(Preliminary and Incomplete)
This Version: April 6, 2012

Tail-Risk Dumping: Bank Bailouts as Optimal Fiscal Operation

Kenichi Ueda *

ABSTRACT

Expectation for bank bailouts is often said to bring inefficiency, especially too much risk taking. However, in some cases, insolvency may be a result of forces exogenous to banks. In these cases, this paper shows that bank bailouts can be optimal from macroeconomic point of view but that it is so only if bankers and borrowers are too much protected by limited liability. I develop a simple general equilibrium model in which productive activities and financial intermediation are determined simultaneously by occupational arbitrage between entrepreneurs and bankers. Entrepreneurs are then sorted out to be borrowers or depositors depending on their draws on idea shocks. The model assumes realistic financial frictions: costly state verification and limited liability. The optimal loan and deposit contracts take a form of a standard debt contract due to costly state verification associated with bank lending. The optimal bank capital is positive and the banking sector is sizable. When a large negative shock hits, borrowers and banks would walk away with some retained assets that the lax limited liability allows. Depositors have to assume all the tail risk (*tail-risk dumping*). To mitigate this tail-risk dumping problem, government-led bailouts of banks, if transparent, can improve welfare as all the funds are distributed to depositors. Then, ex ante occupational choice becomes less risky and everyone better off. A deposit insurance scheme, if funding are supported by a tax contingent on outputs ex post, can mimic the transparent bank bailouts. Some form of liquidity ratio requirement works well, too. In any case, the government needs to raise its revenue (e.g., sales tax, inflation tax) to transfer funds. This

*Kenichi Ueda is a senior economist at the International Monetary Fund. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management. I benefited a lot from discussions from Grace Li. I would like to thank for helpful comments from Nobu Kiyotaki, Tom Sargent, Rob Townsend, and Ivan Werning as well as participants of seminars at Georgetown, IMF, and MIT and the Annual Research Conference at De Nederlandsche Bank and the Annual Research Conference at Georgetown Center for Economic Research.

contingent fiscal operation is optimal. (To Be Addressed: Whether the transfers should be funded by bonds before raising tax revenue.)

JEL Classification Numbers: E44, G21, G28
Keywords: Bailout, prudential regulation, deposit insurance, insolvency, tail risk, endogenous banking sector

I. INTRODUCTION

Bailing out banks, among others, was the key policy adopted in the recent financial crisis by many countries. But, when do banks need to be bailed out? Are there any other better options? Can tighter regulations prevent future bank bailouts? Because systemic importance was stressed as a major reason for bank bailouts, financial sector policies need to be evaluated from a macroeconomic perspective. Because regulations and bailout expectations affect the banking sector size, policy implications is better to be studied recognizing endogeneity of the banking sector size.

Against these backgrounds, I set up a simple general-equilibrium model in which productive activities and financial intermediation are determined simultaneously by arbitrage in the factor markets. The model determines the size of banking sector and bank capital. Then, from a macroeconomic perspective, I evaluate banking sector policies such as bank bailouts and capital adequacy ratio requirement.

Expectation for bank bailouts is often said to bring inefficiency, especially too much risk taking. However, in some cases, insolvency may be a result of forces exogenous to banks: for example, devastative natural disaster (e.g., banks in North coast Japan after large earthquake and tsunami) and large swings in international interest rates for small open developing countries (e.g., Latin American crisis and Thai crisis in early 1980s). In these exogenously induced banking crisis, this paper shows that bank bailouts can be optimal but that it is so only if bankers and borrowers are too much protected by limited liability.

In the model, anyone can choose to work either in the production sector or the financial sector. The financial sector intermediates capital with a fee, which determines the spread between the deposit and the loan rates. This fee is the income of bankers who inspect and verify the situation that a defaulted borrowers face. The expected utility of bankers should equate with the expected utility of entrepreneurs in the production sector.

Because the financial fee is a waste for the production sector, a zero spread, if possible, between the deposit and the loan rates would be the first best. In the hypothetical first best allocation, almost all people engage in production and only a small number of people engage in the financial service. As a result, the optimal bank capital would be tiny, almost zero, relative to the loan size. Any policy intervention, such as capital adequacy ratio requirement, deposit insurance, and bank bailouts would create distortions.

However, the financial fee may be something required to deal with underlying financial frictions. I assume realistic frictions: costly state verification and limited liability with simple renegotiation. The optimal loan and deposit contracts then take a form of a debt contract. Thus, there will be defaults in the equilibrium and the default thresholds are non-contingent on aggregate shocks. This is the “original sin”: If an environment allows to write state contingent contracts freely, there would be no bankruptcy and no reason to worry about insolvency of banks.

In the equilibrium with these financial frictions, banks assume the risk to receive too low loan repayments to honor the deposit contract. This gives banks an incentive to choose a sizable positive capital as a buffer for themselves and thus for depositors. This limits the size of each bank and determine a number of people working as bankers.

In a rare catastrophic occasion, even banks default. Then, depositors must assume all such risks because of noncontingent limited liability of borrowers and banks. I call this *tail-risk dumping* to depositors. While bankrupt ex-bankers and ex-borrowers can enjoy nice retired life based on retained assets and continued income from their human capital, depositors lose all of their life-time savings. This is a *systemic event* that many governments try to avoid ex post. Ex ante, if they know such event happens, depositors would view deposits as risky investments and hesitate to make a large sum of deposits. As a result, capital would be misallocated.

This is a consequence of incomplete institutional arrangement, in particular, lax limited liability. I do not ask why such institution is installed in this paper, but rather ask what is the optimal policy in the presence of such institutional arrangement. In a way, this paper can be viewed as an optimal policy design in the presence of institutionalized “looting” problems (Akerlof and Romer, 1993) by banks and borrowers.

I show that *transparent bank bailouts* can mitigate the tail-risk dumping problem. I define *transparent bank bailouts* as the one that transfers funds only to depositors, without benefiting banks, by taxing everyone, including defaulted borrowers’ and banks’ retained assets ex post. Literally, taxing defaulters directly is politically or legally difficult but in reality taxing them indirectly can be done by a government, for example, through consumption tax or future income tax on human capital.

In essence, a government can relax the lax limited liability constraint, that is, it can make the noncontingent limited liability constraint to be contingent on aggregate shocks. This is because the government has a granted power to tax people. As a result of a bailout, consumption levels become more similar among borrowers, bankers, and depositors in the case of low tail event. Thus, aggregate risks are more equally shared each other. From the ex ante viewpoint, people face less risk in choosing occupation. Therefore, the overall ex ante welfare is improved by transparent bank bailouts. However, there is no need to call for an additional policy such as a capital adequacy ratio requirement.

A deposit insurance with ex post fee adjustment works similarly well by taxing bankers’ retained assets ex post. Some form of liquidity ratio requirement can also mimic the same outcome. These transfers can be made before the tax collection. In this case, government bonds would be used and then consumption tax would fund the transfer. Also, in the monetary economy, cash can be used as transfers and then inflation tax could fund them.

Note that removing limited liability from bankers would not lead to a better equilibrium. An unlimited liability or so-called “double liability” of bankers would make the banking sector too risky compared to the production sector. As a result, the capital ratio would become too large and so the spread income must be higher than the optimal.

In addition to ex ante limited liability, ex post “looting” opportunity may also be available for banks if banks can seize a part of bailout funds. In this case, bailouts are not transparent but include some hidden subsidies to banks. I call this *untransparent bank bailouts*. Some of them may be necessary to persuade bank owners to agree on bailouts (e.g., Landier and Ueda, 2009) but others may well be a result of political influence by bank lobby (e.g., Igan, Mishra, and Tressel, 2011). Again, I do not attempt to theorize the underlying mechanism of such practices in this paper.

However, if banks are enriched by bailouts, bailout expectations will create distorted incentives for people to become bankers rather than productive entrepreneurs. As a result, there will be too many bankers and too little production. Lower production implies lower entrepreneurs’ utility, and so is bankers’ utility through occupational arbitrage in a general equilibrium. I call this *income shifting* problem. This problem requires a policy to limit bank profits, such as a capital adequacy ratio requirement or a bank levy so as not to attract too many people to become bankers. With these regulations, bank bailouts are still considered as optimal response to a tail-risk event in the presence of lax limited liability.

In summary, my paper introduces two new perspectives, namely *tail-risk dumping* and *income shifting* problems, which are complementary to the existing literature. The tail-risk dumping problem associated with lax limited liability is new to the literature. The limited liability is desirable, if not optimal, in other theories. For example, with debt overhang, heavily indebted firms or people can invest or work more efficiently after debts are forgiven. Another example is a virtue of limited liability corporations, with which more high risk high return projects are made. Because of these benefits, limited liability should not be abolished. However, net benefits of limited liability is unconditionally clear only in a partial equilibrium or with risk neutral agents. In a general equilibrium with risk averse agents, some insurance arrangements must be welfare improving for the tail risks that depositors have to assume under the limited liability. This calls for de facto infringement of limited liability when a large negative shock hits an economy. This is the main, robust, message of this paper.

Also, the income shifting problem in a general equilibrium setup is novel, though distortions in the presence of the government protection in the financial system has been known mostly in a partial equilibrium framework. For example, the risk shifting problem induced by deposit insurance requires prudential regulations such as a capital adequacy ratio requirement in Kareken and Wallace (1978), Keeley (1991), and Allen and Gale (2007). The moral hazard problem from expected bailouts requires prudential regulations in Chari and Kehoe (2009) or tax in Kocherlakota (2010) although Chari and Kehoe (2009) admit the bailout of firms via banks is ex post efficient to avoid assumed fixed costs associated with bankruptcy. In a general equilibrium framework, Van den Heuvel (2008) argues that the capital adequacy ratio requirement is costly as it limits the liquidity available in the general equilibrium. Related issue is the effect of competition policy as regulations such as capital adequacy ratio requirement reduces competition. Some argue that risk taking becomes too excessive under freer competition (Allen and Gale, 2000) because monopolistic rents limit the banks’ risk taking behavior. The others argue the opposite (Boyd and De Nicolo, 2005) because bank’s higher monopolistic rents implies firms’ lower rents that lead to higher risk taking at the firm level.

