# Aggregate Uncertainty, Money and Banking*

Hongfei Sun†

Department of Economics
University of Toronto

First version: September, 2004
This version: May 30, 2005

**Abstract**

This paper studies the problem of monitoring the monitor in a model of money and banking with aggregate uncertainty. It shows that when inside money is required as part of the bank loan repayment, a market of inside money is created at the repayment stage and generates information-revealing price that helps discipline the bank. As a result, the bank truthfully reveals its solvency and no cost of monitoring the bank needs to be incurred. Inside money contributes to banking by not only providing liquidity to the economy, but helping improve the efficiency of delegated monitoring as well.

Keywords: Money, Banking, Aggregate Uncertainty

JEL Classifications: E00, G21, D82

*I am grateful to Shouyong Shi for kind encouragement and an enormous amount of valuable comments and suggestions. I thank David Andolfatto, Andrés Erosa, Miquel Faig, Diego Restuccia and Randall Wright for helpful insights and detailed comments. I have also benefited from conversations with seminar participants at the University of Toronto, the 2005 Midwest Macroeconomics Meetings, and the 2005 Annual Meeting of the Canadian Economic Association. This research is partly supported by Shouyong Shi's Bank of Canada Fellowship. However, the opinion expressed here is my own and it does not reflect the view of the Bank of Canada. All errors are my own.

†Mailing address: 150 St. George St., University of Toronto, Toronto, Ontario M5S 3G7, Canada. Telephone: 1-416-595-9396. E-mail address: hf.sun@utoronto.ca.

# 1 Introduction

This paper addresses the problem of *monitoring the monitor* in a model of money and banking with aggregate uncertainty. The problem of monitoring the monitor is "a fundamental problem at the root of intermediation theory" (Gorton and Winton, 2003). It challenges the notion of delegated monitoring: The purpose of delegating the task of monitoring to a single agent (known as the *delegated monitor* or the *intermediary*, which typically refers to a bank) is to avoid duplication of lenders' monitoring costs. However, it does not fundamentally resolve this issue. In principle, even with delegation individual lenders may still need to monitor the delegated monitor, in which case duplication of efforts remains. Therefore, when the cost of monitoring the monitor is present, it is not clear at all whether intermediation dominates direct lending.

My work is complementary to the existing literature on the problem of monitoring the monitor. Diamond (1984) and Williamson (1986) show that this problem can be solved by perfect diversification. That is, there is no need to worry about monitoring the intermediary of an infinite size (in the sense that it can create a risk-free portfolio). However, one is naturally concerned about the fact that the real-world financial intermediaries cannot perfectly diversify the risk of their portfolios. Krasa and Villamil (1992a, b) prove that under fairly general conditions, intermediation can dominate direct lending for a sufficiently large bank. Also in a finite intermediated economy, Winton (1995) demonstrates that bank capitalization, in addition to portfolio diversification, can help banking arrangements dominate direct lending. All these works share the common feature that they are models of banking without any particular role of money.

In contrast, money is essential in my model because it overcomes both the friction of no double coincidence of wants and the friction of limited communication. There is yet another friction in the economy: The solvency of an individual borrower is private information, which makes delegated monitoring potentially useful so as to centralize the task of auditing borrowers. The process of financial intermediation is characterized by the bank's issuing inside money (*i.e.* banknotes), making loans, collecting loan repayments, monitoring borrowers and redeeming banknotes on demand. Furthermore, there is aggregate uncertainty and the bank's income is private information. Thus, banknote-redeemers find it necessary to monitor the bank when it defaults on redemption, which naturally introduces the problem of monitoring the monitor.

As it turns out, for banks with undiversifiable risk, the problem of monitoring the monitor can be solved by inside money being required as part of the loan repayment. According to this mechanism, solvent borrowers must trade output for inside money at the loan repayment stage. The resulting market prices aggregate information of all solvent borrowers and fully reveal the income level of the bank. The note-redeemers no longer need to verify it when the bank announces its insolvency. Furthermore, it is in the bank's own interest to implement the above mechanism. That is, when designing the optimal loan contract, the bank voluntarily "ties its own hands" and provides a market mechanism to prevent itself from cheating on notes redemption. With this contract, there is no need to compensate potential note-redeemers for the cost of monitoring the bank. Therefore, the bank gains its edge in competitive banking by being able to require a lower level of loan repayment, which is certainly desirable to the borrowers.

The economic audience has seen quite a few examples of the phenomenon that additional trading opportunities may reduce welfare. The earliest example was given by Hart (1975) who shows that adding a spot market at a point in time may shrink the space of marketed claims and make agents worse off. Jacklin (1987) studies the role of demand deposit in risk sharing and shows that the introduction of new trading possibilities may break down the incentive compatibility and eliminate potential risk sharing. More recently, in economies with asymmetric information, Dow (1998) and Marin and Rahi (2000) provide examples of welfare-reducing financial innovation (*i.e.* introduction of new securities into the incomplete asset markets) which are driven by revelation of information through prices. In contrast to all of the above, my paper presents a model in which markets are a favorable instrument for incentives of truthful revelation. By introducing a spot market to the loan repayment stage, it becomes incentive compatible for the bank not to misreport its solvency and the revelation of information through prices is unambiguously welfare-improving.

Moreover, the model has the following implications on the contribution of inside money to banking: On one hand, inside money helps create banking revenues. The bank provides liquidity to the economy by issuing inside money and making loans out of it. In the meantime, it performs monitoring services when the loan repayments are due. In return for these functions, the bank is entitled to the claims on borrowers.

On the other hand, inside money being involved in loan repayment reduces the overall cost of banking by allowing the bank to truthfully reveal its solvency. This second role of inside money seems novel relative to the literature and it also distinguishes itself from the role of outside money (*a.k.a.* publicly-issued circulating liabilities). With multiple banks, inside money continues to solve the problem of monitoring the bank because each bank can make its banknotes distinguishable from notes of other banks. By requiring its own banknotes to be part of the repayments, a bank induces a distinctive market of goods for its banknotes at the repayment stage. In this case, there will be as many goods markets and prices as the number of banks, each fully revealing the corresponding bank's income. However, this cannot be done with outside money. That is, outside money can only induce one goods market and the resulting price is not sufficient for differentiating the solvency of individual banks.

The model in this paper follows the framework of Andolfatto and Nosal (2003), where money is motivated by no double coincidence of wants and limited communication friction and intermediation is invoked by limited information friction. A critical assumption in their paper is that despite the finite number of borrowers, there is no aggregate risk in the total output produced by borrowers. This assumption allows the bank to perfectly detect any lying about the borrowers' output as a group. As a result, the bank does not monitor borrowers along the equilibrium path. In contrast, my paper incorporates the aggregate uncertainty and studies banking with nontrivial default risk. In effect, the bank conducts monitoring in equilibrium and the economy is characterized by chances of equilibrium bankruptcy, which had not been previously established in the literature of money and banking. (For a detailed review on recent literature that attempts to integrate theories of money and banking, see Andolfatto and Nosal [2003]).

The remainder of the paper is organized as follows. Section 2 introduces the environment of the model. Section 3 studies the bank loan contract that solves the problem of monitoring the monitor. Section 4 compares various loan contracts and shows that the contract that solves the information problem of a finite bank delivers the optimal banking equilibrium. Section 5 addresses robustness of the result. Section 6 concludes the paper.

## 2 The Model

The model adopts the physical structure in Andolfatto and Nosal (2003). There are three islands in the economy, namely $A$, $B$ and $C$. Each island is populated with a finite number $N$ of individuals. The economy lasts for four periods, where $t = 0, 1, 2, 3$. The physical environment is such that not all individuals can communicate with each other across islands. The timing of communication is the following: at $t = 0$, no communication is available across islands; at $t = 1$, people from islands $A$ and $B$ can freely visit each other; at $t = 2$, islands $B$ and $C$ are in communication; and lastly at $t = 3$, agents from islands $C$ and $A$ can get in touch. Traveling agents return to their native island at the end of their traveling date.

At date 0, each individual in the economy is endowed with a project. Projects on different islands differ in their dates of production. At date date $t \geq 1$, there is only one "harvesting" island. Type $B$ projects (*i.e.* those of individuals on island $B$) deliver output at $t = 1$, type $C$ projects at $t = 2$, and type $A$ projects at $t = 3$. Goods are perfectly divisible but they are perishable across periods. For simplicity, let me assume that the cost of production is zero.

The output of a project is random, being $y > 0$ with probability $1 - \lambda$ and 0 with probability $\lambda$. The parameter $\lambda \in (0, 1)$ is publicly known, but the realization of output is the producer's private information. An outsider of the project can find out the project's output level by costly state verification. The auditing process costs the auditor $\mu > 0$ utils per project.

