?

Intuitively, they will not. They deliver the same good under FI as under Conglomeration. In the two-entrepreneur case, two costs come with the implementation of the benefit of cross subsidization: in state 0, auditing still happens since the liability to investors is not repaid, and in state 2, Ms X obtains rent, since she is then paid by two successful entrepreneurs while one payment suffices to clear the liability to investors. Both costs are worn away when the number of entrepreneurs goes to infinity, because by the LLN, on average the number of successes is almost fixed, and hence both the probability of the liability not being fully repaid and that of Ms X acquiring rents go to 0. This argument is independent of how monitoring is organized, and hence applicable to both FI and Conglomeration. It shows that the two modes are equally good under perfect diversification.

The next section provides a strict analysis and gives a necessary condition for FI to dominate Conglomeration with large but still imperfect diversification.

4 The Case of a Large Number of Entrepreneurs

Now there are N entrepreneurs. Each has an identical and independent project. Ms X, the provider of monitoring service, becomes either the bank that now finances the N projects, or the headquarters of the conglomerate that consists of N division. The costs of auditing the bank is denoted by C_N. We assume $C_N = zN^\alpha$ for some $z > 0$ and $\alpha \leq 1$; that is, the auditing costs do not increase faster than linearly with N. The economy has $N+1$ states, state $s = 0, 1...N$, defined by the number of successful projects and occurring with probability $p_N^s = C_N^s q^s (1-q)^{N-s}$.

Note that the model is isomorphic to that of Diamond (1984). The auditing costs of this model correspond to the non-pecuniary penalties: both are the deadweight loss incurred when default happens, and in both papers, to avoid this loss by utilizing cheap monitoring service is the driving force for the results. Diamond (1984), and the literature following him, only compared FI to IF and showed the incentive costs to be worn out by diversification. This enables FI to dominate IF, which, he claimed, ensures the viability of FI. However, that benefit of diversification is also implemented by Conglomeration, as will be shown, and is thus canceled out in the race between FI and Conglomeration.

What happens under IF is independent of the value of N. Particularly, in the case of large N, a successful entrepreneur outlays $d^I = \frac{R}{S}$, as he did in the case of $N = 2$. Below for the case of large N we first examine FI and Conglomerate, and then compare the two modes.

FI

First examine FI, equipped with the following mechanism. $d_s = d$, $D_s = \begin{Bmatrix} sd, \text{for } s < k \\ kd, \text{for } s \geq k \end{Bmatrix}$, and $l_s = \begin{Bmatrix} 1, \text{for } s < k \\ 0, \text{for } s \geq k \end{Bmatrix}$, for some k. That is, a successful entrepreneur pays the bank with d, independent of s, and the bank repays what it receives to the investors up to kd and is audited definitely in case of failing to fully repay kd. The liability contract of the bank is thus debt, of which the total face value is $F = kd$. This mechanism is the one that was considered by Diamond (1984). It is certainly incentive compatible. It is also partial collusion proof, because the outlay of a successful entrepreneur is independent of the overall state, the report of which Ms X can manipulate. The IR to the investors, (IR-I), is binding.

Lemma 6 *If (IR-X) is not binding, the gross rent to Ms X is at most of order* $\sqrt{N \log N}$.

Proof. See the Appendix. ■

The gross rent is thus not sufficient to cover the total monitoring costs, Nm, when N is large enough. Therefore, (IR-X) is always binding. It follows that each entrepreneur's outlay covers exactly the investment cost (1), the monitoring costs (m), and the average auditing costs. That is,

$$qd = 1 + m + \frac{C_B}{N} \tag{2}$$

Here $C^B = P_d C_N$ is the total auditing costs under FI, with P_d being the probability of default. Suppose P_d goes to 0 (to be verified soon). This, together with $C_N \le zN$, implies that $\frac{C^B}{N} \approx 0$, and hence by (2), $d \approx \frac{1+m}{q}$.[12] The (IR-I) is $(1-P_d)F + P_d Q = N + P_d C_N$, where Q is the expected repayment conditional on default. As $P_d \to 0$, $F \approx N$. In this mechanism, $F = kd$. It follows that $k = \frac{F}{d} \approx \frac{qN}{1+m}$. Then, by the Central Limit Theorem (CLT), the default probability $P_d = \sum_{s \le k-1} p_N^s \approx \Phi(\frac{k-Nq}{\sqrt{Nq(1-q)}}) \approx \Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})$, which indeed approaches 0 quickly. Hence indeed $d \approx \frac{1+m}{q}$, which was the core result of Diamond (1986), namely that on average diversification wipes out average incentive costs and leaves only technical costs (the investment costs and the monitoring costs), and hence FI dominates IF: $\frac{1+m}{q} < \frac{R}{S}$ by Assumption 3.

The auditing costs under FI is $C^B \approx \Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})C_N$.

Conglomeration

We move on to examine Conglomeration. In state s, the conglomerate outlays D_s if honoring the contracts. If it lies that the state is $t < s$, then each of the $s-t$ actually successful but reportedly failed projects is audited with probability l_t, and auditing any one leads to the appropriation of all the projects, worth sR; otherwise it outlays D_t according to the reported state, t. Therefore, the IC for the conglomerate not to misreport state s to be t is

$$(1-(1-l_t)^{s-t}) \cdot sR + (1-l_t)^{s-t} D_t \ge D_s \qquad \text{(Gsth)}$$

Note that even if l_t is small, $(1-(1-l_t)^{s-t})$ could still approach 1 if $s-t$ is large, which represents another advantage of Conglomeration for the case of large N. Let us call it "Spread Advantage": since the collateral of Conglomeration spreads across the N projects, a lie with a large a deviation $(s-t)$ involves many misreportings, which makes the lie vulnerable to be

[12] Hereinafter, the notation $y \approx x$, sometimes denoted as $y = x + o$, means $\frac{y-x}{x} \to 0$ if $x \ne 0$, and $y \to 0$ if $x = 0$. The notation $y = O(x)$ means $y \approx \lambda x$ for some $\lambda > 0$, and $y = o(x)$ means $\frac{y}{x} \to 0$.

detected. The larger the deviation, the bigger the spread advantage, which leads to the following mechanism.