My paper also serves as an improved micro foundation for macroeconomic models with the financial accelerator (e.g., Bernanke, Gertler, and Gilcrist, 1999). These macro models use the costly state verification associated with lending. But, many models do not have a meaningful banking sector and, instead, firms directly borrow from consumers. Similarly, collateral channel models (Kiyotaki and Moore, 1997) is another popular macro-financial model that addresses financial policies and macroeconomic consequences. However, typical collateral channel models do not also separate banks and firms, and borrowers do not bankrupt in an equilibrium. Recent papers do include a banking sector and sometimes assume two frictions together (e.g., Christiano, Motto, and Rostagno, 2010). Still, firms may default but banks do not default on deposit contracts. With these models, it is difficult to discuss the bank insolvency and government bailouts, which are widely argued as the core issues of financial crises (e.g., Reinhart and Rogoff, 2009).

I would also like to emphasize the importance of endogenous size of the banking sector. In most of the macroeconomic models with financial frictions, bank defaults are absent and the capital ratio is not determined endogenously. Naturally, the banking sector size is exogenously given. Then, it is difficult to identify a full scale of distortions created by policies. For example, the capital adequacy ratio requirement would create higher monopoly rents for bankers in an exogenously given banking sector, but such rents would dissipate with endogenous entry of bankers. Only a few papers have investigated the endogenous nature of the financial sector size. The U.S. financial sector has grown over time with increased bankers' wage that compensates increased bankers' income risk (Phillippon, 2008). In an occupational choice model, Bolton, Santos, and Scheinkman (2011) argues that the financial traders attract too much talents due to profitable opportunities in the opaque OTC market. Their paper apparently brings an important argument but does not have much implications regarding policies towards deposit-taking banks.

II. Setup

A. Demography, Utility, and Technology

I analyze a simple one-period model to understand the basic characteristics of distortions in allocation of factors between the production and financial sectors. The economy has a continuum of ex ante identical households in the interval of $[0, 1]$, endowed with initial capital k_0. They choose to be entrepreneurs and bankers endogenously based on individually optimized choice. I denote bankers' population by μ and entrepreneurs' by $1 - \mu$.

Once an agent becomes an entrepreneur, he observes his talent shock or ideas e to carry out production. He then makes the investment decision on endowed capital k_0. He has an option to make a financial decision how much to make a deposit $s(e)$ or to take a loan $l(e)$. The capital reallocation has to be done via bankers. That is, both financial intermediation and production activities are assumed to require specialization—these are time consuming activities so that specialized people serve in each sector.[2]

With the adjusted amount of capital, he produces final goods subject to productivity shocks. There are three types of shocks which each entrepreneur faces: the idiosyncratic talent (or idea) shocks e (e.g., quality of projects or talent matches) from the cumulative distribution $F(e) : [\underline{e}, \overline{e}] \rightarrow [0, 1]$ with mean one; the idiosyncratic productivity shocks ϵ from the distribution $H(\epsilon) : [\underline{\epsilon}, \overline{\epsilon}] \rightarrow [0, 1]$ also with mean one; and the aggregate productivity shocks A from the cumulative distribution $G(A) : [\underline{A}, \overline{A}] \rightarrow [0, 1]$ with mean greater than one.

Entrepreneurs make individual financial and production decisions after observing the idiosyncratic talent shock, but before aggregate and idiosyncratic productivity shocks hit the production process. [For the sake of simplicity and without loss of generality, I assume hereafter that the lowest aggregate shock is zero, $\underline{A} = 0$ but the mean is one. The idiosyncratic productivity shocks ϵ is assumed to be mean one and always above zero, $\underline{\epsilon} > 0$.] So, the combined productivity shock also has the mean one and zero minimum. Also, for the sake of simplicity, we assume only two levels of talent, e^U and e^D (i.e., up or down) with equal probability $1/2$.

The production function is Cobb-Douglas with capital share $0 < \alpha < 1$ as in a standard macroeconomic model. One unit of labor is assumed to be inelastically supplied by each agent to his own project. This implies that:

$$\begin{aligned} y^D &= y(s, e^D, A, \epsilon) = \epsilon A e^D (k_0 - s)^\alpha \quad \text{for those make deposits;} \\ y^L &= y(l, e^U, A, \epsilon) = \epsilon A e^U (k_0 + l)^\alpha \quad \text{for those take loans.} \end{aligned} \tag{1}$$

The production function exhibits diminishing marginal returns to capital. Entrepreneurs who received the high talent e^U have higher expected marginal returns on the endowed capital than the loan rate. Accordingly, they would like to borrow capital l from banks until the expected marginal returns equate to the effective loan rate. On the other hand, entrepreneurs with low talent e^D have lower expected marginal returns on the endowed capital than the effective deposit rate. They will deposit some of his endowed capital s to banks and operate more productive activities in smaller scale. Their expected marginal returns will become equal to the effective deposit rate.[3] With positive spread between the deposit and loan rates, some entrepreneurs might not engage in transaction with banks. However, in the case with two talents, I assume a sufficient difference between the two so that the high talent type always become borrowers (borrowing l) and the low talent type always become depositors (depositing s) for a small spread.

At the end of the period, a high talent entrepreneur consumes what is left after paying back any outstanding loans and a low talent entrepreneur consumes his own outputs and deposits

[2]For the sake of simplicity, labor is not modeled here, but presumably the labor can be allocated either activity. By choosing occupation, an agent is assumed to establish human capital specific to his occupation.

[3]With risk averse utility, deposit and loan size will be also affected by risk sharing considerations.

returned with interests. Consumption for each type can be written given the loan repayment schedule $R^L(l, A, \epsilon)$ faced by a borrower and the deposit repayment schedule $R^D(s, A)$ faced by a depositor.

$$\begin{aligned} c^L &= c^L(l, A, \epsilon) = y(e^U, A, \epsilon) - R^L(l, A, \epsilon) + (1-\delta)k_0 \quad \text{for those take loans;} \\ c^D &= c^D(s, A, \epsilon) = y(e^D, A, \epsilon) + R^D(s, A) + (1-\delta)k_0 \quad \text{for those make deposits.} \end{aligned} \tag{2}$$

Note that capital is not assumed to depreciate completely. Borrowers may use the capital to repay obligations or to consume. Borrowers repay the loan contract R^L using the depreciated endowed capital $(1-\delta)k_0$ together with outputs. Even in the worst case of zero outputs, borrowers still have the depreciated endowed and loaned capital $(1-\delta)(k_0+l)$ and thus they can always repay the depreciated loaned capital $(1-\delta)l$. Then, banks can always repay at least the depreciated deposits $(1-\delta)s$ as they always receive at least the depreciated loaned capital $(1-\delta)l$ from all the borrowers. Of course, banks may not pay. Therefore, depositors always have the depreciated initial capital $(1-\delta)k_0$ to consume except in a case that the net deposit return R^D is deeply negative.

A banker takes deposits s and make loans l. She also invests her own capital $k_0^B = k_0$ as a part of loans to high talent entrepreneurs. Adjusting the relative size, the resource constraint at a bank can be expressed as:

$$\frac{1-\mu}{2\mu} l = \frac{1-\mu}{2\mu} s + k_0^B. \tag{3}$$

A banker receives and consumes the spread between the two repayment schedules with adjusting the relative size. Note that the banks can pool the idiosyncratic loan repayment and thus the deposit repayment depends only on the aggregate shock.

$$c^B(A; R^L, R^D) = \frac{1-\mu}{2\mu} \int \big(R^L(l, A, \epsilon) - \tau \triangle(\epsilon; A)\big)\, dH(\epsilon) - \frac{1-\mu}{2\mu} R^D(s, A) + (1-\delta)k_0^B, \tag{4}$$

where τ is the verification cost and $\triangle(\epsilon; A)$ is the region where banks verify the state.

Suppose that deposit takes a form of debt contract, which proven to be true in Section III. B below. Then, deposit contract has a flat portion and a default region. Let ρ^D denote a deposit rate in case of full repayment. With the market valuation in the asset side but with the face value liability, the ex ante balance sheet of a bank can be written as

$$\frac{1-\mu}{2\mu} \int \int R^L(l, A, \epsilon) dH(\epsilon) dG(A) = \frac{1-\mu}{2\mu} \rho^D s + w(k_0^B), \tag{5}$$

where $w(k_0^B)$ is the accounting valuation for the net worth, which is the sum of the retained initial capital $(1-\delta)k_0^B$ and any expected profits. Ex post, depending on the realization of the aggregate shock, the net worth can become very small as a bank needs to repay deposit in full until it defaults.

Both entrepreneurs and bankers share the common preferences and obtain utility from consumption goods c. Each household maximizes the expected utility $E[u(c)]$ at the end of the period. For the sake of simplicity, I assume the constant relative risk aversion, that is, $u(c) = c^{1-\sigma}/(1-\sigma)$ with positive relative risk aversion parameter $\sigma > 0$. Note that, the utility function $u : R_+ \to R_+$ is increasing $u' > 0$ and concave $u'' < 0$ and satisfies Inada conditions to assure internal solutions.

B. Decentralized Equilibrium

In a Walrasian decentralized market, there is an auctioneer who offers price and matches demand and supply of goods. This paper's decentralized general equilibrium departs from this typical Walrasian equilibrium in two ways. First, there are continuum of nonatomic banks who intermediate capital markets.[4] Second, banks offer a more general form of "price" in the capital market, that is, deposit and loan repayment schedules. In accordance with the specified repayment schedules, the consumption goods are allocated among borrowers, depositors, and bankers.