I assume that output realizations are independent across projects. Since the total number of projects $N$ is finite, the level of aggregate output on each island is also random. The number of successful projects on a harvesting island follows a binomial distribution $ß(N, 1-\lambda)$. To simplify the analysis, I assume that $N$ is large enough so that ß can be approximated by a *truncated normal distribution*,[1] with the following density function:

$$f(n) = \begin{cases} \frac{g(n)}{1-\int_{-\infty}^{0} g(n)dn - \int_{N}^{\infty} g(n)dn}, & \text{if} \quad n \in [0, N] \\ 0, & \text{if} \quad n \in (-\infty, 0) \cup (N, \infty) \end{cases},$$

where $n$ is the number of successful projects on an island and

$$\begin{aligned} g(n) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(n-(1-\lambda)N)^2}{2\sigma^2}\right\}, \\ \sigma &= (1-\lambda)\lambda N. \end{aligned}$$

Define $F(n) \equiv \int_{-\infty}^{n} f(\tau)\, d\tau$ as the cumulative function of the truncated normal distribution. Project risk is realized prior to the arrival of any traveling agent at the beginning of each day.

Agents' preferences are as follows:

$$\begin{aligned} U_A &= c_1 + \varepsilon c_3 - \mu e_A \\ U_B &= c_2 + \varepsilon c_1 - \mu e_B \\ U_C &= c_3 + \varepsilon c_2 - \mu e_C \end{aligned}$$

where $c_t$ denotes individual consumption of date-$t$ goods and $e_i$ the number of projects monitored by an individual type $i$ agent. Assume $0 < \varepsilon < 1$, so that the first term in an individual's utility function is the consumption of her preferred

[1] Note, however, the results of this paper do not hinge on this assumption and are robust to smaller populations.

goods. The configuration of production and consumption is summarized in the following table:

| Type of individuals | $A$ | $B$ | $C$ |
|---|---|---|---|
| Production goods | $t = 3$ | $t = 1$ | $t = 2$ |
| Preferred goods | $t = 1$ | $t = 2$ | $t = 3$ |

To motivate transactions across islands, I also assume $(1 - \lambda) y - \lambda\mu > \varepsilon y$. This condition implies that even after taking into account of the expected auditing cost, agents are strictly better off consuming their preferred goods than consuming their own production goods.

Given the above preferences and technology, there is lack of intertemporal double coincidence of wants in this economy. For example, type $A$ agents who produce at date 3 want to consume date-1 goods, which are produced by type $B$ agents who prefer date-2 goods. Therefore, type $A$ and $B$ agents cannot directly exchange outputs. Furthermore, with the limited communication friction, there is no way for all three types of agents to arrange multilateral exchanges. Thus, there is a potential role for money in this model. An individual may accept an object in exchange for her production goods, in the anticipation that this object can be used in the future to get the goods she prefers. Indeed, privately-issued debt instruments can be held in succession to lubricate the functioning of the economy, helping agents obtain the goods they like and hopefully improving welfare (as opposed to autarky). However, the value of such private money is undermined by the friction of limited information. To overcome this friction, intermediation and monitoring are also necessary. An intermediary can utilize the properties of the distribution of the aggregate output and monitors on behalf of other agents so as to economize the overall cost of monitoring.

# 3 Monitoring the Monitor

## 3.1 The Banking Mechanism

I study the banking arrangement where the bank lends out banknotes to type $A$ agents and expects them to repay their loan at date 3 after successful production. Banknotes each promise a certain amount of date 3 goods and are redeemable on demand. The bank is responsible for any monitoring activity as is contracted. I restrict my attention to contracts that are incentive compatible. In this economy, contracts are enforceable by law on local islands. To invoke the problem of monitoring the monitor, I assume that banking income is private information. It costs an individual $\theta > 0$ utils to monitor the bank and observe the true level of loan repayments it has collected. Any individual who monitors the bank has every right and ability to keep the result to herself unless she has an incentive not to. Indeed, we need to consider the optimal bank contract that solves a *two-sided incentive problem*, *i.e.* a contract that is incentive compatible for both the borrowers and the bank.

### 3.1.1 The Bank Loan Contract

In this section I consider the bank loan contract that requires a combination of banknotes and goods as repayments. This particular repayment scheme will be shown to be critical in solving the incentive problem concerning the bank. Other repayment schemes such as the goods-only repayment and the notes-only repayment, will be discussed in the next section.

The mechanism works as follows. At date 0, a bank arises through competition among all type $A$ agents.[2] The bank writes contracts with type $A$ agents and lends each of them a loan of $\frac{M}{N}$ dollars in banknotes. Conditioning on being a successful producer at date 3, each borrower must repay her loan with $pu$ units of banknotes and $\omega$ units of goods, where $u$ is the amount of goods to be sold in the spot market on island $A$ and $p$ is the spot price. This is called the *repayment scheme*. An unsuccessful producer pays nothing back to the bank. The contract also stipulates the *monitoring scheme* which consists of the set of monitoring states ($S$) and the probability of monitoring individual borrowers ($\alpha$). In particular, the set of auditing states indicates the conditions under which the bank monitors the projects claimed to be unsuccessful. Note that $S \subseteq \{n^a \in \mathbb{R}^+ : n^a \in [0, N]\}$ where $n^a$ is the number of claimed successful projects on island $A$. (when the contract is incentive compatible, $n^a$ also gives the true number of successful projects.) If the bank decides to audit, each of the announced failures is monitored with probability $\alpha$. If any borrower is found cheating, $z$ units of output will be forfeited from her, which is the *punishment scheme*.

Banknotes are redeemable on demand and each unit of them entitles the bearer to $\frac{N\varepsilon y}{M}$ units of goods,[3] which is the face-value of the note. The bank faces limited liability: If the loan repayment collected is less than the total amount of redemption demanded, the bank is only obliged to redeem notes up to the total amount of

[2] Assume that in order to be qualified as a banker, the individual is required to make her project under public surveillance, *i.e.*, to make the outcome of her project publicly observable. This way the tasks of being both a banker and a producer are isolated, and hence the banker is not able to take advantage of her position as an "insider".

[3] The expected payoff to note-redeemers in total must be equal to $N\varepsilon y$, which makes sure that type $C$ (and hence type $B$) agents are willing to accept banknotes in all states of the economy.

the loan repayment realized. In this case, the bank is said to be bankrupted or liquidated. Note-redeemers are served according to the sequential service rule by which payments were made to them on a first-come-first-served basis. They are also entitled to monitoring the bank when it defaults on redemption. Altogether, the bank's choices are $(u,\ \omega,\ z,\ \alpha,\ S)$. All terms of both the contract and the banknote are assumed to be public information.

[ Insert Figure 1. ]

Notice that according to the above repayment scheme, at date 3 all note-holders go to the spot market first, before they make the decision of whether or not to go to the bank to redeem notes. The reason is that they may benefit from the spot trade in case of a high level of aggregate output. In particular, with the probability of $P\left(n^a > \frac{N\varepsilon y}{u}\right)$, the spot price of goods is $p = \frac{M}{un^a} < \frac{M}{N\varepsilon y}$, which means note-holders can get a better deal in the market because the bank only pays one unit of goods for every $\frac{M}{N\varepsilon y}$ units of notes. With probability $P\left(n^a \leq \frac{N\varepsilon y}{u}\right)$, the equilibrium price is given by $p = \frac{m}{un^a} = \frac{M}{N\varepsilon y}$ where $m \leq M$ is the total amount of notes in the market when the price clears the market (see Fig.1) and $m$ is also assumed to be public information.[4] In this case, the price is equalized at the level of the deal offered by the banknotes.[5] Note-holders are indifferent between buying goods in the market and redeeming notes at the bank. Thus, some stay in the market and others are at

[4]This assumption is sensible in our model. The number of note-holders at the beginning of date 3 (who are actually all successful type $C$ agents), $n^c$, can be inferred from the spot price of goods on island $C$. Those of them at the bank to redeem notes can observe the number of individuals in their line-up, $n_b^c$. Then it is straightforward that $m = M\left(1 - n_b^c/n^c\right)$.

[5]Note that the equilibrium price cannot go beyond the implicit price of goods imposed by the bank. Otherwise, note-holders are better off getting goods from the bank and start leaving the market. As a result, the price of goods adjusts downwards until it reaches the level of $M/\left(N\varepsilon y\right)$.

the bank.