$d_s = \begin{Bmatrix} 0, \text{for } s < k \\ d, \text{for } s \geq k \end{Bmatrix}$, $D_s = \begin{Bmatrix} 0, \text{for } s < k \\ kd, \text{for } s \geq k \end{Bmatrix}$, and $l_s = \begin{Bmatrix} l_s, \text{for } s < k \\ 0, \text{for } s \geq k \end{Bmatrix}$, for some k. For state $t \geq k$, it is the same as that under FI. For states $t < k$, differently, D_t is set to equal 0 to make (Gkth) binding for any $t < k$, which makes maximum use of the spread advantage; $d_t = 0$ is to drain Ms X's rent, as $D_t = 0$. This mechanism is partial collusion proof, since when the true state is $s < k$, Ms X never lies that it is a state $t \geq k$, because she has no way to afford kd, and for $s \geq k$, the outlay of a successful entrepreneur is always d, independent of Ms X's report of the overall state.

For any $t < k$, (Gkth) is binding: $(1-(1-l_t)^{k-t})kR = kd$, which implies $l_t = 1-(1-\frac{d}{R})^{\frac{1}{k-t}}$. These l_t will be difficult to handle. But notice that $\frac{d}{R}\frac{1}{k-t} \leq 1-(1-\frac{d}{R})^{\frac{1}{k-t}} \leq \log(1-\frac{d}{R})^{-1}\frac{1}{k-t}$.[13] As we are interested in finding a necessary condition for FI to dominate Conglomeration, let $l_t = \log(1-\frac{d}{R})^{-1}\frac{1}{k-t}$. As under FI and will be shown, $d \to \frac{1+m}{q}$. Thus $\log(1-\frac{d}{R})^{-1} \to \log(1-\frac{1+m}{qR})^{-1} \equiv \gamma$. In state $t < k$, $N-t$ projects fails, each audited with probability $l_t = \frac{\gamma}{k-t}$. The total auditing costs is thus $C^H = \gamma C \sum_{t \leq k-1} p_N^t \frac{N-t}{k-t}$. It is less than γNCP_d, where $P_d \equiv \sum_{t \leq k-1} p_N^t$ is the probability of default.

As under FI, (IR-X) is binding.[14] Suppose P_d and the average auditing costs, $\frac{C^H}{N}$, go to 0 (to be verified). Then, $d \approx \frac{1+m}{q}$ and $k \approx \frac{N}{d} \approx \frac{qN}{1+m}$ by the same argument as for the case of FI. Thus indeed $P_d \approx \Phi(\frac{-m}{1+m}\sqrt{\frac{qN}{1-q}})$ goes to 0 quickly and so does $\frac{C^H}{N} < \gamma CP_d$, as under FI. Therefore indeed $d \approx \frac{1+m}{q}$, which is the point intuitively obtained proceeding this section, namely that the

[13] That is because $(1-x)\mu \leq 1-x^\mu \leq -\mu \log x$ for $0 < x, \mu \leq 1$. For the former inequality, let $f(x) = 1-x^\mu-(1-x)\mu$. $f(1) = 0$, and $f'(x) = -\mu(x^{\mu-1}-1) < 0$ for $x, \mu < 1$. Therefore, $f(x) > 0$ if $x < 1$. For the latter, let $\beta = x^\mu$ and $f(\beta) = -\log\beta - (1-\beta)$. $f(1) = 0$, and $f' = 1-\frac{1}{\beta} < 0$ for $\beta < 1$. Therefore, $f(\beta) > 0$ if $\beta < 1$.

[14] Actually, the gross rent is of the order $\sqrt{N \log\log N}$, if (IR-X) is not binding. The proof is available upon request.

same benefit of diversification is implemented by Conglomeration.

To compare Conglomeration with FI, note that $C^H < \gamma NC \sum_{s \leq k-1} p_N^s \frac{1}{k-s} = \gamma NCP_d \cdot E(\frac{1}{k-s}|s \leq k-1) \equiv \overline{C^H}$. For the expectation, we have

Lemma 7 $E(\frac{1}{k-s}|s \leq k-1) \to \frac{m}{1-q} \log \frac{1-q+m}{m}$, *when* $N \to \infty$ *and* $\frac{k}{N} \to \frac{q}{1+m}$.

Proof. See the appendix. ■

The Comparison

FI has a chance of dominating Conglomeration only if C^B is no bigger than $\overline{C^H}$, which is equivalent to $C_N \leq \gamma NC \cdot \frac{m}{1-q} \log \frac{1-q+m}{m}$ by this lemma. Substitute $\gamma(m, R) = \log(1 - \frac{1+m}{qR})^{-1}$ and do rearrangement. The inequality is equivalent to

$$\frac{NC}{C_N} \geq [\log(1 - \frac{1+m}{qR})^{-1} \frac{m}{1-q} \log \frac{1-q+m}{m}]^{-1} \qquad \text{(FI-C-N)}$$

That is, FI dominates Conglomeration and thus occurs in equilibrium, only if the Number Advantage is beyond the threshold $T_N = [\log(1 - \frac{1+m}{qR})^{-1} \frac{m}{1-q} \log \frac{1-q+m}{m}]^{-1}$. $\frac{\partial T_N}{\partial R} > 0$ and $\frac{\partial L_N}{\partial m} < 0$.[15] This confirms the robustness of the comparative statics results derived in the two-entrepreneur case.

To sum up,

Proposition 4 *When there are a large number of entrepreneurs, both Conglomeration and FI implement the benefit of diversification. And FI dominates Conglomeration only if (FI-C-N) holds, that is, only if the number advantage is bigger than the threshold, which increases with* R *and decreases with* m*, as for the two-entrepreneur case.*

[15] $\log(1 - \frac{1+m}{qR})^{-1}$ increases with m, and so does $m \log \frac{1-q+m}{m}$, since $\{m \log \frac{1-q+m}{m}\}' = \log \frac{1-q+m}{m} - \frac{1-q}{1-q+m} \equiv \log(1+x) - \frac{x}{1+x} > 0$, where $x = \frac{1-q}{m}$. For the last inequality, let $f(x) \equiv \log(1+x) - \frac{x}{1+x}$; then $f(0) = 0$ and $f' = \frac{1}{1+x} - \frac{1}{(1+x)^2} > 0$ for $x > 0$.