Definition 1. *A decentralized equilibrium is the capital allocation, l and s, and the consumption allocation represented by the deposit and loan repayment schedules, $R^D(s, A)$ and $R^L(l, A, \epsilon)$, that satisfy the following conditions:*

- *Given the loan repayment schedule R^L, the expected utility for a high talent entrepreneur (i.e., a borrower) is maximized by her choice of loans l,*

$$V^L(k_0) = \max_l \int \int U\left(c^L(l, A, \epsilon)\right) dG(A)dH(\epsilon). \tag{6}$$

- *Given the deposit repayment schedule R^D, the expected utility for a low talent entrepreneur (i.e., a depositor) is maximized by her choice of deposits s,*

$$V^D(k_0) = \max_s \int \int U\left(c^D(s, A, \epsilon)\right) dG(A)dH(\epsilon). \tag{7}$$

- *Given the reaction of entrepreneurs (i.e., deposit supply and loan demand functions), a banker chooses deposit and loan repayment schedules to maximize his expected utility,*

$$V^B(k_0) = \max_{R^L(l,A,\epsilon), R^D(s,A)} \int U\left(c^B(A; R^L, R^D)\right) dG(A). \tag{8}$$

- *The (pre-production) capital market clears, which is essentially the same as bank's resource constraint,*

$$\frac{1-\mu}{2} l = \frac{1-\mu}{2} s + \mu k_0^B. \tag{9}$$

[4] If banks are finite and possibly monopolistic, strategic actions in both deposit and loan markets could produce a complex equilibrium (Ueda, 2006). Such strategic behaviors are assumed away in this paper.

- *The (after-production) consumption goods market clears for any realization of aggregate shock A,*

$$\begin{aligned} &\frac{1-\mu}{2}\int \left(c^L(l,A,\epsilon)+c^D(s,A,\epsilon)\right)dH(\epsilon)+\mu c^B(A;R^L,R^D) \\ =&\frac{1-\mu}{2}\int \left(y(e^U,A,\epsilon)+y(e^D,A,\epsilon)\right)dH(\epsilon). \end{aligned} \tag{10}$$

- *The ex ante arbitrage condition for occupational choice to become a banker or an entrepreneur before observing talent $e=e^H$ or e^L holds,*

$$V^B(k_0)=V^E(k_0)\equiv\frac{1}{2}V^D(k_0)+\frac{1}{2}V^L(k_0). \tag{11}$$

Note that, with lower banker population μ, by construction, banker's consumption monotonically increases for any realization of A, *ceteris paribus*. This implies that the banker's utility $V^B(A)$ increases monotonically when there are fewer banks (i.e., lower μ), *ceteris paribus*.

III. Optimal Contracts

A. Loan Contract

A bank is assumed to offer an exclusive contract represented by the repayment schedule $R^L(l,A,\epsilon)$ to a borrower. I assume costly state verification (CSV, Townsend, 1979) and limited liability. Also, for the sake of simplicity, I assume a simple renegotiation of payments when default happens.

Assumption 1. ***[Micro Structure of Loan Market]***
(i) [Costly State Verification] Realization of combined productivity shock ϵA is private information but can be verified by a banker with verification cost τ. Note that even aggregate shock is not public information if not inspected.[5]
(ii) [Limited Liability] A defaulter can retain $\lambda>0$ portion of their endowed capital after depreciation.
(iii) [Simple Renegotiation] A bank has a sole bargaining power to recover loans from a defaulter except for the retained λ portion.

Lemma 1. ***[Restriction on Loan Contracts]***
The default threshold and the loan rate ρ^L cannot be aggregate shock contingent. Rather, they are contingent on the combined shock. The repayment in the default region is linear in state. It is determined as all the outputs of a defaulter minus the minimum assets $\lambda(1-\delta)k_0$ that the defaulter keeps.

[5] This captures the situation in which default decision cannot wait until, say, the finalized GDP figures are released by a government. The fact that the deposit is not contingent on the aggregate shock (e.g., GDP) is the "original sin" for the large scale bank insolvency problem. Why it is so merits another research.

Proof. Under Assumption 1 (i), after a measurable set of borrowers are inspected, the aggregate shock should be revealed. However, unless inspected, the aggregate shocks are not revealed. Therefore, the threshold of inspection cannot be contingent on the aggregate shock. Above the threshold, a bank do not know the level of the aggregate shock, so that the loan rate ρ^L cannot be contingent on aggregate or idiosyncratic shocks.

Assumption 1 (ii) implies that the retained assets under limited liability is not contingent on any shocks directly.[6]

In case of default, there is a modeling issue regarding how to allocate the outputs in two parties in renegotiation, in which bargaining powers matter. Here, for the sake of simplicity, Assumption 1 (iii) states that the creditors can seize the assets, except some portion that defaulters can flee with. Since the combined shock determines the output level, the loan recovery rate is contingent on the combined level of the aggregate and idiosyncratic shocks.[7] *Q.E.D.*

In the previous literature with CSV, banks determines the threshold of default, under which repayment becomes contingent on shocks. However, in this paper, the limited liability implies that the borrowers may determine the threshold when they default. Indeed, the latter is the case.

Lemma 2. *There is a unique default threshold θ^L defined for combined shocks ϵA. It is determined by borrowers for any given loan rate ρ^L.*

Proof. I consider only truth-telling strategies, which is indeed supported in an equilibrium as shown below.

Let θ^L_N denote the default threshold determined by a borrower. At this threshold, borrower's consumption from defaulting and not-defaulting should be equated. This decision is after production (in the renegotiation stage) so that the decision is made given some loan size l.

$$\theta^L_N e^H (k_0 + l)^\alpha + (1-\delta)k_0 - (\rho^L + \delta)l = \lambda(1-\delta)k_0 \tag{12}$$

Let θ^L_{CSV} denote the default threshold determined by a bank in a hypothetical situation in which a borrower cannot declare default but a bank decides after looking at the output report

[6]This stems from a deeper assumption on borrowers' ability to walk away from debt obligation. Defaulted borrowers is assumed to be able to flee from their residences or factories with carrying some of their assets. This is usually true in many counties, formally or informally. This is the foundation of the limited liability constraint, exogenously given to the model.

[7]The analysis in this paper should also be robust to other simple allocation of bargaining powers.

from a borrower. In this case, bank's consumption from letting firms default or not should be the same at this hypothetical threshold.[8]

$$\theta^L_{CSV} e^H (k_0 + l)^\alpha + (1 - \lambda)(1 - \delta)k_0 - \tau = (\rho^L + \delta)l. \tag{13}$$

The relative size of two thresholds are given simply by the cost of verification:

$$\theta^L_{CSV} > \theta^L_N \Leftrightarrow \tau > 0. \tag{14}$$

This implies that in the region where combined shock realization is below or equal to θ^L_N, a bank has an incentive to inspect with paying cost τ. Moreover, a bank optimally adopts the threshold determined by the borrower's limited liability constraint, that is, $\theta^L = \theta^L_N$. By adopting this, a bank gains over the case with adopting θ^L_{CSV}. By using borrower's self-selected default threshold θ^L_N, banks do not have to pay τ for realized states between θ^L_N and θ^L_{CSV}. With this banks' policy, between θ^L_N and θ^L_{CSV}, borrowers also do not have to lower their consumption at the retained asset level, which is lower than the full repayment case. *Q.E.D.*

Remark. *Banks need to offer the loan repayment schedule that is consistent with the borrowers' default decision, (12).*

In summary, the optimal loan repayment schedule $R^{L*}(l, A, \epsilon)$ has default region $[0, \theta^L]$ and flat full-pay loan rate ρ^L. Moreover, the optimal loan contract implies that
(i) [Pooling Idiosyncratic Shocks] Each bank pools idiosyncratic shocks of borrowers perfectly.
(ii) [Public Information] Loan repayments are public information and thus the aggregate shocks are revealed to everyone except for the case that all borrowers repay in full.

The optimal loan repayment schedule $R^{L*}(l, A, \epsilon)$ is described as follows:

- Above the threshold θ^L, a borrower repays in full, $(\rho^L + \delta)l$.
- A defaulter retains $\lambda(1 - \delta)k_0$.
- A defaulter's repayment schedule has an intercept term and a linearly increased portion with respect to the realized (combined) productivity shock ϵA:

$$\epsilon A e^H (k_0 + l)^\alpha + (1 - \lambda)(1 - \delta)k_0. \tag{15}$$

- At the threshold θ^L, this repayment (15) equals to the full loan repayment, $(\rho^L + \delta)l$.

[8]For a banker, against the fixed cost τ to verify the state, the marginal benefit of verifying the state is strictly decreasing in state. If reported outputs are suspiciously low, they should be investigated, but otherwise they would not carry out costly investigation. Thus, there is a threshold of verification. Without limited liability, the optimal contract with CSV usually implies the risk sharing between two parities below the default threshold. The literature so far identifies that it becomes a straight line for risk neutral agents (see a review by Fulghieri and Goldman, 2008).

Accordingly, a defaulted entrepreneur consumes only

$$c^L(l, A, \epsilon) = \lambda(1-\delta)k_0, \quad \text{if} \quad \epsilon A \in [\underline{\epsilon A}, \theta^L]. \tag{16}$$

Otherwise, his consumption is

$$c^L(l, A, \epsilon) = \epsilon A e^H (k_0 + l)^\alpha + (1-\delta)k_0 - (\rho^L + \delta)l, \quad \text{if} \quad \epsilon A \in [\theta^L, \overline{\epsilon A}]. \tag{17}$$

B. Deposit Contract

After production occurs, borrowers decide defaulting or not and then banks decide defaulting or not. Depositors have to anticipate their behaviors before they make deposit decision. Naturally, loan contracts and amounts are optimally chosen contingent on realization of shocks and then deposit contracts and amounts are chosen expecting such loan market outcome.

Lemma 3. *There is a flat portion in an equilibrium deposit repayment schedule.*

Proof. Remark to 2 implies that each bank insures against idiosyncratic shocks of borrowers. By Lemma 1, the loan default threshold and the loan rate is not contingent on the aggregate shocks. Therefore, when all borrowers repay the loans in full, bank revenue is flat, non-contingent on any shocks. Moreover, the aggregate shocks are not revealed in this case. Hence, even if depositors have equity-type claim on bank revenue, the returns have a flat portion above some threshold level of aggregate shock realization. *Q.E.D.*

For the deposit market, I do not assume costly state verification but assume the similar assumptions for limited liability as only bankers are supposed to specialized to conduct verification on states.