### 3.1.2 Market as the Source of Banking Information

According to the timing of our model, the bank is not able to use the banknotes collected as loan repayment to buy goods in the date 3 spot market. This is because after the bank collects all the repayments, the spot market has already shut down and all sellers had gone to the bank to repay loans. Therefore, notes as repayment bear no real value to the bank at date 3. To avoid the issue of banknote-overissue, it is also assumed that the bank is only able to issue notes at date 0 and the total amount $M$ is publicly observable. Thus, the bank cannot issue additional notes at $t > 0$ to dilute the purchasing power of those already in circulation.

The banking income is given by $\omega n^a$. Therefore, as long as $n^a$ is known by the public, the bank need not be monitored. According to the loan repayment which requires a mix of notes and goods, at date 3 all successful type $A$ agents first go to the market to sell $u$ units of goods in exchange for banknotes and then go to the bank to repay their loans. The spot price of goods is simply given by $p = \frac{m}{un^a}$. Since both $m$ and $u$ are public information, this price level reveals the true level of banking income $\omega n^a = \frac{\omega m}{pu}$.

Therefore, the spot market generates information-revealing prices that give the note-holders the accurate information about the bank's income. The bank is left on no grounds to cheat.[6] Thus, by requiring inside money to be part of the repayment, the bank intentionally "ties its own hands" and provides a market mechanism to

[6] One might think that it might be beneficial for the bank to bribe the borrowers into faking a bankruptcy before the spot market opens. However, that is not true because the bank never wants to reveal an intent to cheat before any loan is repaid. Otherwise, it becomes justifiable for borrowers to renegotiate for not repaying their loans.

prevent itself from deviating. In summary, inside money being required as a means of loan repayment is critical for solving the incentive problem of a bank. As a result, borrowers must obtain inside money in order to make repayments. A market of goods for inside money is created at the loan repayment stage to help reveal information and discipline the bank.

#### 3.1.3 The Banking Problem

The bank chooses a contract, $\mathbb{C}_1 = (u,\ \omega,\ z,\ \alpha,\ S)$, to maximize the expected utility of each individual borrower:

$$(PB1) \qquad \max_{(u,\ \omega,\ z,\ \alpha,\ S)} \qquad (1-\lambda)\,y + \varepsilon\,(1-\lambda)\,(y-u-\omega)$$

which is subject to the following constraints:

(i) The participation constraint of the bank:

$$\varepsilon\left\{(1-\lambda)\,N\omega - \int_0^{\frac{N\varepsilon y}{u}} \min\left[N\varepsilon y - un^a, \omega n^a\right] dF\,(n^a)\right\}$$
$$-\mu\left\{\int_0^N \left[\Lambda\,(n^a)\,(N-n^a)\right] dF\,(n^a) - \lambda\int_0^N \Lambda\,(n^a)\,dF\,(n^a)\right\} = 0 \qquad (1)$$

(ii) The participation constraint of potential note-redeemers:

$$u\int_{\frac{N\varepsilon y}{u}}^N n^a dF\,(n^a) + \int_0^{\frac{N\varepsilon y}{u}} \min\left[N\varepsilon y, (u+\omega)\,n^a\right] dF\,(n^a) = N\varepsilon y \qquad (2)$$

(iii) The incentive compatibility constraint of borrowers:

$$y-u-\omega \geq \int_0^N \left[\Lambda\,(n^a)\,(y-z) + (1-\Lambda\,(n^a))\,y\right] dF\,(n^a) \qquad (3)$$

(iv) Other normalizing constraints:

$$
\begin{aligned}
S &\subseteq \left\{ n^a \in \mathbb{R}^+ : n^a \in [0, N] \right\}; \\
\overline{S} &= \left\{ n^a \in \mathbb{R}^+ : n^a \in [0, N] \right\} \backslash S; \\
\Lambda\left(n^a\right) &= \begin{cases} \alpha, & \text{if } \; n^a \in S \\ 0, & \text{if } \; n^a \in \overline{S} \end{cases}; \\
u \;\geq\; 0; \;\; \omega \geq 0; \;\; & 0 \leq u + \omega \leq y; \;\; 0 \leq z \leq y; \;\; 0 \leq \alpha \leq 1.
\end{aligned}
$$

The bank's participation constraint simply says that by competitive banking the expected gain of banking is driven down to zero. Due to limited liability, the bank's total payout to the note-redeemers is given by $\min\left[N\varepsilon y - un^a, \omega n^a\right]$. The bank only monitors when the number of successful projects is within the set $S$. Note that the banker does not incur any cost of monitoring her own project and thus it should be excluded from the overall expected cost of monitoring.

The participation constraint of potential note-redeemers states that the expected payoff to type $C$ note-holders as a whole must be equal to $N\varepsilon y$, in order to guarantee the notes to end up circulating. It consists of the two potential sources of benefits to the note-holders, from the spot market and from note-redemption. All the other constraints intend to make the contract feasible.

**Lemma 1** *The optimal set of monitoring states is given by* $S^* = \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$.

**Proof (sketch).** First, if all borrowers report success, there is no need for monitoring. Hence, $S^* \subseteq \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$. It then follows that for any $S \subset \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$, there exist states of the economy in which the bank does not monitor any announced failures. No contract can be incentive compatible with such contingencies because the potential loan repayers can benefit from a fair lottery which decides who actually repay the loans. The lottery works as long as the number of individuals with successful projects is larger than the lowest number required by the bank for not monitoring any. The latter corresponds to the number of agents chosen by the lottery. This way, some can be spared from repaying the loan and the bank can do nothing about it. Therefore, the optimal set of monitoring states is $S^* = \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$, under which individuals do not find the above lottery worthwhile. ■

**Proposition 2** *Assume* $0 \leq \varepsilon(1-\lambda) - \lambda\frac{\mu}{y} - \varepsilon^2$*, there exists a unique solution,* $\mathbb{C}_1^*$*, to (PB1). The optimal bank contract consists of the following terms:* $u^* + \omega^* \equiv \rho^* = \frac{\varepsilon^2 y}{\varepsilon(1-\lambda)-\lambda\frac{\mu}{y}}$, $z^* = y$, $\alpha^* = \frac{\rho^*}{z^*} = \frac{\varepsilon^2}{\varepsilon(1-\lambda)-\lambda\frac{\mu}{y}}$, $S^* = \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$, *and a unique* $u^* \in (\varepsilon y, \rho^*)$ *(see Fig.2) that solves*

$$u = \frac{N\varepsilon y\left(1 - \int_{\frac{N\varepsilon y}{\rho^*}}^{\frac{N\varepsilon y}{u}} dF(n^a)\right) - \rho^* \int_0^{\frac{N\varepsilon y}{\rho^*}} n^a dF(n^a)}{\int_{\frac{N\varepsilon y}{u}}^{N} n^a dF(n^a)}. \tag{4}$$

**Proof.** See Appendix *A*. ■

[ Insert Figure 2. ]

**Understanding the optimal bank contract.** The terms of the optimal contract are quite intuitive. Firstly, the punishment scheme is pushed up to the harshest level possible, $z^* = y$. Since we are looking at incentive compatible contracts, the more severe the punishment, the less attempting for one to cheat. The optimal monitoring probability $\alpha^*$ turns out to be constant and does not depend on the aggregate output. The equivalent level of loan repayments in terms of goods $(u^* + \omega^* \equiv \rho^*)$ has the following properties:

$$\frac{\partial \rho^*}{\partial \lambda} > 0; \quad \frac{\partial \rho^*}{\partial \mu} > 0; \qquad \frac{\partial \rho^*}{\partial \varepsilon}, \quad \frac{\partial \rho^*}{\partial y} \begin{cases} \leq 0, & \text{if} \quad \varepsilon\,(1-\lambda) \leq \lambda \frac{\mu}{y} \\ > 0, & \text{if} \quad \varepsilon\,(1-\lambda) > \lambda \frac{\mu}{y} \end{cases} .$$

Both the probability of project failure $(\lambda)$ and the disutility of monitoring a project $(\mu)$ have a positive effect on $\rho^*$. Intuitively, all else equal, a higher chance of project failure or a higher utility loss associated with monitoring raises the expected monitoring cost of the bank; and hence, it must be compensated by higher level of loan repayments.

However, the effects of the utility factor of less preferred goods $(\varepsilon)$ and the output of a successful project $(y)$ on $\rho^*$, respectively, are ambiguous. In fact, they both have a positive and a negative effect. A higher level of $\varepsilon$ increases the utility gain induced by a given amount of repayment to the bank, which reduces the pressure on the repayment level. This constitutes the negative effect. In the meantime, however, a higher $\varepsilon$ also makes the goods more valuable to the borrowers. They would find it more rewarding to cheat. The bank would then have to exert more efforts on monitoring and eventually be compensated by a higher level of repayment. And that explains the positive effect. When $\varepsilon$ is small, the negative effect outweighs the positive one; while as $\varepsilon$ gets larger, it will eventually be dominated by the positive effect.