Note that $\lim_{m \to 0} T_N = \infty$. This, however, does not mean that when $m = 0$, FI never dominates Conglomeration in any circumstances, but that it never if C_N is in the order of N. It can be shown that when $m = 0$, FI dominates Conglomeration only if C_N is in the order of at most $\sqrt{N}$.[16] $m = 0$ is a very special case where (IR-X) is never binding; then the optimal k is decided by the trade-off between the auditing costs and the rent to Ms X, as for the two-entrepreneur case (see the discussion following Lemmas 3 and 4).

5 Conclusion and Future Research

This paper proposes that the organization of financing is decided by the allocation of the liability to repay investors. It examines the economics of the liability allocation in a Townsend economy with two entrepreneurs: it is costly to audit their outputs but cheap to monitor. The paper analyzes all the possible modes of financing, each defined by the according allocation of the liability. The real race is between two modes. One is FI, where the liability is taken by the monitor alone, who hence becomes the bank; the other is Conglomeration, where the liability is taken by the conglomerate composed of the monitor and the entrepreneurs, the former running the headquarters to monitor the latter. Conglomeration has the collateral advantage, namely that its collateral is the pool of the projects, which contains as a part and is thus worth more than the bank asset, the collateral of FI; the higher the collateral's value, the less the auditing needed. On the other hand, FI has the number advantage, namely that in case of default, the investors audit *one* bank asset under FI while *many* entrepreneur-projects under Conglomeration, which saves auditing costs. So FI arises in equilibrium if and only if the number advantage is beyond the threshold.

Furthermore, the paper compares FI with Conglomeration for the case of a large number of entrepreneurs. The economy is isomorphic to Diamond (1984). Both Conglomeration and FI

[16] The proof is available upon request.

implement the benefit of diversification, which, therefore, does not drive FI, as Diamond (1984) claimed and the literature well accepted. Rather FI is driven by the number advantage and arises only if the advantage is beyond a threshold.

Both thresholds increase with R, the return rates of a successful project, and decrease with m, the monitoring costs. As R captures the size of the projects and m captures how complex they are, the comparative statics results predict that the smaller the projects, and/or the more complex the projects, the higher the chance they are financed by the bank.

The only difference between FI and Conglomeration is in the collateral, derived from the difference in the liability allocation. In all other respects, the two modes are the same. Under both, funds flow in the same way, passed by the monitor; the information structure is the same, with each entrepreneur knowing of the output of his own, the monitor all the outputs, investors none before auditing; and there are the same incentive and collusion problems. If FI is to be justified from the cost side of Conglomeration, then this paper suggests that the costs must be either specific to Conglomeration or better handled by FI. It is not obvious that such costs exist.

To consider liability allocation provides a general angle of examining the organization of financial markets. We can carry out an analogous analysis for each service that addresses certain market friction, for example, the service of ex ante screening the quality of projects. The provider of screening service can either sell her knowledge directly to investors and thus become the rating agency, or use her knowledge to invest with investors' capital and thus become the bank. The two modes differ, again, in who takes the liability to repay investors.

Moreover, to consider liability allocation may provide a new angle of delineating the boundary of the firm and examining its nature. The literature starting with Grossman and Hart (1986) and Hart and Moore (1990) delineates the boundary of the firm according to the allocation of ownership of physical capital. This paper suggests that it be delineated according to the allocation of liabilities to a third party, for example, the investors. In this paper, therefore, the entrepreneurs and the monitor are within one firm under Conglomeration, and are not under FI. More generally, if party A takes (uncontractible) liabilities of party B's work, then B is an employee of A, while if B himself takes them, he is an independent contractor. Where the incentive to avoid these liabilities matters, we may reach a theory in the manner of Grossman and Hart (1986).

6 Appendices

The Proof of Lemma 3:

It suffices to show the optimal solution without taking into account (IR-X) gives Ms X expected gross rent of q^2B; therefore, if $q^2B > 2m$, (IR-X) is indeed unbinding. First let us make up the two constraints that are unbinding and skipped in the main context. The IC for the bank not to misreport state 2 to be 0 is

(G20b): $l_0 \cdot 2d_2 \geq D_2$.

And in state 1, if Ms X arranges the partial collusion in which she gives ϵ to the failed entrepreneur to buy his silence when declaring state 2, she is not audited ($l_2 = 0$) and obtains $d_2 - D_2 - \epsilon$, which should be no bigger than $d_1 - D_1$ for any $\epsilon > 0$ to disincentivize the collusion. The partial collusion proof in state 1 is thus

(P1b): $d_1 - D_1 \geq d_2 - D_2$

Note that (P2b) is binding in the optimization. Otherwise, consider the mechanism $D_1 = D_2 = d_1 = 2d_2; l_0 = 1, l_1 = l_2 = 0$. It implements the benefit of cross subsidization, but gives no rent to Ms X, which is impossible to achieve by Lemma 2. This mechanism is thus infeasible, but it satisfies all the constraints but (P2b). The binding (P2b) is:

(A1): $2d_2 - D_2 = (1 - l_1)(d_1 + d_2 - D_1)$.

It follows that (P1b) is not binding: $d_1 - D_1 \geq |_{d_1 \geq D_1 \text{ by } (LL)} (1 - l_1)(d_1 - D_1) = |_{(\text{A1})} d_2 - D_2 +$

$l_1 d_2 \geq d_2 - D_2$. For the investors' problem of finding the optimal auditing strategy, (G21b) is binding; otherwise, they would lower l_1, which only loosens (IR-Ib) of their problem (it tightens (P2b), which they do not care, however). The binding (G21b) is

(A2) $D_2 = 2l_1 d_2 + (1 - l_1)D_1$.