Assumption 2. *[Micro Structure of Deposit Market]*
(i) [Default Trigger] Default of a banker is defined when he cannot repay the flat deposit rate in full.
(ii) [Limited Liability] If a banker defaults, he can retain $\lambda > 0$ portion of the invested capital after depreciation.
(iii) [Simple Renegotiation] Depositors have a sole bargaining power to recover deposits from a defaulted banker except for his retained portion.

The optimal deposit contract can be characterized almost directly from Assumption 2, similar to the case with the loan contract.

Lemma 4. *The optimal deposit contract looks like a standard debt contract. The deposit repayment schedule $R^D(s, A)$ has a flat portion representing full-pay deposit rate ρ^D above the threshold θ^D defined for aggregate shock A. Below θ^D is the bank default region, and repayment depends on the realization of aggregate shocks A. In this case, depositors seize the remaining bank assets except that bankers retain λ portion of their endowed capital (after depreciation).*

Proof follows directly from Lemma **??** and Assumption 2.

Remark. *The non-default region is larger than the region in which aggregate state is not revealed.*

By construction, the region of flat deposit repayment is at least as large as the region where the aggregate shock is not revealed. However, if defaulted, a bank loses everything except for the retained assets because of Assumption 2. This penalty provides a banker an incentive to repay deposits in full out of his own capital even some borrowers default. Because of this capital buffer of a banker, banker's default threshold is lower than borrowers' average default threshold (on aggregate shock). And, deposit repayment schedule is not contingent on aggregate shock even in the region between the borrower's average default threshold and banker's default threshold.

Note that the uniqueness and characterization of the deposit contract $R^D(s, A)$ is the key issue and determined in an equilibrium as shown in sections below. Accordingly, the equilibrium deposits and depositor's utility are determined endogenously in an equilibrium as well.

The banker's deposit repayment schedule $R^D(s, A)$ is defined as follows. If it can, a banker repays full obligation to a depositor:

$$R^D(s, A) = (\rho^D + \delta)s, \quad \text{if } A \in [\theta^D, \overline{A}]. \tag{18}$$

A banker can also retain λ portion of their endowed capital when they default. In this case, similar to the entrepreneur's case, the repayment function to a depositor has an intercept term and a linearly increased portion, correcting for the relative size of the banking sector:

$$R^D(s, A) = \frac{2\mu}{1-\mu}\left((1-\lambda)(1-\delta)k_0^B + B(A)\right), \quad \text{if } A \in [\underline{A}, \theta^D]. \tag{19}$$

where $B(A)$ denote a banker's gross income. It can be expressed as a function of the aggregate shocks only, after correcting for the relative size of the banking sector:

$$\begin{aligned} B(A) &= \frac{1-\mu}{2\mu}\int_{\underline{\epsilon}}^{\overline{\epsilon}} \left(R^L(A,\epsilon) - \tau\triangle(\epsilon; A)\right) dH(\epsilon) \\ &= \frac{1-\mu}{2\mu}\left(1 - H\left(\frac{\theta^L}{A}\right)\right)(\rho^L + \delta)l \\ &+ \frac{1-\mu}{2\mu}\int_{\underline{\epsilon}}^{\frac{\theta^L}{A}} \left((1-\lambda)(1-\delta)k_0 + \epsilon A e^H (k_0 + l)^\alpha - \tau\right) dH(\epsilon), \end{aligned} \tag{20}$$

In a region where a borrower does not default (i.e., $\epsilon A > \theta^L$), repayment from the borrower is constant $(\rho^L + \delta)$. On the other hand, in a region where a borrower default (i.e., $\epsilon A < \theta^L$), repayment from the borrower is increasing with the aggregate shock A with the floor, which is the collateral value $(1-\lambda)(1-\delta)k_0$. Overall, banker's gross income $B(A)$ is increasing in aggregate shock A.

C. Bankers' Choice and Consumption

Banks are fully competitive. In the beginning of the period, bankers offer deposit and loan rates by equating their expected utility from the spread income to their reservation utilities, which are based on the expected income of being an entrepreneur.[9] When banks offer deposit and loan rates, they rationally expect the possibility of defaults of both borrowers and themselves.

A banker's net income is the gross income net of (size-corrected) repayments. In case of default, it is just the retained capital

$$c^B(D,A) = \lambda(1-\delta)k_0^B, \quad \text{if} \quad A \in [\underline{A}, \theta^D]. \tag{21}$$

If every borrower repays and a banker repays, a banker obtains the fee income,

$$c^B(D,A) = \frac{1-\mu}{2\mu}(\rho^L - \rho^D)s + (1-\delta+\rho^L)k_0^B, \quad \text{if} \quad A \in [\frac{\theta^L}{\underline{\epsilon}}, \overline{A}]. \tag{22}$$

In between, a banker repays deposits in full to depositors but receives less than full loan payments from borrowers,

$$c^B(D,A) = B(A) + (1-\delta)k_0^B - \frac{1-\mu}{2\mu}(\rho^D + \delta)s, \quad \text{if} \quad A \in [\theta^D, \frac{\theta^L}{\underline{\epsilon}}]. \tag{23}$$

IV. DECENTRALIZED ALLOCATION

Although the utility function and the production function take orthodox forms, the deposit and loan repayment schedules have kinks. A natural question is whether equilibrium is unique. I show it is by a constructive analysis on (i) the partial equilibrium in the loan market, (ii) the partial equilibrium in the deposit market, and (iii) the general equilibrium.

A. Loan Market Partial Equilibrium

For any given amount of loans l, the *iso-loan supply function* can be drawn on the θ^L-ρ^L plane based on the default condition, (12):

$$\rho^L + \delta = \theta^L e^U \frac{(k_0 + l)^\alpha}{l} + (1-\lambda)(1-\delta)\frac{k_0}{l}. \tag{24}$$

Lemma 5. *The iso-loan supply function is monotonically increasing.*

[9] I assume no changes in deposit and loan contracts within the period but negotiation after default is expected.

Proof. The slope is the derivative of the right hand side of the iso-loan supply function with respect to θ^L:

$$e^U \frac{(k_0 + l)^\alpha}{l} > 0. \tag{25}$$

Q.E.D.

On the same θ^L-ρ^L plane, *iso-loan demand function* can be also drawn. It is based on the optimal decision by borrowers. Let $\eta \equiv \epsilon A$ denote the combined shock with the cdf $M \equiv G \circ H$. The first order condition for the borrower's problem (6) is

$$\int_{\theta^L}^{\overline{\epsilon A}} \left(\alpha \eta e^U (k_0 + l)^{\alpha - 1} - (\rho^L + \delta)\right) U'(c^L) dM(\eta) = 0. \tag{26}$$

This is essentially the optimal leverage problem for a limited-liability entrepreneur who borrows capital until the marginal productivity of capital equals to the loan rate but only for the non-default region. The *iso-loan demand function* of the loan rate ρ^L with respect to the default threshold θ^L given loan amount l is expressed as an implicit function of

$$\chi(\theta^L, l) \equiv \int_{\theta^L}^{\overline{\epsilon A}} \left(\alpha \eta e^H (k_0 + l)^{\alpha - 1} - (\rho^L + \delta)\right) U'(c^L) dM(\eta). \tag{27}$$

Lemma 6. *For any given amount of loans l, the iso-loan demand function on the θ^L-ρ^L plane is monotonically decreasing.*

Proof. The derivative of the iso-loan demand function with respect to θ^L is negative:

$$\frac{\partial \chi(\theta^L, l)}{\partial \theta^L} = \left(-\alpha \theta^L e^U (k_0 + l)^{\alpha - 1} - (\rho^L + \delta)\right) U'(\underline{c}^L) M(\theta^L) < 0, \tag{28}$$

where $\underline{c}^L$ is evaluated at $\eta = \theta^L$. The derivative with respect to ρ^L is

$$\begin{aligned}
\frac{\partial \chi(\theta^L, l)}{\partial \rho^L} &= \int_{\theta^L}^{\overline{\epsilon A}} \left(-U'(c^L) - (\alpha \eta e^H (k_0 + l)^{\alpha - 1} - (\rho^L + \delta)) l U''(c^L)\right) dM \\
&= - \int_{\theta^L}^{\overline{\epsilon A}} U'(c^L) dM - \sigma l \int_{\theta^L}^{\overline{\epsilon A}} (\alpha \eta e^H (k_0 + l)^{\alpha - 1} - (\rho^L + \delta)) \frac{U'(c^L)}{c^L} dM \\
&< 0.
\end{aligned} \tag{29}$$

Note that inside the second integral in the penultimate line has a “weight” of U'/c^L, which has higher weights for the lower realization of shocks and lower weights for the higher realization of shocks compared to the weight U' in the (27). Because the second integral is different only in this “weight” from the borrower's first order condition (27) valued at zero, the second integral must be negative.

In summary,

$$\frac{d\rho^L}{d\theta^L} = -\frac{\partial \chi / \partial \theta^L}{\partial \chi / \partial \rho^L} < 0. \tag{30}$$

Q.E.D.

Proposition 1. *In a partial equilibrium of the loan market, there exists a unique loan contract (loan repayment schedule), characterized by a unique set of loan rate and default threshold (ρ^L, θ^L), for each loan amount l. Moreover, the equilibrium loan amount l^* is strictly decreasing in loan rate ρ^L.*

Proof. On the θ^L-ρ^L plane, because the iso-loan supply curve is increasing (Lemma 5) and the iso-loan demand curve is decreasing (Lemma 6), there is a unique set of the loan rate and the default threshold (ρ^L, θ^L) to satisfy both demand and supply.

To see the equilibrium relationship between the loan amount l and the default threshold θ^L, I investigate how two curves shift with higher loan amount. For the iso-loan demand function,

$$\begin{aligned}&\frac{\partial \chi(\theta^L, l)}{\partial l}\\&= \int_{\theta^L}^{\overline{\epsilon A}} (\alpha - 1)\alpha\eta e^H (k_0 + l)^{\alpha-2} U'(c^L) dM + \int_{\theta^L}^{\overline{\epsilon A}} (\alpha\eta e^H (k_0 + l)^{\alpha-1} - (\rho^L + \delta))^2 U''(c^L) dM\\&< 0.\end{aligned} \tag{31}$$

Because $\partial\chi(\theta^L, l)/\partial\rho^L < 0$ and $\partial\chi(\theta^L, l)/\partial\theta^L < 0$ as shown already, considering the implicit functions, it is obvious that an increase in the loan amount is accompanied by decline in both loan rate ρ^L and threshold θ^L. That is, the iso-loan demand function shifts down to the origin on the θ^L-ρ^L plane with higher loan amount.