For a given level of repayment $\rho$, it becomes less heavy a burden on the borrowers as $y$ increases. Thus, borrowers are less reluctant to repay the loan and it is relatively easier for the bank to audit, which is the negative effect of $y$ on the repayment level. On the other hand, with the optimal monitoring probability being $\alpha^* = \frac{\rho}{y}$, for a given $\rho$, the probability of being audited becomes smaller as $y$ increases. Borrowers are faced with a rather mild auditing scheme and hence are more attempted to cheat. This describes the positive effect because the bank has to put more effort into monitoring as $y$ increases. Similarly to $\varepsilon$, for small $y$, the negative effect outweighs the positive one; while as it gets larger, it will eventually be dominated by the positive effect.

### 3.1.4 Contributions of Inside Money to Banking

Let me remark on the contributions of inside money to banking in the model. The bank issues inside money as a symbol of claims on the bank. The value of inside money to the bank comes in two ways. On one hand, by issuing inside money and making loans out of it, the bank is entitled to claims on borrowers, which are essentially the "goods" content of the loan repayments. In this sense inside money helps bring revenue to the bank. The bank deserves such claims because it provides liquidity and monitoring services to the economy. On the other hand, inside money adds value by skimming the cost of banking, as the problem of monitoring the bank can be solved by requiring inside money as part of the repayments. This second role of inside money seems novel relative to the literature.

# 4 Money and Goods as Alternative Repayments

The previous section established the theory that bank loan contracts with inside money involved in repayment solves the problem of monitoring the monitor. However, I have restricted the repayment scheme to a combination of goods and banknotes. It is not clear, *a priori*, whether this combination of repayments is better than repayments in only banknotes or in only goods. In this section, I examine contracts with those alternative repayment schemes and compare the associated allocations with those from the combination repayment.

## 4.1 Notes-Only Repayment

Rather than a combination of banknotes and goods, the bank might only ask for a certain amount of banknotes as loan repayment. However, as is explained in the previous section, the timing of this model is such that notes repayment bears no real value to the bank. And it has been assumed that the bank is not allowed to issue extra banknotes at date $t > 0$. It follows that loans must be repaid either only with goods or with a combination of both goods and notes, so as to compensate the bank for its monitoring efforts. The fact that notes-only repayment scheme does not work here is a particular result of the above assumption. Without it, the notes-only repayment works equivalently to the combined repayment.[7] Either way, the point is, markets of inside money help reveal banking information as long as it is involved in loan repayment.

[7] On the assumption that total money stock (both date-0 and date-3 issuances) is public information, the contract asks the borrowers to suppy $\rho^*$ units of goods to the market to obtain notes for loan repayment. At date 3, the bank issues $\frac{\omega^*}{u^*}M$ units of new notes, where $M$ is the amount issued at date 0 and values of $\rho^*$, $\omega^*$ and $u^*$ are the same as those given by $\mathbb{C}_1^*$. When the aggregate output is high enough so that spot price is no higher than $\frac{M}{N\varepsilon y}$ (the bank's offer), then all successful type $C$ agents and the banker buy goods in the market; Otherwise, the banker becomes the only buyer and all successful type $C$ agents go to the bank asking for redemption. Market prices still reveal banking income.

## 4.2 Goods-Only Repayment

Now we are ready to consider goods-only repayment. Here I shall specify in details the associated banking mechanism and demonstrate the fact that goods-only contracts are plagued with the problem of monitoring the monitor.

### 4.2.1 The Bank Loan Contract

The general picture is similar to the combination contract, with the main differences lying in the face-values of banknotes and repayment schemes. In particular, at date 0, a bank arises through competitive banking on island $A$. It contracts with each of the non-bankers and loans out $M$ dollars in banknotes to them. Loan repayment is conditional on the project success at date 3. The repayment scheme is such that each loan has to be repaid with $x$ units of goods. Monitoring is carried out on the contingencies specified by the set of monitoring states ($S$), with $\alpha$ being the probability of monitoring an individual who has announced failure. The punishment scheme is that upon monitoring $z$ units of output discovered will be forfeited. The face-value of each dollar of banknote is $\frac{H}{M}$ units of goods. Again, the bank is subject to limited liability and notes-redemption on demand. Now the bank's choices are $(x,\ z,\ \alpha,\ H,\ S)$.

Because loans are now repaid only with goods, the spot market does not function on island $A$. The borrowers go straight to the bank to repay their loan with output. Therefore, when the bank defaults on redemption, note-redeemers who are not served have the incentive to monitor the bank, in order to justify the claimed bankruptcy. The optimal set of monitoring states for note-holders is then given by $S_C^* = \left\{n^{a\prime} \in \mathbb{R}^+ : n^{a\prime} \in \left[0, \frac{H}{x}\right)\right\}$, where $n^{a\prime}$ is the number of successful projects reported by the bank. Without loss of generality, I assume that agents equally share the goods found hidden by the bank.

**Lemma 3** *With the goods-only contract, the equilibrium with a truth-revealing bank is unique and involves all the unserved note-redeemers monitor the bank with probability one.*

**Proof.** See Appendix $B$. ■

Therefore, with goods-only repayments, the problem of monitoring the monitor exists and there is duplication of efforts on monitoring the bank.

#### 4.2.2 The Banking Problem

The bank chooses, $\mathbb{C}_2 = (x,\ z,\ \alpha,\ H,\ S)$, to maximize the expected utility of each individual borrower:

$$(PB2) \qquad \max_{(x,\ z,\ \alpha,\ H,\ S)} \qquad (1-\lambda)\,y + \varepsilon\,(1-\lambda)\,(y-x)$$

which is subject to the following constraints:

(i) The participation constraint of the bank:

$$\begin{aligned}&\varepsilon\left\{(1-\lambda)\,Nx - \int_0^N \min\left[H, xn^a\right] dF\left(n^a\right)\right\}\\ &-\mu\left\{\int_0^N \left[\alpha\left(n^a\right)\left(N-n^a\right)\right] dF\left(n^a\right) - \lambda\int_0^N \alpha\left(n^a\right) dF\left(n^a\right)\right\} = 0 \qquad (5)\end{aligned}$$

(ii) The participation constraint of note-redeemers. In particular, ex ante total payoff for type $C$ agents to hold banknotes is no less than $n^c\varepsilon y$ for any number of successful projects on island $C$, $n^c$:

$$\int_0^N \min\left[H, xn^a\right] dF\left(n^a\right) - \theta n^c \int_0^{\frac{H}{x}} \left(1 - \frac{xn^a}{H}\right) dF\left(n^a\right) \geq n^c \varepsilon y, \quad \forall\, n^c \tag{6}$$

(iii) The incentive compatibility constraint of the borrowers:

$$y - x \geq \int_0^N \left[\alpha\left(n^a\right)(y - z) + \left(1 - \alpha\left(n^a\right)\right) y\right] dF\left(n^a\right) \tag{7}$$

(iv) Other normalizing constraints:

$$\begin{aligned}
S &\subseteq \left\{n^a \in \mathbb{R}^+ : n^a \in [0, N]\right\}; \\
\overline{S} &= \left\{n^a \in \mathbb{R}^+ : n^a \in [0, N]\right\} \backslash S; \\
\alpha\left(n^a\right) &= \begin{cases} \alpha, & \text{if} \quad n^a \in S \\ 0, & \text{if} \quad n^a \in \overline{S} \end{cases}; \\
0 \leq\ & x \leq y; \quad 0 \leq z \leq y; \quad 0 \leq H \leq Ny; \quad 0 \leq \alpha \leq 1.
\end{aligned} \tag{8}$$

The problem is similar to the one in the earlier section and is self-explanatory. What needs particular attention is that constraint (6) compensates note-holders for their expected cost of monitoring the bank. Since this condition must hold for any realized $n^c$, it implies

$$\int_0^N \min\left[H, xn^a\right] dF\left(n^a\right) - \theta N \int_0^{\frac{H}{x}} \left(1 - \frac{xn^a}{H}\right) dF\left(n^a\right) = N\varepsilon y. \tag{9}$$

**Lemma 4** *The optimal set of auditing states is given by* $S^* = \left\{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\right\}.$

**Proof.** The lottery argument in the proof of Lemma 1 also applies here. The only difference is that here people consider setting up the lottery before going to the bank rather than the spot market. ■