The entrepreneurs will choose d_1 such that (G10b) is binding; otherwise, they would lower d_1, which loosens (P2b), and only tightens (P1b), which we saw is not binding. The binding (G10b) is

(A3): $D_1 = l_0 d_1$.

Lastly, (IR-Ib) is binding:

(A4): $q^2 D_2 + 2q(1-q)D_1 = 2 + C_2[1-q)^2 l_0 + 2q(1-q)l_1]$.

(G20b) is to be verified not binding. The entrepreneurs' problem is:

Problem *: $\min B = qd_2 + (1-q)d_1$, s.t. (A1)-(A4).

Substituting D_2 and D_1 of (A1) with (A2) and (A3) respectively, we get a link between d_1 and d_2: $(1-l_1)(d_2 - d_1) = 0$. Then, two cases arise, either $l_1 = 1$, or $d_2 = d_1$. They are examined one by one, and the latter first.

If $d_2 = d_1 = B$, by (A2) $D_1 = l_0 B$ and by (A3) $D_2 = (2l_1 + l_0 - l_0 l_1)B$. Substitute these into (A4), we have $B = \frac{1}{q}\frac{2+C_2[(1-q)^2 l_0 + 2q(1-q)l_1]}{2ql_1 + (2-q)l_0 - ql_0 l_1}$. Applying $(\frac{a+bx}{c+dx})' = \frac{bc-ad}{(c+dx)^2}$, we have $\frac{\partial B}{\partial l_0} = \frac{1}{q}\frac{-2(2+q+ql_1)+2q(1-q)C_2 l_1(ql_1-1)}{[2ql_1+(2-q)l_0-ql_0 l_1]^2}$. Obviously $-2(2+q+ql_1) \leq 0$ and $ql_1 - 1 < 0$. Therefore, $\frac{\partial B}{\partial l_0} < 0$ and the optimal $l_0 = 1$. Similarly, $\frac{\partial B}{\partial l_1}|_{l_0=1} = \frac{-2+(1-q)C_2(3-q)}{[2ql_1+(2-q)l_0-ql_0 l_1]^2} > 0$, where $C_2(1-q)(3-q) > 2$ because $C_2 \geq C \geq \frac{1}{1-q}$, the last inequality derived from Assumption 3: $2(1-q)C - (1-q)C \geq qR - (1-q)C \geq 1$. Therefore, the optimal $l_1 = 0$. Substitute $l_0 = 1$ and $l_1 = 0$ into the formula of B, and we have $B = \frac{2+C_2(1-q)^2}{2q-q^2} = d_2 = d_1$, and into (A2) and (A3) we have $D_1 = D_2 = d_1 = B$.

If $l_1 = 1$, by (A2) $D_2 = 2d_2$. Since $D_2 \leq 2l_0 d_2$ ((G20b)), we get $l_0 = 1$. By (A3) $D_1 = d_1$. That is, all the revenues paid into the bank is passed to investors. Then the LHS of (A4) equals

$2qB$. As $l_0 = l_1 = 1$, the RHS of (A4) equals $2+C_2[(1-q)^2+2q(1-q)]$. Therefore, $B = \frac{2+C_2(1-q^2)}{2q}$ in this case.

Compare the two cases. $\frac{2+C_2(1-q^2)}{2q} > \frac{2+C_2(1-q)^2}{2q-q^2} \Leftrightarrow C_2(1-q)(3-q) > 2$, which has been proved true. Thus, the first case gives the solution of the minimization problem. Therefore, the optimal mechanism, without taking into account (IR-X), is $d_2 = d_1 = D_1 = D_2 = B \equiv \frac{2+C_2(1-q)^2}{2q-q^2}$, $l_0 = 1$ and $l_1 = 0$. (G20b) is satisfied. Ms X obtains $2d_2 - D_2 = B$ in state 2. Therefore, if $q^2B > 2m$, the IR for Ms X is indeed unbinding.

Q.E.D.

The Proof of Lemma 4

It suffices to show the optimal solution without taking into account (IR-X) gives Ms X expected gross rent of q^2H; therefore, if $q^2H > 2m$, (IR-X) is indeed unbinding. First we make up the neglected constraints. If the conglomerate reports state 0 in state 2, with probability $2l_0 - l_0^2$, one project at least is audited, and then the whole collateral, worth $2R$, is lost to the investors (here the implicit off-equilibrium assumption is that they commit to auditing the rest of the collateral whenever uncovering a fraud on one project, in order to appropriate the whole collateral). The IC for the conglomerate not to misreport state 2 to be 0 is

(G20h): $(2l_0 - l_0^2) \cdot 2R \geq D_2$.

The partial collusion proof in state 1 is the same as (P1b):

(P1h): $d_1 - D_1 \geq d_2 - D_2$.

We first pin down $l_0^H(D_1, D_2)$ and $l_1^H(D_1, D_2)$. (G21h) is binding, for the investors to minimize l_1, which pins down $l_1^H = \frac{D_2-D_1}{2R-D_1}$. l_0 is present in both (G10h) and (G20h), which implies $l_0 \geq \frac{D_1}{R}$ and $l_0 \geq 1 - \sqrt{1 - \frac{D_2}{2R}}$ respectively. Thus the minimum $l_0^H = \max(\frac{D_1}{R}, 1 - \sqrt{1 - \frac{D_2}{2R}})$. Thus,

(B1): $(l_0^H, l_1^H) = (\max(\frac{D_1}{R}, 1 - \sqrt{1 - \frac{D_2}{2R}}), \frac{D_2-D_1}{2R-D_1})$.