The derivative of the right hand side of the iso-loan supply curve (24) with respect to loan amount l is

$$\theta^L e^U \frac{(k_0 + l)^{\alpha-1}}{l} \left(\alpha - \frac{k_0 + l}{l}\right) - (1 - \lambda)(1 - \delta)\frac{k_0}{l^2} < 0. \tag{32}$$

Note that $\alpha < 1$ and $(k_0 + l)/l > 1$. Therefore, the iso-loan supply curve shifts down with a higher loan amount.

Because both curves shift down with higher loan amount on the θ^L-ρ^L plane, the equilibrium loan amount l is decreasing with loan rate ρ^L.

Q.E.D.

B. Deposit Supply by Depositors

Proposition 2 (Optimal Deposit Size). *Given a deposit contract $R^D(s, A)$, a depositor decides deposit amount s to maximize his utility (7). This is uniquely determined.*

Proof. The first order condition with respect to deposits s is

$$\begin{aligned}&\int_{\underline{A}}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(c^D(s,A,\epsilon))\alpha\epsilon Ae^L(k_0-s)^{\alpha-1}dH(\epsilon)dG(A)\\ =&\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(c^D(s,A,\epsilon))\frac{\partial R^D(s,A)}{\partial s}dH(\epsilon)dG(A),\end{aligned} \tag{33}$$

where

$$\begin{aligned}c^D(s,A,\epsilon)&=\epsilon Ae^L(k_0-s)^\alpha+(1-\delta)k_0+(\rho^D+\delta)s, \quad \text{if } A\geq\theta^D;\\ &=\epsilon Ae^L(k_0-s)^\alpha+(1-\delta)(k_0-s)+(1-\lambda)(1-\delta)k_0^B, \quad \text{if } A<\theta^D.\end{aligned} \tag{34}$$

is the consumption of a depositor. Note that the derivative of $R^D(s,A)$ with respect to s is equal to $(\rho^D+\delta)$ in the nondefault region and $-(1-\delta)$ otherwise. Then, the right hand side of (33) is equal to

$$\begin{aligned}&=(\rho^D+\delta)\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(c^D(s,A,\epsilon))dH(\epsilon)dG(A)\\ &-(1-\delta)\int_{\underline{A}}^{\theta^D}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(c^D(s,A,\epsilon))dH(\epsilon)dG(A).\end{aligned} \tag{35}$$

The first order condition (33) essentially is the optimal portfolio problem of allocating capital so as to equate the internal marginal product from own business to the outside opportunity, which is the deposit to banks. Similar to loan size determination by borrowers, there is unique solution of deposit size. *Q.E.D.*

For given k_0 and parameter values of the production and utility as well as the deposit repayment schedule R^D, the utility level is determined in the equilibrium by optimally chosen deposit s. For a specific utility level, the first order condition (33) gives us the deposit supply as a function of the deposit repayment schedule (i.e., the default threshold θ^D and the full-pay deposit rate ρ^D). For given full-pay loan rate ρ^L, the full-pay deposit rate is just lower as much as the spread, that is, $\rho^D=\rho^L-\pi$. Here, the (Hicks) deposit supply function can be expressed as $s^s(\theta^D,\pi|\overline{u})$ in the θ^D-π plane.

If more capital is allocated to the deposit, less capital is allocated to own business and raises the marginal product of capital. Then, the left hand side of (33) increases. On the other hand, more capital would be allocated to the deposit if the deposit contract becomes safer (i.e., the default threshold θ^D becomes smaller) for the same deposit rate ρ^D. Thus, the deposit supply function $s^s(\theta^D,\pi|\overline{u})$ is decreasing in default threshold θ^D. Similarly, it is decreasing in spread π, that is, increasing in deposit rate ρ^D $(=\rho^L-\pi)$.

On the θ^D-π plane, given spread π, deposit supply s^s is decreasing in the threshold θ^D and vice versa. Here, for any given deposit supply level, spread π as a function of default threshold θ^D is strictly decreasing. That is, if the deposit rate is lowered, depositors ask to lower default threshold for compensation to deposit the same amount in a bank compared to returns form own business. The θ^D-π deposit supply line represents an *iso-deposit supply curve*. A higher deposit rate and a lower default probability give the depositors higher utility. The deposits (and the utility level) are larger if the iso-deposit supply curve is closer to the origin (lower spread and lower threshold).

Given the FOC-satisfying deposit level $\bar{s}$ fixed, the slope of this iso-deposit supply curve on the θ^D-π plane is determined by the first order condition (33). Let $\Phi(\theta^D, \pi)$ denote the right-hand-side minus the left-hand-side of the first order condition—it is zero at the optimum. For the same level of deposits $\bar{s}$, the slope the iso-deposit supply curve is:[10]

$$\frac{d\pi}{d\theta^D} = \frac{-\partial\Phi(\theta^D, \pi)/\partial\theta^D}{\partial\Phi(\theta^D, \pi)/\partial\pi}. \tag{36}$$

Here, with $U' > 0$ and $U'' < 0$,

$$-\frac{\partial\Phi(\theta^D, \pi)}{\partial\theta^D} = (\rho^D+\delta)g(\theta^D)\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(\epsilon\theta^D e^L(k_0-\bar{s})^\alpha+(1-\delta)k_0+(\rho^D+\delta)\bar{s})dH(\epsilon) > 0, \tag{37}$$

where $g(A)$ is pdf for cdf $G(A)$, and

$$\frac{\partial\Phi(\theta^D, \pi)}{\partial\pi} = -\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'dH(\epsilon)dG(A) - \bar{s}\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}}(\rho^D+\delta-\alpha\epsilon Ae^L(k_0-s)^{\alpha-1})U''dH(\epsilon)dG(A). \tag{38}$$

So far, the model is too general and difficult to characterize. Hereafter, I focus on the empirically relevant range of parameter values by the following assumptions. Mostly they are restrictions on productivity shock distributions to assure convexity of the deposit supply and existence of pure strategies. Without them, the most of the argument still goes through if lotteries or mixed strategies are introduced to convexify the deposit supply curve.

Assumption 3. *[Regularity Assumptions]*
(i) [Left Tail for Aggregate Shocks]: The pdf is increasing in the neighborhood of low aggregate shocks that trigger a bank default, that is, $g'(\theta^D) > 0$. Also, in that neighborhood, the elasticity of the pdf is higher than the half of the relative risk aversion, that is,

$$\frac{g'(\theta^D)\theta^D}{g(\theta^D)} > \frac{\sigma}{2}.$$

(ii) [Return Bound]: The average marginal product of capital conditional on the highest realization of aggregate shock for a high talent entrepreneur is bounded above,

$$\alpha\overline{A}e^U k_0^{\alpha-1} \leq 2-\delta.$$

[10] The derivative changes at default threshold θ^D depending on whether depositors expect default or not by banks at the threshold. I assume commitment of payment by banks at the threshold and therefore the derivative is taken from the "right" side.

Lemma 7. *The iso-deposit supply curve on the θ^D-π plane is strictly decreasing and strictly concave in threshold θ^D.*

This Lemma implies that, for a low threshold of default, depositors tolerate a high spread (i.e., a low deposit rate).

Proof. The iso-deposit supply curve is strictly decreasing if the denominator (38) is negative—because the numerator 37 is positive, it means that the overall slope (36) is negative.

$$\begin{aligned}
&-\frac{\partial\Phi(\theta^D,\pi)}{\partial\pi}\\
&=\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U' dH(\epsilon)dG(A) + \bar{s}\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} (\rho^D+\delta-\alpha\epsilon A e^L(k_0-s)^{\alpha-1})U'' dH(\epsilon)dG(A)\\
&=\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U' + \bar{s}(\rho^D+\delta-\alpha\epsilon A e^L(k_0-s)^{\alpha-1})\frac{U''}{U'}U' dH(\epsilon)dG(A)\\
&=\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} U' - (\rho^D+\delta-\alpha\epsilon A e^L(k_0-s)^{\alpha-1})\frac{\bar{s}}{c^D}\frac{-c^D U''}{U'}U' dH(\epsilon)dG(A)\\
&>\sigma\frac{\bar{s}}{\tilde{c}^D}\int_{\theta^D}^{\overline{A}}\int_{\underline{\epsilon}}^{\overline{\epsilon}} \left((1-\delta)+(\alpha\epsilon A e^L(k_0-s)^{\alpha-1})-\rho^D)\right) U' dH(\epsilon)dG(A).\\
&>-\Phi(\theta^D,\pi)=0.
\end{aligned} \tag{39}$$

In the penultimate line, $\tilde{c}^D$ denotes the consumption level conditional on the realization of highest aggregate and idiosyncratic shocks. Note that the integral portion of the formula is very similar to the first order condition (33) but the aggregate shocks are truncated at θ^D to calculate the expected return from own business. Therefore, it is positive and its reverse is negative,

$$\frac{\partial\Phi(\theta^D,\pi)}{\partial\pi}<0. \tag{40}$$

For the concavity, I show that the slope (36) is strictly decreasing with θ^D, or equivalently $d^2\pi/(d\theta^D)^2<0$.

$$\frac{d^2\pi}{(d\theta^D)^2}=\frac{-\partial^2\Phi(\theta^D,\pi)/\partial\theta^D\partial\theta^D}{\partial\Phi(\theta^D,\pi)/\partial\pi}-\frac{-\partial\Phi(\theta^D,\pi)/\partial\theta^D}{(\partial\Phi(\theta^D,\pi)/\partial\pi)^2}\frac{\partial^2\Phi(\theta^D,\pi)}{\partial\pi\partial\theta^D}. \tag{41}$$