**Proposition 5** *Assume* $\varepsilon\lambda\frac{\theta}{y} \leq \varepsilon(1-\lambda) - \lambda\frac{\mu}{y} - \varepsilon^2$, *there exists a unique solution,* $\mathbb{C}_2^*$, *to (PB2). The optimal bank contract consists of the following terms:* $S^{**} = \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$, $z^{**} = y$, $\alpha^{**} = \frac{x^{**}}{y}$, *and unique* $x^{**} \in (\varepsilon y, y]$ *and* $H^{**} \in (N\varepsilon y, Ny)$ *(see Fig.3 and Fig.4) that solve*

$$\begin{cases} H - \left(1 + \frac{\theta N}{H}\right) \int_0^{\frac{H}{x}} (H - xn^a)\, dF(n^a) - N\varepsilon y = 0 & (a) \\ \left(\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\frac{N-1}{N}\right) x - \frac{\varepsilon\theta}{H} \int_0^{\frac{H}{x}} (H - xn^a)\, dF(n^a) = \varepsilon^2 y. & (b) \end{cases} \quad (10)$$

**Proof.** See Appendix $C$. ■

[ Insert Figure 3 and Figure 4. ]

### 4.2.3 Summary of Results

Table 1 compares the results from the goods-only contract and the combination contract. With the goods-only contract, cost of monitoring includes both the cost of monitoring borrowers and the cost of monitoring the bank (whenever necessary), while the combination contract does not involve the latter. As a result, the goods-only contract must impose on the borrowers a strictly higher level of repayment in order to cover its higher expected costs. Not surprisingly, type $A$ agents are strictly better off with the combination contract and the bank will choose to implement it rather than the goods-only contract. This explains why the bank would intentionally provide a market mechanism to discipline itself: It gains edges in competitive banking by offering loan contracts with the least stringent requirement on repayment, which of course is attractive to all borrowers. Both the combination and the goods-only contracts offer the same level of expected utility to type $B$ and $C$ agents, respectively. In terms of overall welfare, the combination contract strictly dominates the goods-only contract.

Furthermore, the goods-only contract does not induce a spot market on island $A$ and involves a higher probability of bankruptcy. Note that the higher the probability of bankruptcy, the more likely note-holders get nothing back from the bank. Thus, if agents were risk averse, *ceteris paribus*, type $C$ agents would also prefer the combination contract.

[ Insert Table 1. ]

Lastly, notice particularly that

$$x^{**} \to \rho^*, \quad \text{as } N \to \infty.$$

The advantage of the combination repayment scheme fades away as the aggregate uncertainty diminishes. In the limit economy, it does not matter whether bank loans are repaid with goods, or notes, or a mix of both. Intuitively, one would not have to worry about monitoring the monitor in an infinite economy after all, according to Diamond (1984) and Williamson (1986).[8]

[8] Andolfatto and Nosal (2003) discuss the efficiencies of non-banking private money arrangements (*NBPMA*) and banking mechanisms, respectively. Similar results apply here in the presence of aggregate uncertainty. The banking mechanism improves efficiency, as opposed to the *NBPMA*, because of its remarkable advantage in reducing cost of monitoring. Also, it is in the interest of all type $A$ agents to compete and become the banker because a banking mechanism offers a strictly higher level of expected utility to them.

# 5 Discussion

In this section I discuss the robustness of the main results with respect to the existence of multiple banks and the introduction of outside money.

## 5.1 Multiple Banks

A key result of the model is that, by requiring inside money as part of the repayments, a market of goods for inside money generates prices that fully reveal the bank's private information. One wonders whether this result is robust to the existence of multiple banks. With multiple banks, if market prices reveal only the total number of loan repayers in the economy, then it is not sufficient for agents to calculate the solvencies of individual banks.

This potential problem of monitoring multiple banks arises only when all banks issue the same banknotes. However, such indistinguishable banknotes are hardly realistic or optimal for the banks. Recall that an important purpose for a bank to issue inside money and to require it as part of the repayments is to reduce its own incentive to cheat at the repayment stage. To achieve this goal, each bank will find it optimal to make its banknotes distinguishable from other banks' notes. By requiring its own banknotes to be part of the repayments, a bank induces a distinctive market of goods for its banknotes at the repayment stage. In this case, there will be as many spot markets as the number of banks. The price in each market fully reveals the information of the corresponding bank. Again the problem of monitoring multiple banks is solved.

It is also interesting to consider arbitrage across markets as there have been concerns that such behavior might mess up the market prices and jeopardize our result. Let us consider the scenario with two banks, bank $E$ and bank $F$. Let $p_k$ and $p'_k$ (where $k = E, F$) denote the spot price of goods and the redemption value, respectively, corresponding to each of the banks. Suppose a buyer in market $E$ were to arbitrage and let $\psi$ be the amount of bank-$E$ notes she initially has. The individual arbitrages by first obtaining goods in market $E$. She then sells the goods in market $F$ and redeems the notes at bank $F$. Therefore, the gain of arbitrage is given by

$$\frac{\psi}{p_E}\frac{p_F}{p'_F} \leq \frac{\psi}{p_E}, \quad \text{as} \;\; p_F \leq p'_F.$$

The right-hand side of the above inequality is the amount of goods she can get if she does not arbitrage. Obviously, it never pays off for one to attempt an arbitrage no matter what the relative price across markets $(\frac{p_E}{p_F})$ is .

The intuition is that the result of an arbitrage in such a model is to obtain goods from another bank (as in the previous example, the individual with bank-$E$ notes ends up getting goods from bank $F$). Since goods are never cheaper at a bank than at its corresponding market ($p_F \leq p'_F$ according to Fig.1), the arbitrageur can obtain no more than whatever amount of goods she brings into the target market $(\frac{\psi}{p_E})$, which is exactly the amount she can get without any arbitrage. By the same logic, no seller has the incentive to sell her goods in a market of no counterpart to the bank she borrows from. All the above arguments can be applied to situations of more than two banks. In any case, no one has the incentive to arbitrage or sell goods in markets other than the market she is initially assigned to. Market prices still have no trouble in revealing the information of each existing bank.

## 5.2 Inside vs. Outside Money

The above discussions also help us distinguish the role of inside money from that of outside money. Suppose that inside money is banned and outside money is imposed upon the model economy. In the case of a unique bank, requiring outside money as part of the repayments also eliminates the cost of monitoring the bank. However, this is not true when there exist multiple banks. Recall that with multiple banks inside money solves the problem of monitoring the bank precisely because each bank can make its banknotes distinguishable from other banks' notes. This cannot be done with outside money. That is, outside money cannot induce prices in the goods market that differentiate the solvencies of individual banks.

# 6 Conclusion

This paper studied the problem of monitoring the monitor in a model of money and banking with aggregate uncertainty. The bank is characterized by a finitely-sized portfolio and is subject to default risk. There are strategic interactions between the bank and the agents and among the non-banker agents themselves. In effect, the bank conducts monitoring and the model is characterized by chances of bankruptcy in equilibrium.

The mechanism of inside money being required as a means of bank loan repayment is shown to solve the information problem of a bank. As a result, a market of goods for inside money is created at the loan repayment stage and generates information-revealing prices that help discipline the bank. Furthermore, inside money distinguishes itself from outside money by its ability to provide incentives for truthful revelation even with the existence of multiple banks.

Inside money contributes to banking in two ways: on one hand, inside money brings in banking revenue by entitling claims on borrowers to the bank, for the liquidity and monitoring services it provides to the economy. On the other hand, by solving the problem of monitoring the bank, inside money helps reduce the overall cost of banking and improve the efficiency of delegated monitoring.