We then move on to pin down Ms X's rent. Let $m_1 = d_1 - D_1$ (the rent to Ms X in state 1), $m_2 = 2d_2 - D_2$ (the rent in state 2), and $V = 2q(1-q)m_1 + q^2m_2$ (the total gross rent). Using these notations, (P2h) becomes $m_2 \geq 2(1-l_1)m_1 + D_2 - 2l_1R$, and (P1h) becomes $m_1 \geq \frac{m_2-D_2}{2}$. m_i is nonnegative by (LL). To minimize V, $m_1 = 0$ and $m_2 = D_2 - 2l_1R$. That is, (P2h) is binding and (P1h), equivalent to $0 \geq -l_1R$, is unbinding. The m_2 so obtained is nonnegative: $D_2 - 2l_1R = \frac{(2R-D_2)D_1}{2R-D_1} \geq 0$ by (B1). Thus,

(B2): $(m_1, m_2) = (0, D_2 - 2l_1R)$.

Lastly, the (IR-Ih) is binding. Substitute (B1) into the binding (IR-Ih),

(B3): $q^2D_2 + 2q(1-q)D_1 = 2 + 2[(1-q)^2 l_0^H(D_1, D_2) + q(1-q)l_1^H(D_1, D_2)]C$.

(B3) implicitly defines a function $D_2(D_1)$. $\{(D_1, D_2)|D_2 = D_2(D_1)\}$ is then the set of all feasible securities. If the repayment in state 1 (D_1) decreases, as compensation, the repayment in state 2 (D_2) has to increase, that is, $D_2' < 0$. Here, and for the rest of the proof, "′" represents the full derivative with respect to D_1.

Last, let $C^H \equiv 2C[(1-q)^2 l_0 + q(1-q)l_1]$ be the total auditing costs. Then the total financial costs consist of the investment costs (2) and X's rent (V) and C^H. The entrepreneurs' problem becomes $\min_{D_1, D_2} V + C^H$, s.t. (B1)-(B3).

Lemma 4 asserts that the minimization happens at $D_1 = D_2$. As $D_1 \leq D_2$ is assumed, to prove the lemma, it suffices to show that $(V + C^H)' < 0$ everywhere. As $l_0 = \max(1 - \sqrt{1 - \frac{D_2}{2R}}, \frac{D_1}{R})$, we consider two cases depending whether $1 - \sqrt{1 - \frac{D_2}{2R}} \leq \frac{D_1}{R}$ or not.

Consider first the case where $1 - \sqrt{1 - \frac{D_2}{2R}} \leq \frac{D_1}{R}$ and thus $l_0 = \frac{D_1}{R}$. To get an explicit trade-off between the rent (V) and the auditing costs (C^H), notice that $V = q^2m_2 = q^2D_2 - 2q^2Rl_1 \Rightarrow q^2D_2 = V + 2q^2Rl_1$. And $l_0 = \frac{D_1}{R} \Rightarrow D_1 = l_0R$. Let $l \equiv ql_1 + (1-q)l_0$ and substitute these into (B3), we get $V + 2q^2l_1R + 2q(1-q)l_0R = 2 + 2(1-q)Cl \Leftrightarrow V + 2qRl = 2 + 2(1-q)Cl \Leftrightarrow V + 2(qR - (1-q)C)l = 2$. As $l = \frac{C^H}{2(1-q)C}$, it follows that $V + \frac{qR-(1-q)C}{(1-q)C}C^H = 2$. Then $(V + C^H)' = \frac{qR-2(1-q)C}{qR-(1-q)C}V'$. By Assumption 2, $qR - 2(1-q)C \leq 0$. Thus $(V + C^H)' < 0 \Leftrightarrow V' > 0$. $V = q^2m_2$ and $m_2 = D_2 - 2l_1R = \frac{(2R-D_2)D_1}{2R-D_1}$. Then, $V' \propto m_2' > 0$, since $\frac{\partial m_2}{\partial D_1} > 0$, $\frac{\partial m_2}{\partial D_2} < 0$, and $D_2' < 0$. Hence $(V + C^H)' < 0$ in this case.

Consider the case where $1-\sqrt{1-\frac{D_2}{2R}} > \frac{D_1}{R}$ and thus $l_0 = 1-\sqrt{1-\frac{D_2}{2R}}$. Then $\frac{dl_0}{dD_2} > 0$, and $l_0' = \frac{dl_0}{dD_2}D_2' < 0$. As $V + C^H = q^2D_2 - 2q^2Rl_1 + 2C[(1-q)^2l_0 + q(1-q)l_1] = q^2D_2 - 2qSl_1 + 2C(1-q)^2l_0$ (remember $S = qR-(1-q)C$), $(V+C^H)' = q^2D_2' - 2qSl_1' + 2C(1-q)^2l_0' < q^2D_2' - 2qSl_1'$. In order to show $(V+C^H)' < 0$, it suffices to prove that $q^2D_2' - 2qSl_1' < 0$. By (B3), $q^2D_2' + 2q(1-q) = 2(1-q)^2Cl_0' + 2q(1-q)Cl_1'$. q^2D_2' is smaller than the LHS of this equation, and the RHS is smaller than $2q(1-q)Cl_1'$, as $l_0' < 0$. Therefore, $q^2D_2' < 2q(1-q)Cl_1'$. Then, $q^2D_2' - 2qSl_1' < 2q(1-q)Cl_1' - 2qSl_1' = 2q[(1-q)C - S]l_1' < 0$; for the last inequality we applies $(1-q)C - S = 2(1-q)C - qR \geq 0$ and $l_1' < 0$ ($l_1 = \frac{D_2-D_1}{2R-D_1}$ by (B1) so that $\frac{\partial l_1}{\partial D_1} < 0$ and $\frac{\partial l_1}{\partial D_2} > 0$). Therefore, $(V+C^H)' < 0$ in this case.

To sum up, the solution to the entrepreneurs' problem is $D_1 = D_2 = H$. Accordingly, by (B1), $l_1 = 0; l_0 = \frac{D}{R}$. Substituting all these into (B3), we have $[q^2 + 2q(1-q)]H = 2 + 2C(1-q)^2\frac{H}{R}$, which implies $H = \frac{2R}{q^2R+2(1-q)S}$. By (B2), $m_1 = 0; m_2 = D_2 = H$. Then $d_1 = D_1 + m_1 = H; d_2 = \frac{m_2+D_2}{2} = H$. Indeed the gross rent to Ms X is q^2H; therefore, if $q^2H > 2m$, (IR-X) is unbinding and the optimal mechanism is as specified above.