The numerator of the first term is

$$
\begin{aligned}
&-\frac{\partial^2\Phi(\theta^D,\pi)}{\partial\theta^D\partial\theta^D}\\
=&(\rho^D+\delta)g'(\theta^D)\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(\epsilon\theta^D e^L(k_0-s)^\alpha+(1-\delta)k_0+(\rho^D+\delta)s)dH(\epsilon)\\
+&(\rho^D+\delta)g(\theta^D)\int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon e^L(k_0-s)^\alpha U''(\epsilon\theta^D e^L(k_0-s)^\alpha+(1-\delta)k_0+(\rho^D+\delta)s)dH(\epsilon)\\
=&(\rho^D+\delta)\int_{\underline{\epsilon}}^{\overline{\epsilon}}\left(g'(\theta^D)U'(c^D(s,\theta^D,\epsilon))+g(\theta^D)\epsilon e^L(k_0-s)^\alpha U''(c^D(s,\theta^D,\epsilon))\right)dH(\epsilon)\\
=&(\rho^D+\delta)\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(c^D(s,\theta^D,\epsilon))\\
&\left(g'(\theta^D)+g(\theta^D)\frac{c^D(s,\theta^D,\epsilon)U''(c^D(s,\theta^D,\epsilon))}{U'(c^D(s,\theta^D,\epsilon))}\frac{y(e^D,\theta^D,\epsilon)}{\theta^D c^D(s,\theta^D,\epsilon)}\right)dH(\epsilon)\\
=&(\rho^D+\delta)\int_{\underline{\epsilon}}^{\overline{\epsilon}} U'(c^D(s,\theta^D,\epsilon))\left(g'(\theta^D)-\frac{g(\theta^D)}{\theta^D}\sigma\frac{y(e^D,\theta^D,\epsilon)}{c^D(s,\theta^D,\epsilon)}\right)dH(\epsilon)\\
>&(\rho^D+\delta)U'(\overline{c}^D)\int_{\underline{\epsilon}}^{\overline{\epsilon}}\left(g'(\theta^D)-\frac{g(\theta^D)}{\theta^D}\sigma\frac{y(e^D,\theta^D,\epsilon)}{c^D(s,\theta^D,\epsilon)}\right)dH(\epsilon)\\
>&(\rho^D+\delta)U'(\overline{c}^D)\left(g'(\theta^D)-\frac{g(\theta^D)}{2\theta^D}\sigma\right)\\
>&0,
\end{aligned}
\tag{42}
$$

where consumption $\overline{c}^D$ in the last three lines denote the consumption evaluated at the highest possible values, that is with $\epsilon=\overline{\epsilon}$, under $A=\theta^D$. The penultimate line follows that the average depositor's consumption c^D conditional on the aggregate shock at θ^D is strictly twice larger than the depositor's own output y for any realization of the idiosyncratic shocks. This is because at θ^D a depositor receives full repayments from a bank and she can also consume her own output and the (depreciated) initial capital. Note that at θ^D, average borrowers are defaulting by definition, that is, the average output of borrowers are less than the depreciated initial capital, and so does the average own-output of depositors. That is,

$$
\begin{aligned}
c^D &= y+(\rho^D+\delta)+(1-\delta)k_0\\
&> y+(1-\delta)k_0\\
&> 2y.
\end{aligned}
\tag{43}
$$

Finally, by Assumption 3 (i), the last big parenthesis is strictly positive.

The denominator of the first term of the derivative of the slope (41) is negative as shown in (38). Thus, the first term is negative overall. The last portion of the second term of the derivative of the slope (41) is

$$
\begin{aligned}
&\frac{\partial^2 \Phi(\theta^D, \pi)}{\partial \pi \partial \theta^D} \\
=&\int_{\underline{\epsilon}}^{\overline{\epsilon}} U' \left((1-\delta) + (\alpha\epsilon\theta^D e^L (k_0 - s)^{\alpha-1} - \rho^D)\right) \frac{\bar{s}}{c^D} \frac{-c^D U''}{U'} dH(\epsilon) \\
>&\frac{\bar{s}}{\overline{c}^D} \sigma U'(\overline{c}^D) \int_{\underline{\epsilon}}^{\overline{\epsilon}} \left(1 - \delta - \rho^D + \alpha\epsilon\theta^D e^L (k_0 - s)^{\alpha-1}\right) dH(\epsilon) \\
=&\frac{\bar{s}}{\overline{c}^D} \sigma U'(\overline{c}^D) \left(1 - \delta - \rho^D + \alpha\theta^D e^L (k_0 - s)^{\alpha-1}\right) \\
>&0.
\end{aligned} \tag{44}
$$

where consumption inside U' and U'' in the first line are evaluated at $A = \theta^D$; and $\overline{c}^D$ in the second line and below denote the consumption evaluated at the highest possible values, that is with $\epsilon = \overline{\epsilon}$, under $A = \theta^D$. The penultimate line uses the assumption that the mean of idiosyncratic shock ϵ is one. The last line follows Assumption 3 (ii) and $1 + \rho^D \leq 1 + \rho^L < \alpha\overline{A}e^U k_0^{\alpha-1}$, the average autarkic marginal product of capital for a high talent entrepreneur contingent on the highest aggregate shock. That is, $1 - \delta - \rho^D > 0$.

The first portion of the second term of the derivative of the slope (41) is positive as shown in (37). With the minus sign in front, the second term is negative overall. Because both the first and second terms of the derivative of the slope (41) are proven to be negative, the iso-deposit supply curve is strictly concave. *Q.E.D.*

C. Credible Deposit Demand by Bankers

The deposit market is under unfettered competition. A banker offers a deposit contract, which specify deposit repayment function. However, as long as the spread income is positive, a banker is happy to take as much deposits as possible. That is, the deposit demand by a bank is inelastic to any given profitable pair of spread and default thresholds.

A banker faces a credibility constraint: He needs to repay deposit in full $\rho^D s$ under a deposit contract $R^D(s, A)$ up to the default threshold θ^D. This means that there is a technical trade-off between the threshold θ^D and the spread π. Given certain spread π, higher default threshold θ^D (i.e., easier to default on deposits) gives a banker more profits but the threshold cannot be higher than the contract specifies. That is, the contract needs to be renegotiation proof or credible. This implies that a banker maximizes his utility by choosing the threshold θ^D for any given full pay deposit rate such that consumption under default is equal to consumption under full deposit repayment:

$$\lambda(1-\delta)k_0^B = B(\theta^D) + (1-\delta)k_0^B - \frac{1-\mu}{2\mu}(\rho^D + \delta)s. \tag{45}$$

Given deposits s, the initial capital k_0^B, and banker population μ, this renegotiation proof constraint (45) should hold and appears as *credible deposit contract offer curve* $s^d(\theta^D, \pi)$ on the θ^D-π plane.

For the sake of simplicity, I assume the following property of the pdf function for the idiosyncratic shock.

Assumption 4. *[Right Tail for Idiosyncratic Shocks] In the right tail, for* $\forall \epsilon > mean(\epsilon) = 1$, *the pdf is decreasing* $h'(\epsilon) < 0$ *but not steeply—the elasticity is less than two,* $|\epsilon h'/h| \leq 2$ *and the elasticity of the cdf is less than one,* $\epsilon h/H \leq 1$.

Lemma 8. *The credible deposit contract offer curve is strictly decreasing and strictly convex on the* θ^D-π *plane.*

Proof. On the θ^D-π plane, given a deposit level s, the credible deposit contract offer curve appears as,

$$\pi - \rho^L - \delta = -\frac{2\mu}{(1-\mu)s}\left(B(\theta^D) + (1-\lambda)(1-\delta)k_0^B\right). \tag{46}$$

Using (20), it is expressed as

$$\begin{aligned}\pi - \rho^L - \delta = &-\frac{2\mu}{(1-\mu)s}(1-\lambda)(1-\delta)k_0^B \\ &- \frac{1}{s}\left(1 - H\left(\frac{\theta^L}{\theta^D}\right)\right)(\rho^L + \delta)l \\ &- \frac{1}{s}\int_{\underline{\epsilon}}^{\frac{\theta^L}{\theta^D}} \left((1-\lambda)(1-\delta)k_0 + \epsilon\theta^D e^U (k_0 + l)^\alpha - \tau\right) dH(\epsilon).\end{aligned} \tag{47}$$

Multiply both sides by s and take a derivative of the right hand side with respect to θ^D:

$$\begin{aligned}&-(\rho^L+\delta)l\frac{\theta^L}{\theta^{D2}}h + \left((1-\lambda)(1-\delta)k_0 + \frac{\theta^L}{\theta^D}\theta^D e^U(k_0+l)^\alpha - \tau\right)\frac{\theta^L}{\theta^{D2}}h \\ &- \int_{\underline{\epsilon}}^{\frac{\theta^L}{\theta^D}} \epsilon e^U (k_0+l)^\alpha dH(\epsilon) \\ =&-\left[(\rho^L+\delta) - \left((1-\lambda)(1-\delta)k_0 + \theta^L e^U(k_0+l)^\alpha - \tau\right)\right]\frac{\theta^L}{\theta^{D2}}h \\ &- \int_{\underline{\epsilon}}^{\frac{\theta^L}{\theta^D}} \epsilon e^U (k_0+l)^\alpha dH(\epsilon) \\ =&-\tau\frac{\theta^L}{\theta^{D2}}h - \int_{\underline{\epsilon}}^{\frac{\theta^L}{\theta^D}} \epsilon e^U (k_0+l)^\alpha dH(\epsilon) < 0,\end{aligned} \tag{48}$$

where pdf h is evaluated at θ^L/θ^D. This is negative and thus the credible deposit contract offer curve is strictly decreasing. Note that, the last line follows the fact that the full loan repayment $\rho^L + \delta$ is equal to the all the outputs plus net-of-retained portion of depreciated capital of a borrower at the default threshold of a loan (i.e., ϵA is at θ^L).