# Appendix

## A Proof of Proposition 2

**Proof.** According to Lemma 1,

$$\int_{\overline{S}^*} dF(n^a) \simeq 0; \quad \text{and} \quad \int_{S^*} dF(n^a) \simeq 1.$$

$$\int_{S^*} (N - n^a)\, dF(n^a) = \int_0^{N-1} (N - n^a)\, dF(n^a) \simeq \int_0^{N} (N - n^a)\, dF(n^a) = \lambda N.$$

The approximations hold because $N$ is assumed to be sufficiently large. Constraint (1) boils down to

$$\begin{aligned}
& \varepsilon \left\{ (1-\lambda) N\omega - \int_0^{\frac{N\varepsilon y}{u}} \min\left[N\varepsilon y - un^a, \omega n^a\right] dF(n^a) \right\} \\
& -\mu\alpha \int_{S^*} (N - n^a - \lambda)\, dF(n^a) = 0 \\
\Rightarrow \quad & \varepsilon \left\{ (1-\lambda) N\omega - \int_0^{\frac{N\varepsilon y}{u}} \min\left[N\varepsilon y - un^a, \omega n^a\right] dF(n^a) \right\} - \mu\alpha\lambda (N-1) = 0.
\end{aligned} \tag{11}$$

Notice that $n^a \notin S^*$ only happens when $n^a = n^{a\prime} = N$, where every type A agent is successful and tells the truth. Hence, the incentive compatibility constraint simplifies to $u + \omega \leq \alpha z$. Notice that all else equal, a higher value of $\alpha$ means a strictly lower utility payoff (the left-hand side of equation [11]) to the bank. Therefore, for any given level of $(u + \omega)$, the bank will choose $\alpha$ such that $u + \omega = \alpha z$. Also for a given $(u + \omega)$, the higher the value of $z$ the lower $\alpha$ is needed. Hence the bank will choose the highest possible value of $z$, $z^* = y$ (which satisfies the constraint $0 \leq z \leq y$), such that $\alpha$ is minimized and the profit is maximized with lowest possible expected monitoring costs. It all boils down to $\alpha^* = \frac{u+\omega}{z} = \frac{\rho}{y}$.

Multiplying equation (2) with $\varepsilon$ and adding it to equation (1), we have:

$$\varepsilon (1-\lambda) N\rho - N\varepsilon^2 y - \frac{\rho}{y}\mu\lambda (N-1) = 0. \tag{12}$$

Re-arranging this equation we get $\rho^* \simeq \frac{\varepsilon^2 y}{\varepsilon(1-\lambda)-\lambda\frac{\mu}{y}}$. Again conditions $0 \leq \rho \leq y$ and $0 \leq \alpha \leq 1$ are equivalent. According to these conditions, the existence of solution requires $0 \leq \varepsilon (1-\lambda) - \lambda\frac{\mu}{y} - \varepsilon^2$.

From equation (2) we can solve for $u$ as is shown by equation (4). Define $\xi (u)$ as the right-hand side of equation (4). Then,

$$\begin{aligned} \xi (u) &= \frac{\int_0^{\frac{N\varepsilon y}{\rho^*}} (N\varepsilon y - \rho^* n^a) dF (n^a)}{0} = \infty > u, \qquad \forall \; u \in [0, \varepsilon y] \\ \xi (\rho^*) &= \frac{N\varepsilon y - \rho^* \int_0^{\frac{N\varepsilon y}{\rho^*}} n^a dF (n^a)}{\int_{\frac{N\varepsilon y}{\rho^*}}^{N} n^a dF (n^a)} < \rho^*. \end{aligned}$$

Also, $\xi' (u) < 0, \; \forall \; u \in [0, \varepsilon y]$ and $\xi' (\rho^*) > 0$. The sign of $\xi'' (u)$ is indecisive. By continuity, for $u \in (\varepsilon y, \rho^*)$, there exists at least one solution to the equation $u = \xi (u)$.

To see the uniqueness of $u^*$ and hence $\omega^*$, notice that we can solve for them from equations (2) and (12). Re-arrange them and define

$$\begin{cases} K (u, \omega) \equiv \; (1-\lambda) Nu + \int_{\frac{N\varepsilon y}{u+\omega}}^{\frac{N\varepsilon y}{u}} (N\varepsilon y - un^a) dF (n^a) + \omega \int_0^{\frac{N\varepsilon y}{u+\omega}} n^a dF (n^a) - N\varepsilon y = 0 \\ L (u, \omega) \equiv \left(\varepsilon (1-\lambda) - \lambda\frac{\mu}{y}\frac{N-1}{N}\right) (u+\omega) - \varepsilon^2 y = 0 \end{cases}.$$

The Jacobian matrix of this system of equations is:

$$\begin{aligned}
J &= \begin{vmatrix} (1-\lambda)N - \int_{\frac{N\varepsilon y}{u+\omega}}^{\frac{N\varepsilon y}{u}} n^a dF(n^a) & \int_0^{\frac{N\varepsilon y}{u+\omega}} n^a dF(n^a) \\ \varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\frac{N-1}{N} & \varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\frac{N-1}{N} \end{vmatrix} \\
&= \left(\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\frac{N-1}{N}\right)\left((1-\lambda)N - \int_0^{\frac{N\varepsilon y}{u}} n^a dF(n^a)\right) \\
&\neq 0, \quad \forall \;\; u > \varepsilon y.
\end{aligned}$$

We have shown that any solution of $u$ to the system satisfies $u \in (\varepsilon y, \rho^*)$. Therefore, there exist unique solutions $(u^*, \omega^*)$ to the system of $K(u,\omega) = 0$ and $L(u,\omega) = 0$. This also implies a unique solution to $u^* = \xi(u^*)$ which can be a result of the possible situations as is shown in Fig.2. ■

# B Proof of Lemma 3

**Proof.** To report to the note-holders, the bank chooses the probability of cheating and *the extent to cheat*, denoted as $\delta$ and $\Gamma(\delta)$, respectively. In particular, the bank cheats by claiming not having enough income to redeem all banknotes while it actually is able to. The extent to cheat is given by the ratio of all notes that the bank voluntarily redeems. It denotes how much of the true information the bank reveals. In particular,

$$\Gamma(\delta) = \begin{cases} \gamma, & \text{if} \quad 0 < \delta \leq 1 \\ 1, & \text{if} \quad \delta = 0 \end{cases};$$
$$0 \leq \delta \leq 1; \quad 0 \leq \gamma \leq 1.$$

The unserved note-redeemers choose the probability $\Phi\left(n^{a\prime}\right)$ with which to monitor the bank:

$$\Phi\left(n^{a\prime}\right)=\begin{cases}\phi, & \text{if} \quad n^{a\prime}\in S_C^* \\ 0, & \text{if} \quad n^{a\prime}\notin S_C^*\end{cases}.$$

Note that the bank may try to "bribe" the auditors into not revealing what they have found to the non-auditors, *i.e.* those who are not served but do not go auditing the bank. In particular, the bank can offer to pay the auditors the exact amount they are entitled to if they truthfully reveal the banking information, and ask them not to tell the truth. Since the auditors are indifferent between telling and not telling, without loss, I assume they do accept such an offer.

The timing of this game is given by the following:

1. The bank announces bankruptcy and the number of successful projects, $n^{a\prime}$;

2. Auditors come to monitor the bank;

3. The bank offers bribery to the auditors;

4. The auditors decide whether to accept the bribery and announce the result of the audit accordingly.

Let us first consider an unserved note-redeemer's best responses. Define $n^c$ as the number of successful type $C$ agents (or, the note-holders). Given $n^c$ and $n^{a\prime}$, she chooses $\phi$ to maximize her expected payoff:

$$\max_{\phi} \quad \phi\left\{\frac{\Delta}{\int_{n^{a\prime}}^{N} dF\left(n\right)}\int_{n^{a\prime}}^{N}\left(\frac{\min[H,xn]-xn^{a\prime}}{\left(1-\frac{xn^{a\prime}}{H}\right)n^c}\right)dF\left(n\right)-\theta\right\}$$

where $\Delta$ denotes the potential auditors' belief of the chances that the bank has cheated. I assume that agents take $\Delta$ as given and it is observable only after the bank has announced its financial status. The best responses boil down to

$$\phi^* \begin{cases} = 1, & \text{if} \quad \Delta > \theta/W \\ \in [0,1], & \text{if} \quad \Delta = \theta/W \\ = 0, & \text{if} \quad \Delta < \theta/W \end{cases}$$

where

$$W \equiv \frac{H}{n^c \int_{n^{a\prime}}^{N} dF(n)} \int_{n^{a\prime}}^{N} \left( \min[1, \ \frac{x(n - n^{a\prime})}{H - xn^{a\prime}}] \right) dF(n).$$

Let $\widetilde{\phi}$ be the bank's expectation of potential auditors' choice of $\phi$, the probability of an individual goes auditing the bank. The bank is assumed to take $\widetilde{\phi}$ as given when making decisions. It chooses $\delta$ and $\gamma$ to minimize its expected payout to the note-holders:

$$\min_{(\delta,\ \gamma)} \quad \left\{ \delta \left[ \Gamma(\delta) + \widetilde{\phi}(1 - \Gamma(\delta)) \right] \min[H, xn^a] + (1-\delta) \min[H, xn^a] \right\}.$$

The first term denotes the expected payout of the bank if it cheats. In such a case, it reveals and pays a fraction $\Gamma(\delta)$ of the amount it owes to all note-holders, *i.e.* $\min[H, xn^a]$. Those who are not paid each come to audit the bank with probability $\widetilde{\phi}$. The bank successfully bribes the auditors and gets to keep the rest of the output to itself. The non-auditors get nothing. With probability $(1-\delta)$, the bank reveals the truth, which is given by the second term in the above. Given $\widetilde{\phi}$, the optimal responses are given as follows:

$$\gamma^* \begin{cases} = 0, & \text{if} \quad 0 \leq \widetilde{\phi} < 1 \\ \in [0,1], & \text{if} \quad \widetilde{\phi} = 1 \end{cases} ; \tag{13}$$

$$\delta^* \begin{cases} = 1, & \text{if } \ 0 \leq \widetilde{\phi} < 1 \\ \in [0,1], & \text{if } \ \widetilde{\phi} = 1 \end{cases} . \tag{14}$$

Intuitively, as long as $0 \leq \widetilde{\phi} < 1$, not all the interested note-holders are expected to come. Then it is always beneficial for the bank to cheat and try to bribe the auditors; accordingly, $\delta = 1$ and $\gamma = 0$.