Q.E.D.

Hereinafter, p_N^s is denoted as p_s for simplicity, if without confusion.

The Proof of Lemma 6

The IR for the investors is binding, as follows:

(IR-I): $(k\sum_{s\geq k} p_s + \sum_{s\leq k-1} sp_s)d = N + C_N \sum_{s\leq k-1} p_s$.

(IR-I) determines a function $d(k)$. When (IR-X) is not binding, the optimal k that minimizes d is decided by the trade-off between the auditing costs and the rent to Ms X: auditing happens in states $s \leq k-1$ and Ms X receives rent in states $s > k$.

To find the first order condition of the minimization, let us simplify (IR-I). Divide both sides by N. By the Central Limit Theorem (CLT), $\frac{s-Nq}{\sqrt{Nq(1-q)}} \sim N(0,1)$, with the dense and cumulative distribution functions being $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ and $\Phi(x)$ respectively. And let $k = Nq + h\sqrt{Nq(1-q)}$. Then, $d(k)$ leads the following function.

(C1): $d(h) = \frac{1+\frac{C_N}{N}\Phi(h)}{q+\sqrt{\frac{q(1-q)}{N}}(h(1-\Phi(h))-\phi(h))}$.

Given N, the optimal h satisfies the first order condition

(C2): $\frac{C_N}{N}\phi(h)[q+\sqrt{\frac{q(1-q)}{N}}(h(1-\Phi(h))-\phi(h))] = [1+\frac{C_N}{N}\Phi(h)]\sqrt{\frac{q(1-q)}{N}}(1-\Phi(h))$.

Remember $C_N = zN^\alpha$ for some $\alpha \in (0,1]$. Suppose $\frac{C_N}{N}\Phi(h) = o(1)$, which is obvious for $\alpha < 1$ and to be verified for $\alpha = 1$. Then, the RHS of (C2) $\approx \sqrt{\frac{q(1-q)}{N}}(1-\Phi(h))$. Suppose $\frac{h}{\sqrt{N}} = o(1)$ (to be verified later), which implies that the LHS $\approx \frac{C_N}{N}\phi(h)q$. (C2) asymptotically becomes:

(C3): $\frac{qz}{\sqrt{q(1-q)}}N^{\alpha-0.5} = \frac{1-\Phi(h)}{\phi(h)}$.

Claim: The solution of (C3) is $h^* = \left\{\begin{array}{ccc} \frac{\sqrt{q(1-q)}}{qz}N^{0.5-\alpha}+o & & \alpha<0.5 \\ \widehat{h} & \text{if} & \alpha = 0.5 \\ -\sqrt{(2\alpha-1)\log N}+o & & \alpha > 0.5 \end{array}\right\}$, where $\widehat{h}$ is a constant.

Proof: If $\alpha = 0.5$, (hb) becomes $\frac{qz}{\sqrt{q(1-q)}} = \frac{1-\Phi(h)}{\phi(h)}$. $\lim_{h\to-\infty}\frac{1-\Phi(h)}{\phi(h)} = \infty$ and by L'Hospital's rule, $\lim_{h\to+\infty}\frac{1-\Phi(h)}{\phi(h)} = \lim_{h\to+\infty}\frac{-\phi(h)}{-\phi(h)h} = 0$. $\frac{1-\Phi(h)}{\phi(h)}$ is decreasing, since $\{\frac{1-\Phi(h)}{\phi(h)}\}' = \frac{(1-\Phi(h))h-\phi(h)}{\phi(h)}$ and $(1-\Phi(h))h-\phi(h) = -\int_h^\infty(t-h)\phi(t)dt < 0$. Therefore the equation has a unique solution, $\widehat{h}$.

If $\alpha < 0.5$, (C3) implies that when $N\to\infty$, $\frac{1-\Phi(h)}{\phi(h)} = \frac{qz}{\sqrt{q(1-q)}}N^{\alpha-0.5} \to 0$. Thus $h\to+\infty$. It follows that $\frac{1-\Phi(h)}{\phi(h)} \approx \frac{1}{h}$, as $\lim_{h\to+\infty}h(\frac{1-\Phi(h)}{\phi(h)}-\frac{1}{h}) = \lim_{h\to+\infty}\frac{h(1-\Phi(h))}{\phi(h)} - 1|_{\text{L'Hospital}} = \lim_{h\to+\infty}\frac{1-\Phi(h)}{-h\phi(h)} = \lim_{h\to+\infty}\frac{1}{-h}\frac{1-\Phi(h)}{\phi(h)} = 0$. Therefore, $\frac{qz}{\sqrt{q(1-q)}}N^{\alpha-0.5} \approx \frac{1}{h} \Rightarrow h \approx \frac{\sqrt{q(1-q)}}{qz}N^{0.5-\alpha}$.

If $\alpha > 0.5$, (hb) implies that when $N\to\infty$, $\frac{1-\Phi(h)}{\phi(h)}\to\infty$, and thus $h\to-\infty$, which implies $1-\Phi(h)\to 1$. Therefore, $\frac{qz}{\sqrt{q(1-q)}}N^{\alpha-0.5} = \phi(h)^{-1} = \sqrt{2\pi}e^{\frac{h^2}{2}}$, as $\phi(h) = \frac{1}{\sqrt{2\pi}}e^{-\frac{h^2}{2}}$. Take log operation on both sides$\Rightarrow \frac{h^2}{2} \approx (\alpha-0.5)\log N \Rightarrow h \approx -\sqrt{(2\alpha-1)\log N}$.

q.e.d.

Let $x = \frac{C_N}{N}\Phi(h^*)$ and $y = \sqrt{\frac{q(1-q)}{N}}(h(1-\Phi(h))-\phi(h))|_{h=h^*}$, so $d(h^*) = \frac{1+x}{q+y}$ by (C1). Then, $x = o(1)$ indeed; for $\alpha = 1$, $x = O(\Phi(h^*)) = |_{\text{the claim}} O(\Phi(-\sqrt{\log N})) = o(1)$. And by the claim, $y = O(\frac{h}{\sqrt{N}}) = o(1)$. Therefore, $d \to \frac{1}{q}$.