As for the convexity, take the second derivative with respect to θ^D:

$$\tau \frac{\theta^L}{\theta^{D3}} \left(\frac{\theta^L}{\theta^D} h' + 2h \right) + \frac{\theta^L}{\theta^D} e^U (k_0 + l)^\alpha \frac{\theta^L}{\theta^{D2}} h > 0, \tag{49}$$

where pdf h and its derivative h' are evaluated at θ^L/θ^D. Note that θ^L/θ^D is by construction above one, which is the mean value of idiosyncratic shocks. Thus, the inside of the large bracket in the first term is positive or zero by Assumption 4. The last term is apparently positive. So, the second derivative is positive overall, that is, the credible demand contract offer curve is strictly convex. *Q.E.D.*

D. Deposit Market Partial Equilibrium

It may be obvious but the demand and supply needs to be equal in an equilibrium even if the supply is limited by the resource constraint and the deposit contract allows the possibility of default.

Lemma 9. *The deposit contract* (θ^D, π, s) *is an equilibrium only if it is a tangential point of the iso-deposit supply curve and the credible deposit contract offer curve. That is, the deposit supply equals to the deposit demand,* $s^s(\theta^D, \pi) = s^d(\theta^D, \pi) = s^*$.

Proof. Lemma 7 (strictly decreasing and concave iso-deposit supply curve) and Lemma 8 (strictly decreasing and convex credible deposit contract curve) provide a unique tangential point of these curves. However, both are fixed in the $\theta^D - \pi$ plane at a specific level of deposits. Consider the case in which both curves are plotted at s^*, the level that satisfy the banker's resource constraint. There are three cases:

(i) two curves are tangent each other at (θ^{D*}, π^*);

(ii) two curves are apart and never cross; or

(iii) two curves cross twice at (θ^{D1}, π^1) and (θ^{D2}, π^2).

In case (i), the tangential point is where the deposit supply and demand are equal and there is no Pareto superior allocation. Therefore, it is an equilibrium in the deposit market.

In case (ii), depositors do not agree with the deposit contract offer. This is not an equilibrium.

In case (iii), a banker could offer more attractive contract, which is still on the same credible deposit contract offer curve, but it is an inner point of depositors' iso-deposit supply curve. Then, there would be deposit rationing. In other words, bankers can offer a little lower deposit rate for the same threshold and can still take deposit as much as before. This means that the current repayment schedule R^D is not maximizing bankers' utility. Therefore, this is not an equilibrium. *Q.E.D.*

There still remains, however, how two curves can be tangential each other for the same level of deposits. Both the credible deposit contract offer curve and the iso-deposit supply curve are fixed in the $\theta^D - \pi$ plane at a specific level of deposits.

Lemma 10. *The credible deposit contract offer curve shifts downward with an increase in loan rate ρ^L and associated decline in loan default threshold θ^L consistent with a loan market equilibrium described in Proposition 1.*

Proof. Multiply both sides of (47) by $(1-\mu)s/2\mu$ and take a derivative of the right hand side with respect to ρ^L including its effect on θ^L. The sign is equal to $-dB(\theta^D)/d\rho^L$.

$$\frac{dB(\theta^D)}{d\rho^L} = (1-H)l - h(\rho^L+\delta)\frac{l}{\theta^D}\frac{\partial\theta^L}{\partial\rho^L} + \left((1-\lambda)(1-\delta)k_0 + \frac{\theta^L}{\theta^{D2}}e^U(k_0+l)^\alpha\frac{\partial\theta^L}{\partial\rho^L} - \tau\right)H, \tag{50}$$

where H and h are evaluated at θ^L/θ^D and note that $\partial\theta^L/\partial\rho^L < 0$ (Proposition 1). Rearranging this,

$$\begin{aligned}\frac{dB(\theta^D)}{d\rho^L} &= (1-H)l - H\tau \\ &\quad + H\left(\frac{\theta^L}{\theta^{D2}}\frac{\partial\theta^L}{\partial\rho^L}e^U(k_0+l)^\alpha - \frac{h}{\theta^D H}\frac{\partial\theta^L}{\partial\rho^L}(\rho^L+\delta)l + (1-\lambda)(1-\delta)k_0\right).\end{aligned} \tag{51}$$

The fist line is positive as long as τ is small relative to the loan amount l. By comparing the second line to bankers' loan contract offer (i.e., default decision by borrowers) (12), if the coefficient of the output in the first term in the bracket is larger than or equal to the coefficient of the loan rate with depreciation in the second term, whole second line is positive or zero. This condition is

$$\frac{\theta^L}{\theta^{D2}}\frac{\partial\theta^L}{\partial\rho^L} \geq \frac{lh}{\theta^D H}\frac{\partial\theta^L}{\partial\rho^L}, \tag{52}$$

equivalently,

$$\frac{\theta^L}{\theta^D} \geq \frac{h}{H}, \tag{53}$$

that is,

$$\frac{\theta^L}{\theta^D} \geq \frac{h\theta^L/\theta^D}{H}. \tag{54}$$

This condition is satisfied if the elasticity of the cdf of idiosyncratic shock at $\theta^L/\theta^D > 1$ is less than one (because $\theta^L/\theta^D > 1$). It is indeed so by Assumption 4.

Therefore, $-dB(\theta^D)/d\rho^L < 0$. *Q.E.D.*

Proposition 3. *In an equilibrium, a unique set of loan rate and associated loan default threshold (ρ^L, θ^L) supports a unique set of deposit rate and associated deposit default threshold (ρ^D, θ^D) for some deposit amount $\overline{s}$.*

Proof. Because the change in loan rate ρ^L does not bring any effects on the iso-deposit supply curve, Lemma 10 implies that, by appropriately choosing the loan rate and associated loan default threshold (ρ^L, θ^L), there is a unique set of deposit rate and associated deposit default threshold (ρ^D, θ^D), at which point the iso-deposit supply curve and the credible deposit contract curve are tangential each other for some deposit amount $\overline{s}$. *Q.E.D.*

Corollary 1. *In the deposit market partial equilibrium, the deposit amount is increasing in the loan rate ρ^L that supports an equilibrium deposit contract.*

Proof. Closer the iso-deposit supply curves are to the origin of the θ^D-π plane, higher the deposit rate and lower the deposit default threshold, implying higher the deposit supply. Then, Lemma 10 implies that the higher loan rate is needed to match with the higher deposit supply by shifting down the credible deposit contract curve. *Q.E.D.*

E. General Equilibrium

Proposition 4. *Given a banker population μ, there exists unique equilibrium loan rate ρ^L, deposit amount s, and loan amount l. Moreover, these pin down also unique equilibrium loan default threshold θ^L, deposit rate ρ^D (and spread π), and deposit default threshold θ^D.*

Proof. In the deposit market partial equilibrium, the deposit amount s is increasing with loan rate ρ^L (Corollary 1), while in the loan market partial equilibrium the loan amount l is decreasing with loan rate ρ^L (Proposition 1). These "deposit" curve and "loan" curve can be drawn in the ρ^L-(s, l) plane.

Because in the general equilibrium, the resource constraint needs to be met:

$$l^* = s^* + \frac{2\mu}{1-\mu} k_0^B. \tag{55}$$

Shifting up the "deposit" curve by $2\mu/(1-\mu)k_0^B$, the unique general equilibrium loan rate and loan amount (ρ^{l*}, l^*) is found as the cross point of the shifted-up deposit curve and the loan curve. And, the equilibrium deposit amount is derived by (55).

Moreover, given (ρ^{l*}, l^*), the default condition (i.e., loan supply function) pins down the equilibrium loan default threshold θ^{L*}. Given these equilibrium loan market variables and given the equilibrium deposit amount, the cross point of the iso-deposit demand curve and the credible deposit contract curve provides the general equilibrium deposit rate and default threshold (ρ^{D*}, θ^{D*}) as well as associated spread π^*. *Q.E.D.*

Proposition 5. *The banker population μ is determined uniquely in the decentralized equilibrium.*

Proof. (TBD: Sketch) First, the entrepreneur value V^E is decreasing with higher banker population μ. In the ρ^L-(s,l) plane, higher banker population μ implies lower loan rate ρ^L, larger loan amount l per borrower, and smaller deposit amount s per depositor. In the deposit market, in the θ^D-π plane, lower loan rate shifts up the credible deposit contract offer curve and smaller deposit amount also shifts up the iso-deposit supply curve. This means that the new spread π is higher (deposit rate ρ^D is lower) and the deposit default threshold θ^D is also higher. In the loan market, in the θ^L-ρ^L plane, higher loan means lower loan rate and ambiguous effects on the loan default threshold θ^L. Overall, the borrower gains and the depositor loses. From the expected utility point of view, however, choosing entrepreneur as a occupation becomes more riskier and thus less value. Second, the banker value V^B is also decreasing with an increase of banker population. A banker's income is essentially a spread income times relative size, $\pi * s * (1-\mu)/\mu$. If almost all choose to be banker $\mu \approx 1$, then only a tiny portion of people produce and exchange capital, and therefore $V^B \approx 0$. While in the other extreme $\mu \approx 0$, as long as some positive spread is paid, a banker's income become very high due to high level of capital exchange and more importantly due to almost infinite leverage (i.e., relative size).

Both V^E and V^B are decreasing but the slope of the latter is steeper. At $\mu \approx 1$, banker's utility $V^B \approx 0$ as shown above. But, the entrepreneur is essentially at autarkic level, thus $V^E > 0$. At $\mu \approx 0$, banker's utility $V^B \to \infty$ while the entrepreneur's is bounded by the finite value achieved in the Walrasian equilibrium allocation. *Q.E.D.*

V. SOCIALLY OPTIMAL ALLOCATION

A. Welfare Theorem

The optimal allocation is characterized as follows.

Definition 2. *The constrained social optimal allocation is the solution to the social planner's problem in which the social planner faces the same restrictions as the private agents given in Assumptions 1 to 4. Specifically, the social planner maximizes ex ante utility of a banker $V^B(k_0)$ subject to the occupational arbitrage condition (11).*[11] *This maximization is carried out by optimally choosing the capital allocation, l^o and s^o, the consumption allocation represented by the loan repayment schedule $R^{Lo}(l, A, \epsilon)$ and the deposit repayment schedule $R^{Do}(s, A)$, and the the number of banks, μ^o.*

Proposition 6. *The decentralized equilibrium achieves the constrained social optimum. That is, it achieves the constrained social optimal allocation given banker's population μ, which then is determined optimally.*

The proof is straightforward and omitted. There is no externality to break the link between the decentralized equilibrium and the social optimum in the model. First, there is no

[11] Maximizing V^E with a condition $V^E = V^B$ will yield the same allocation.

technological externality that affect Lemma 9. Second, because the occupational arbitrage equates the banker's and entrepreneur's utility and everyone is fully employed, there is no externality associated with the number of bankers in Proposition 5.