A *rational expectation equilibrium* is defined as the following:

1. Given $\widetilde{\phi}$, the bank chooses $(\delta, \gamma)$ to minimize its expected payout to the note-holders;

2. Given $(n^c, n^{a\prime}, \Delta)$, individual note-holders who are interested in auditing the bank choose $\phi$ to maximize their expected payoff;

3. $\phi = \widetilde{\phi}$.

In sum, there exist a continuum of equilibria of this game. However, the only one with a truth-telling bank consists of the optimal strategies of $\phi^* = 1$, $\delta^* = 0$ and $\Gamma(\delta)^* = 1$. In particular, all note-redeemers who are not served monitor the bank with probability one. The bank rationally expects this ($\widetilde{\phi} = \phi$) and fully reveals the truth ($n^{a\prime} = n^a$). The existence of this particular equilibrium requires $\Delta > \theta/W$. Indeed, it depends crucially on the potential auditors' common judgment in regard to the bank's honesty. A high level of $\Delta$ represents low public faith in the bank and triggers an audit involving all the interested individuals. ■

# C Proof of Proposition 5

**Proof.** By Lemma 4, $S^* = \{n^a \in \mathbb{R}^+ : n^a \in [0, N-1]\}$. And again, $z^{**} = y$ and $\alpha^{**} = \frac{x}{y}$. Re-arrange equation (9) we get equation (10.a), which implies that $H > N\varepsilon y$. Substitute equation (9) into the participation constraint of the bank, equation (5), and it simplifies to equation (10.b) and hence we have:

$$\begin{aligned} x &= \frac{\varepsilon^2 y + \frac{\varepsilon\theta}{H}\int_0^{\frac{H}{x}} (H - xn^a)\, dF(n^a)}{\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\frac{N-1}{N}} \\ &\simeq \frac{\varepsilon^2 y + \frac{\varepsilon\theta}{H}\int_0^{\frac{H}{x}} (H - xn^a)\, dF(n^a)}{\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}} \equiv \xi(x). \end{aligned} \tag{15}$$

Then it follows that

$$\begin{aligned} &\xi(x) > \varepsilon y \geq x, \qquad \forall\, x \in [0, \varepsilon y] \\ &\xi'(x) = \frac{-\varepsilon\theta \int_0^{\frac{H}{x}} n^a dF(n^a)}{H\left(\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\right)} < 0 \\ &\xi''(x) = \frac{\varepsilon\theta}{H\left(\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}\right)}\frac{H^2}{x^3} > 0. \end{aligned}$$

There exists a unique solution $x^{**} \in (\varepsilon y, y]$ to $x = \xi(x)$ if and only if $\xi(y) \leq y$ (see Fig.3), *i.e.*

$$\begin{aligned} &\frac{\varepsilon^2 y + \frac{\varepsilon\theta}{H}\int_0^{\frac{H}{y}} (H - yn^a)\, dF(n^a)}{\varepsilon(1-\lambda) - \lambda\frac{\mu}{y}} \leq y, \qquad \forall\, H \\ \Rightarrow\ &\varepsilon^2 y + \frac{\varepsilon\theta}{H}\int_0^{\frac{H}{y}} (H - yn^a)\, dF(n^a) \leq \varepsilon(1-\lambda)\, y - \lambda\mu, \qquad \forall\, H. \end{aligned}$$

Notice that the term $\frac{\varepsilon\theta}{H}\int_0^{\frac{H}{y}}(H - yn^a)\,dF(n^a)$ is strictly increasing in $H$. Under the condition $H \le Ny$, a sufficient condition for both existence and uniqueness is then given by:

$$\varepsilon\lambda\frac{\theta}{y} \le \varepsilon(1-\lambda) - \lambda\frac{\mu}{y} - \varepsilon^2.$$

Re-arrange equation (15) and substitute it into equation (10.a), we get

$$H = \left(1 - \lambda - \frac{\lambda\mu}{\varepsilon y}\right)Nx + \int_0^{\frac{H}{x}}(H - xn^a)\,dF(n^a) \equiv G(H).$$

We have shown that $H > N\varepsilon y$, hence,

$$\begin{aligned} &G(N\varepsilon y) > N\varepsilon y \\ &G(Ny) = Ny - \frac{\lambda\mu}{\varepsilon y}Nx < Ny \\ &G'(H) = \int_0^{\frac{H}{x}} dF(n^a) > 0 \\ &G''(H) = \frac{1}{x} > 0. \end{aligned}$$

Therefore, there exists a unique solution $H^{**} \in (N\varepsilon y, Ny)$ to $H = G(H)$ (see Fig.4).

■

# References


[1] Andolfatto, D. and E. Nosal, 2003, "A theory of money and banking," Federal Reserve Bank of Cleveland, working paper 03-10.

[2] Angeletos, G. and I. Werning, 2005, "Crises and prices: information aggregation, multiplicity and volatility," working paper.

[3] Bernanke, B. and M. Gertler, 1989, "Agency costs, net worth, and business fluctuations," *American Economic Review* **79**, 14-31.

[4] Cavalcanti, R., A. Erosa, and T. Temzelides, 1999, "Private money and reserve management in a random-matching model," *Journal of Political Economy* **107(5)**, 929-945.

[5] Cavalcanti, R. and N. Wallace, 1999a, "Model of private bank-note issue," *Review of Economic Dynamics* **2(1)**, 104-136.

[6] Cavalcanti, R. and N. Wallace, 1999b, "Inside and outside money as alternative media of exchange," *Journal of Money, Credit, and Banking* (August), 443-457.

[7] Diamond, D., 1984, "Financial intermediation and delegated monitoring," *Review of Economic Studies* **51**, 393-414.

[8] Diamond, D. and P. Dybvig, 1983, "Bank runs, deposit insurance and liquidity," *Journal of Political Economy* **91**, 410-419.

[9] Dow, J., 1998, "Arbitrage, hedging and financial innovation," *Review of Financial Studies* **11**, 739-755.

[10] Faig, M., 2004, "Money and banking in an economy with villages," manuscript.

[11] Freeman, S., 1996, "Clearinghouse banks and banknote over-issue," *Journal of Monetary Economics* **38**, 101-115.

[12] Freeman, S., 1996, "The payments system, liquidity, and rediscounting," *American Economic Review* **86**, 1126-1138.

[13] Friedman, M., 1960, "A program for monetary stability," Fordham University Press, New York.

[14] Gorton, G. and J. Haubrich, 1987, "Bank deregulation, credit markets, and the control of capital," *Carnegie-Rochester Conference Series on Public Policy* **26**, 289-334.

[15] Gorton, G. and A. Winton, 2003, "Financial intermediation," *Handbook of the Economics of Finance*, edited by G. Constantinides, M. Harris and R. Stulz (Amsterdam: Elsevier North-Holland; ISBN: 044450298X), Volume 1A, Chapter 8.

[16] Hart, O. D., 1975, "On the optimality of equilibrium when markets are incomplete," *Journal of Economic Theory* **11**, 418-443.

[17] He, P., L. Huang, and R. Wright, 2003, "Money and banking in search equilibrium," manuscript.

[18] Jacklin, Charles J. 1987, "Demand deposits, trading restrictions, and risk sharing," *Contractual Arrangements for Intertemporal Trade*, edited by Edward C. Prescott and Neil Wallace (Minnesota Studies in Macroeconomics, vol. 1. Minneapolis, Minn.: University of Minnesota Press), 26–47.