The gross rent to Ms X is $V_N = \sum_{s \geq k} d(s-k)p_s$. Apply the CLT, $d \approx \frac{1}{q}$ and $k = Nq + h\sqrt{Nq(1-q)}$, and let $s = Nq + t\sqrt{Nq(1-q)}$. We have $V_N \approx \frac{\sqrt{Nq(1-q)}}{q}\int_{h^*}^{\infty}(t-h^*)\phi(t)dt$. The integration equals $\phi(h^*) - h^*(1-\Phi(h^*))$. It converges to $\phi(\widehat{h}) - \widehat{h}(1-\Phi(\widehat{h}))$, for $\alpha = 0.5$, and $\approx -h^*$ for $\alpha > 0.5$ ($h^* \to -\infty$); if $a < 0.5$, as $h^* > 0$, the integration is smaller than $\phi(h^*)$, which multiplied by $\sqrt{N}$ goes to 0, as $h^* = O(N^{0.5-\alpha})$. Therefore, $V_N = \left\{ \begin{array}{ccc} o(1) & & \alpha < 0.5 \\ O(\sqrt{N}) & \text{if} & \alpha = 0.5 \\ O(\sqrt{(2\alpha-1)N\log N}) & & \alpha > 0.5 \end{array} \right\}$. This gives the order of the gross rent to Ms X, which is at most in the order of $\sqrt{N \log N}$.

Q.E.D.

The Proof of Lemma 7

As $\frac{k}{N} \to \frac{q}{1+m}$, it suffices to prove the lemma for $N = k\frac{1+m}{q}$. Let $\theta = \frac{q}{1+m} = \frac{k}{N}$. An intuitive proof of the lemma is as follows. $E(\frac{1}{k-s}|s \leq k-1) = \frac{p_{k-1}\cdot 1 + p_{k-2}\cdot\frac{1}{2} + \ldots + p_0\frac{1}{k}}{p_{k-1}+p_{k-2}+\ldots+p_0} = \frac{1+\frac{p_{k-2}}{p_{k-1}}\cdot\frac{1}{2}+\ldots\frac{p_0}{p_{k-1}}\cdot\frac{1}{k}}{1+\frac{p_{k-2}}{p_{k-1}}+\ldots\frac{p_0}{p_{k-1}}}$. For given N, $\frac{p_{k-i}}{p_{k-1}} = \frac{C_N^{k-i}q^{k-i}(1-q)^{N-k+i}}{C_N^{k-1}q^{k-1}(1-q)^{N-k+1}} = (\frac{1-q}{q})^{i-1}\frac{(k-1)\cdot(k-2)\cdot\ldots(k-i+1)}{(N-k+i)\cdot(N-k+i-1)\cdot\ldots(N-k+2)}$, where $C_N^i = \frac{N!}{i!(N-i)!}$ is the number of combinations. Given that k and N are large, $\frac{k-1}{N-k+i} \approx \frac{k-2}{N-k+i-1} \approx \ldots\frac{k-i+1}{N-k+2} \approx \frac{k}{N-k} = \frac{\theta}{1-\theta}$. Then $\frac{p_{k-i}}{p_{k-1}} \approx (\frac{(1-q)\theta}{q(1-\theta)})^{i-1} \equiv \lambda^{i-1}$, where $\lambda \equiv \frac{(1-q)\theta}{q(1-\theta)} < 1$ as $\theta < q$. Then, $E(\frac{1}{k-s}|s \leq k-1) \approx \frac{1+\lambda\cdot\frac{1}{2}+\ldots\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+\ldots\lambda^{k-1}} = \frac{1-\lambda}{\lambda}\frac{\int_0^{\lambda}\frac{1-t^k}{1-t}dt}{1-\lambda^k} \to \frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda}$ when $k \to \infty$.

For a strict proof, we are going to show that for the comparison between $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1)$ and $\frac{1-\lambda}{\lambda}\log\frac{1}{1-\lambda}$, both "$\geq$" and "$\leq$ " hold true, so that they must be equal. For the " $\geq$ " part, note that $\frac{p_{k-i+1}}{p_{k-i}} = \frac{q}{1-q}\cdot\frac{k-i+1}{N-k+i} < \frac{qk}{(1-q)(N-k)} = \lambda$ for any $i = 2, 3\ldots k$. The following lemma is useful to establish " $\geq$ ".

Lemma A2: If $\frac{a_{i+1}}{a_i} < \lambda$, then $\frac{a_1+a_2\frac{1}{2}+\ldots a_k\frac{1}{k}}{a_1+a_2+\ldots a_k} \geq \frac{1+\lambda\cdot\frac{1}{2}+\ldots\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+\ldots\lambda^{k-1}}$.

Proof: By mathematical induction. For $k = 1$, the inequality surely holds true. Assume that the lemma holds true for a given k. Consider the case for $k + 1$. Let $V_k = \frac{a_1+a_2\frac{1}{2}+...a_k\frac{1}{k}}{a_1+a_2+...a_k}$ and $W_k = \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$. By the induction assumption $V_k \geq W_k$. Both V_k and W_k are a convex combination of $1, \frac{1}{2}, ..., \frac{1}{k}$, and are thus bigger than $\frac{1}{k+1}$. Notice that $\frac{a_{k+1}}{a_1+a_2+...a_{k+1}} < \frac{\lambda^k}{1+\lambda+...\lambda^k}$, as it $\Leftrightarrow \frac{a_1}{a_{k+1}} + \frac{a_2}{a_{k+1}} + ...1 > \lambda^{-k} + \lambda^{-(k-1)} + ...1$, which is true because $\frac{a_i}{a_{k+1}} = \frac{a_i}{a_{i+1}} \cdot \frac{a_{I+1}}{a_{i+2}} ... \frac{a_k}{a_{k+1}} > (\frac{1}{\lambda})^{k+1-i}$ for any $i = 1, 2, .., k$. Then $V_{k+1} = \frac{a_1+a_2+...a_k}{a_1+a_2+...a_{k+1}} V_k + \frac{a_{k+1}}{a_1+a_2+...a_{k+1}} \frac{1}{k+1} > \frac{1+\lambda+...\lambda^{k-1}}{1+\lambda+...\lambda^k} V_k + \frac{\lambda^k}{1+\lambda+...\lambda^k} \frac{1}{k+1} \geq \frac{1+\lambda+...\lambda^{k-1}}{1+\lambda+...\lambda^k} W_k + \frac{\lambda^k}{1+\lambda+...\lambda^k} \frac{1}{k+1} = W_{k+1}$, where for the first inequality we apply $V_k > \frac{1}{k+1}$ and $\frac{a_{k+1}}{a_1+a_2+...a_{k+1}} < \frac{\lambda^k}{1+\lambda+...\lambda^k}$. q.e.d.