[19] Kiyotaki, N. and J. H. Moore, 2000, "Inside money and liquidity," manuscript, London School of Economics.

[20] Kiyotaki, N. and J. H. Moore, 2001, "Evil is the root of all money," Claredon Lecture, London School of Economics.

[21] Krasa, S. and A. Villamil, 1992a, "Monitoring the monitor," *Journal of Economic Theory* **30**, 197-221.

[22] Krasa, S. and A. Villamil, 1992b, "A theory of optimal bank size," *Oxford Economic Papers* **44**, 725-749.

[23] Marin, J. and R. Rahi, 2000, "Information relevation and market incompleteness," *Review of Economic Studies* **67**, 455-481.

[24] Shi, S., 1996, " Credit and money in a search model with divisible commodities," *Review of Economic Studies* **63**, 627-652.

[25] Townsend, R. M., 1979, "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory* **21**, 1-29.

[26] Townsend, R. M., 1987, "Economic organization with limited communication," *American Economic Review* **77**, 954-971.

[27] Williamson, S., 1986, "Costly monitoring, financial intermediation, and equilibrium credit rationing," *Journal of Monetary Economics* **18**, 159-179.

[28] Williamson, S., 1987, "Financial intermediation, business failures, and real business cycles," *Journal of Political Economy* **95**, 1196-1216.

[29] Winton, A., 1995, "Delegated monitoring and bank structure in a finite economy," *Journal of Financial Intermediation* **4:2**, 158-187.

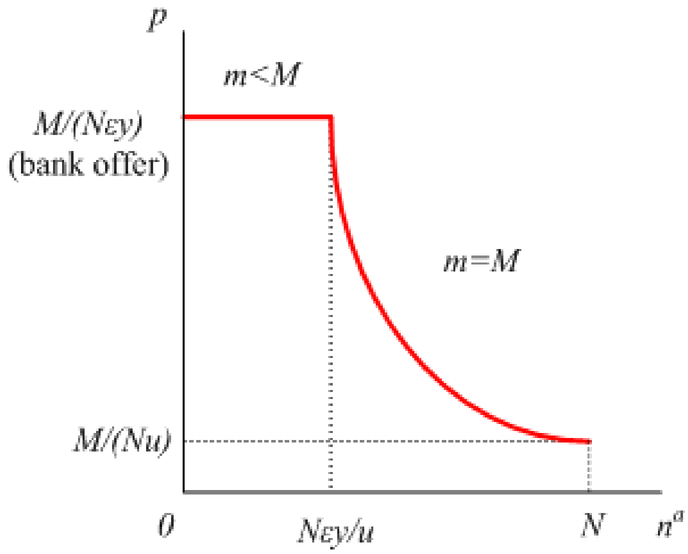


Figure 1 Spot price of island $A$

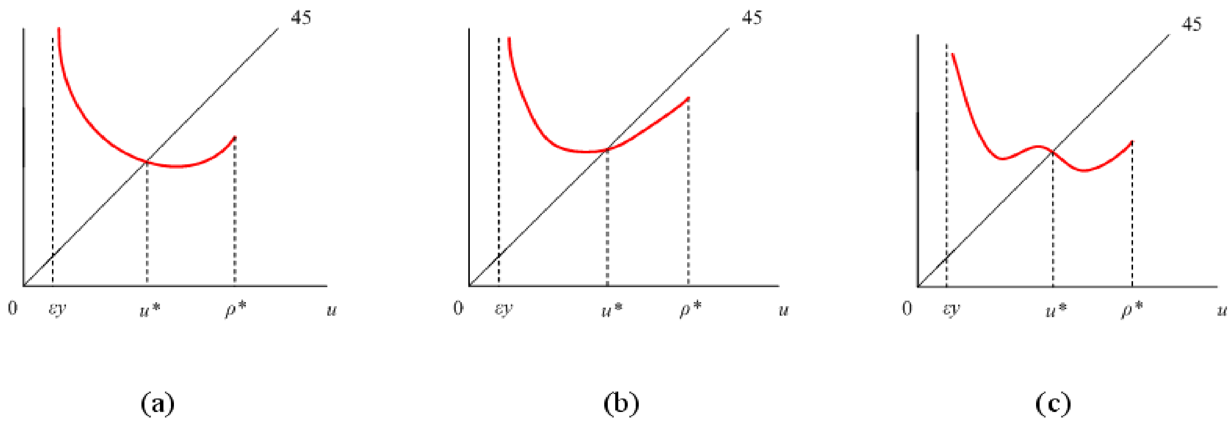


Figure 2 Illustrations of possible solutions of $u^*$

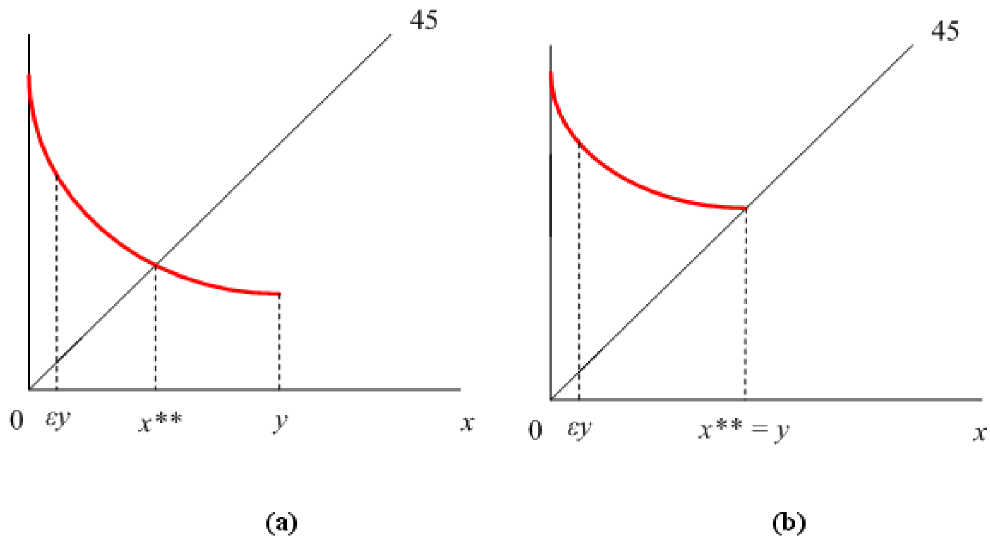


Figure 3  Illustration of possible solution of $x^{**}$

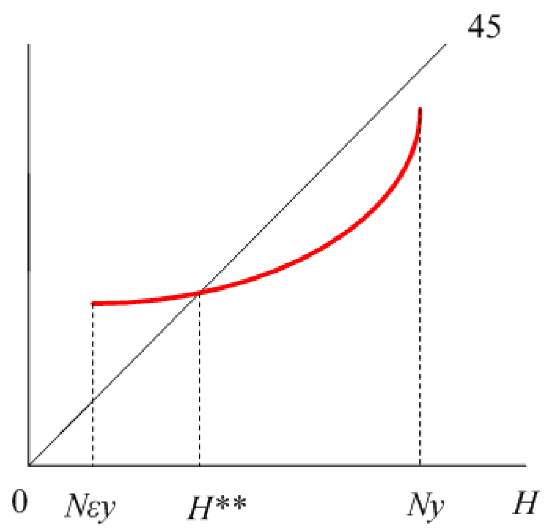


Figure 4  Unique solution of $H^{**}$

**Table 1. Comparison of Bank Loan Contracts**

| | $\mathbb{C}_1^*$ **(Combination)** | | $\mathbb{C}_2^*$ **(Goods-only)** |
|---|---|---|---|
| **Real value of repayments** | $\rho^*$ | $<$ | $x^{**}$ |
| **Expected monitoring costs** | $\alpha^* \mu\lambda (N-1)$ | $<$ | $\alpha^{**} \mu\lambda (N-1)$ |
| **Probability of bankruptcy** | $P\left(n^a \leq \frac{N\varepsilon y}{\rho^*}\right)$ | $<$ | $P\left(n^a \leq \frac{H^{**}}{x^{**}}\right)$ |
| **Expected utility to** $A$ | $(1-\lambda)\left[y+\varepsilon\left(y-\rho^*\right)\right]$ | $>$ | $(1-\lambda)\left[y+\varepsilon\left(y-x^{**}\right)\right]$ |
| **Expected utility to** $B$ | $(1-\lambda) Ny$ | $=$ | $(1-\lambda) Ny$ |
| **Expected utility to** $C$ | $N\varepsilon y$ | $=$ | $N\varepsilon y$ |
| **Local spot markets** | $A,\ B,\ C$ | | $B,\ C$ |