By Lemma A2, $E(\frac{1}{k-s}|s \leq k-1) = \frac{p_{k-1}\cdot 1+p_{k-2}\cdot\frac{1}{2}+...+p_0\frac{1}{k}}{p_{k-1}+p_{k-2}+...+p_0} \geq \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{k-1}\cdot\frac{1}{k}}{1+\lambda+...\lambda^{k-1}}$. Let N and $k = \frac{q}{1+m}N$ go to ∞ on both sides, and we have $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1) \geq \frac{1-\lambda}{\lambda} \log \frac{1}{1-\lambda}$.

To prove $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1) \leq \frac{1-\lambda}{\lambda} \log \frac{1}{1-\lambda}$, let us restore notation p_N^s for its simplification p_s. For any $L < k$, $\frac{p_N^{k-1}\cdot 1+p_N^{k-2}\cdot\frac{1}{2}+...+p_N^0\frac{1}{k}}{p_N^{k-1}+p_N^{k-2}+...+p_N^0} < \frac{p_N^{k-1}\cdot 1+p_N^{k-2}\cdot\frac{1}{2}+...+p_N^{k-L}\frac{1}{L}}{p_N^{k-1}+p_N^{k-2}+...+p_N^{k-L}}$, because the former is the convex combination of the latter and the smaller terms, $\frac{1}{L+1}, \frac{1}{L+2} ... \frac{1}{K}$. For this inequality, keep L fixed and let N (and $k = \frac{q}{1+m}N$) go to infinity. Then the left hand side goes to $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1)$. The right hand side goes to $\frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{L-1}\cdot\frac{1}{L}}{1+\lambda+...\lambda^{L-1}}$, because $\frac{p_N^{k-i}}{p_N^{k-1}} = (\frac{1-q}{q})^{i-1} \frac{(k-1)\cdot(k-2)\cdot...(k-i+1)}{(N-k+i)\cdot(N-k+i-1)\cdot...(N-k+2)} \to \lambda^{i-1}$ for any i no bigger than the given L. Therefore, for any given L, $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1) \leq \frac{1+\lambda\cdot\frac{1}{2}+...\lambda^{L-1}\cdot\frac{1}{L}}{1+\lambda+...\lambda^{L-1}}$. Let L go to infinity, $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1) \leq \frac{1-\lambda}{\lambda} \log \frac{1}{1-\lambda}$.

Therefore, $\lim_{N\to\infty} E(\frac{1}{k-s}|s \leq k-1) = \frac{1-\lambda}{\lambda} \log \frac{1}{1-\lambda}$. Substitute $\lambda = \frac{(1-q)\theta}{q(1-\theta)} = \frac{(1-q)\frac{q}{1+m}}{q\frac{1+m-q}{1+m}} = \frac{1-q}{1-q+m}$. Then $\frac{1-\lambda}{\lambda} \log \frac{1}{1-\lambda} = \frac{m}{1-q} \log \frac{1-q+m}{m}$.

Q.E.D.

References

[1] P. Bond, Bank and Nonbank Financial Intermediation, J. Finance, LIX (5) (2004) 2489-2529.

[2] D. Diamond, Financial Intermediation and Delegated Monitoring, Rev. Econ. Stud. 51 (3) (1984) 393-414.

[3] D. Diamond, Financial Intermediation and Delegated Monitoring: A Simple Case, Econ. Quart. Federal Reserve Bank of Richmond, 82 (1996).

[4] J. Diaz-Gimenez, E. Prescott, T. Fitzgerald, F. Alvarez, Banking in computable general equilibrium economies, *J. Econ. Dynamics & Control*, 16, 533-559 (1992).

[5] S. Grossman, O. Hart, The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Ownership, *J. of Polit. Economy*, 94, 691-719 (1986).

[6] O. Hart, J. Moore, Property Rights and the Nature of the Firm, *J. of Polit. Economy*, 98, 1119-1158 (1990).

[7] M. Hellwig, Financial Intermediation with Risk Aversion, Rev. Econ. Stud. 67 (4) (2000) 719-742.

[8] B. Holmstrom, J. Tirole, Financial Intermediation, Loanable Funds, and the Real Sector, Quart. J. Econ. CXII (1) (1997) 663-691.

[9] G. Gorton, A. Winton, Financial Intermediation, mimeo, 2002, http://fic.wharton.upenn.edu/fic/papers/02/0228.pdf .

[10] S. Krasa, A. P. Villamil, Monitoring the Monitor: an Incentive Structure for a financial intermediary, J. Econ. Theory 57 (1) (1992) 197-221.

[11] D. Mookherjee, I. Png, Optimal Auditing, Insurance, and Redistribution, Quart. J. Econ. 104 (2) (1989) 399-415.

[12] R. M. Townsend, Optimal Contracts and Competitive Markets with Costly State auditing, J. Econ. Theory 21 (1979) 265-293.

[13] S. D. Williamson, Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing, J. Monetary Econ. 18 (2) (1986) 159-179.

[14] A. Winton, Costly State auditing and Multiple Investors: the Role of Seniority, Rev. Finan. Stud. 8 (1) (1995) 91-123.