The two factors of the loan contract, that is, threshold θ^L and loan rate ρ^L, are determined almost mechanically by technological factors such as the verification cost and limited liability. As such, there is no difference between the monopolist banker's solution and competitive bankers' solution on relative allocation of two factors (and thus implicit relative price). This carries over to the social planner's solution. If I introduce price in the decentralized economy a la Prescott and Townsend (1984), the price of the equilibrium loan contract is indeed "one"— there is no surcharge providing the loan contract and no rents are earned by banks.

As for the deposit contract, the relative weight of risk θ^D and return ρ^D is uniquely determined at the tangential point of the deposit supply and demand. This carries over to the social planner's problem.

The occupational arbitrage constraint implies that bank's spread income plus the rents from selling deposit contracts are equated with the income of entrepreneurs. The extra rents to sell deposit contracts are uniquely determined at zero. And, there is no distinction between social and private solutions.

B. Walrasian Equilibrium as the Limit

Throughout the paper, I consider the general cases in which the verification cost is strictly positive and the limited liability constraint is binding. However, for a reference, in this section I analyze a "complete" market case as the limit of the model. Here, the verification cost is assumed to be almost zero and there is no limited liability. Therefore, the loan and deposit contracts take a form of equity contracts as the limit of debt contracts. This result follows in spirit that of Townsend (1978) on costly bilateral exchange.

Proposition 7. *If there is almost no financial frictions, that is, the verification cost is close to zero $\tau \approx 0$ and the retained assets after default is zero $\lambda = 0$, then deposit and loan contracts can take a form of equity (complete contract). The equilibrium with complete deposit and loan contracts mimics the Warlasian competitive equilibrium and is the first best. Specifically, the equilibrium deposit and loan rates are the same in the limit and so do the deposit and loan amounts. A small number of (i.e., measure zero) banks intermediate the capital. The optimal capital ratio of banks are (almost) zero. As a result, consumption is almost perfectly shared equally among all households.*

Proof. (Sketch). The positive spread between deposit and loan rates is a friction to the production sector. The smaller the spread, the larger is the outputs. Thus, a smallest number of banks should intermediate capital from the social planner's point of view. In the limit, the deposit-loan rate spread becomes zero with the same deposit and loan size. With zero spread and compete contracts, the idiosyncratic shocks are shared perfectly among all agents. As Proposition 6 still applies, this allocation is also supported by a decentralized equilibrium. *Q.E.D.*

C. Optimal Bank Size and Capital Ratio

A remaining question is the optimal size of the banking sector in an economy with sizable financial frictions. A depositor faces shock-contingent income up to the threshold θ^D and then flat income ρ^D from full deposit repayment. With large enough variations in the aggregate shocks, there are sizable chances that deposits are not repaid in full. Then, the depositors would prefer the more insured contract that provides a same expected return with a lower deposit rate but also with a lower default risk. A banker also prefers to provide this insurance contract, implying that he will have strictly positive capital.

Proposition 8. *In the constrained social optimum and hence in the decentralized equilibrium, the banking sector size is sizable $\mu^* > 0$. This implies that the optimal capital ratio is strictly positive.*

Proof. Suppose that the optimal banking sector size μ is measure zero (as is the case with the optimal complete market in Proposition 7). Then, the spread $\hat{\pi}$ is (almost) zero. The loan and deposit repayment schedules are the same by arbitrage. Specifically, there is one threshold $\hat{\theta}^L = \hat{\theta}^D$ below which both borrowers default and bankers default. I show below that this deposit contract is not constrained Pareto optimal.

First, a depositor prefers the deposit contract which gives the same expected return with less volatile repayment. Here, less volatile means a lower full repayment (i.e., lower deposit rate $\underline{\rho}^D$) but with lower threshold and higher repayment in case of banker's default. This contract (partially) insures depositors' income for the combined idiosyncratic and aggregate productivity shocks and thus it is preferred by a risk averse depositor.

Second, the partially insured contracts can be indeed designed so that the overall repayment has the same expected values. A banker can make such a contract by limiting deposits and loans given his endowed capital. Essentially he uses his capital as a buffer to depositors, so that the default threshold θ^D will be lowered. Because the new contract is assumed to give the same expected repayments to depositors, the banker is offering a lower deposit rate $\underline{\rho}^D$ for the same loan rate $\hat{\rho}^L$. This change of the deposit rate is denoted by the increase in the spread $\triangle\pi$ from zero.

Third, I can show that a banker strictly prefers the new contract—the lower expected value with less volatile deposit repayments—given prevailing loan contracts $R^{L*}(l, A, \epsilon)$ and the amount of deposit $\hat{s}$ per depositor. Consider drawing the original and new deposit repayment schedules with the aggregate shock realization on the x-axis and the repayment on the y-axis. Let $\underline{\theta}^D$ denote the new threshold of default by the banker.

(i) The depositor's gain is strictly larger than the the parallelogram made by the original and new contracts below the new threshold of the default. At the worst case at the origin, $\epsilon A = 0$, repayment is the seized asset adjusted for relative size $w(1-\lambda)(1-\delta)k_0^B$. Under the measure zero banking sector, this is equal to zero. So, the depositor's gain at the origin is $\hat{w}(1-\lambda)(1-\delta)k_0^B$, where $\hat{w}$ is the relative size of banking sector under the new contract. This is the amount that shifts up linearly the recovery rate of deposit contract in the default region. The new default region is from zero to $M(\underline{\theta}^D)$. Overall gains include the triangle area stemming from the difference of default thresholds under the original and new contracts. The parallelogram excluding such triangle area is smaller than the overall gain. Therefore,

$$Gain > \underline{Gain} = \hat{w}(1-\lambda)(1-\delta)k_0^B M(\underline{\theta}^D). \tag{56}$$

(ii) On the other hand, the depositor's loss is strictly lower than the upperbound of the loss, which is measured by the rectangle made by the change in the spread and the cumulative probability above the new threshold. That is,

$$Loss < \overline{Loss} = \triangle\pi(1 - M(\underline{\theta}^D)). \tag{57}$$

(iii) Because the new contract is supposed to have the same expected returns for a depositor, the gain and the loss must be the same. Therefore, the lowerbound of the gain must be strictly lower than the upperbound of the loss.

$$\triangle\pi > \hat{w}(1-\lambda)(1-\delta)k_0^B \frac{H(\underline{\theta}^D)}{1 - H(\underline{\theta}^D)}. \tag{58}$$

(iv) Because the current banking sector size is almost zero, I evaluate this at the limit $\hat{\mu} \to 0$ to see the profit of stemming from a new contract only slightly different from the current one. The limit value is positive:

$$\lim_{\hat{\mu} \to 0} \triangle\pi > 0. \tag{59}$$

(v) This implies that the increase in the spread by introducing the new contract is strictly positive—the spread is literary an insurance premium. With the new contract, a banker will have a higher income per deposit in case not defaulting in addition to a lower default probability.

(vi) By limiting loans and deposits, the total income could become less because the total income is affected by the spread times deposits. But, recall that the spread under the original contract is zero. Thus, the total income also increases from zero with the new contract, which can be only slightly different from the original contract.

In summary, both a depositor and a banker prefer a new contract, given the same loan contract (i.e., the same utility for borrowers). Therefore, the optimal capital ratio must be positive in the constrained social optimum. *Q.E.D.*

VI. Policy Implications

A. Bank Bailouts

I consider an expected bailout of bankers by a government. Specifically, a bailout policy is defined as guaranteeing a banker's income in case that a banker would default without the bailout policy in order to enable a banker to repay deposits in full. The government finances the transfer to a banker by taxing everyone ex post. This description represents the actual bailouts (see e.g., Landier and Ueda, 2009).

Assumption 5. *A government can collect tax from those who defaulted.*

It is legally and politically difficult to tax those who defaulted. However, in reality, the bailout funds are financed by government bonds, which the government repays over time, for example, by inflation tax on monetary assets or income tax on human capital (though not modeled here explicitly). In any case, those who defaulted will end up contributing the bailout expenditure in reality.

Definition 3. *A "transparent" bailout transfers funds to depositors via bankers without benefitting bankers, while an "untransparent" bailout benefits bankers directly.*

A transparent bailout is a good insurance for depositors at the cost of borrowers. Depositors can have the perfectly constant deposit repayment with the bailout policy for any realizations of the aggregate shocks but they also need to pay tax $\kappa(A)$ ex post to finance the bailout policy for negative aggregate shocks. The net-of-tax deposit repayments are not perfectly constant:

$$R^D_{BO}(s, A) = (\rho^D + \delta)s - \kappa(A), \quad \text{for } \forall A \in [\underline{A}, \overline{A}]. \tag{60}$$

A borrower needs to pay tax, too. His consumption is changed to

$$\epsilon A e^H (k_0 + l)^\alpha - R^L(l, A, \epsilon) - \kappa(A). \tag{61}$$

Under a transparent bailout, when a banker faces default, the government transfers funds to a banker just to repay the deposit in full so that the banker's pre-tax consumption schedule $c^B(A)$ is the same as before, described in equations (21) to (23).

Transfer occurs only when the aggregate shock is lower than the banker's default threshold, $\theta^D < A$. Per entrepreneur transfer is the difference between the banker's retained asset (21) and what a banker would consume without retained asset (23):

$$\kappa(A) = \frac{2\mu}{1-\mu}\left((\rho^D + \delta)\frac{1-\mu}{2\mu}s - B(A) - (1-\lambda)(1-\delta)k_0^B\right), \quad \text{if} \quad A \in [\underline{A}, \theta^D_{BO}]. \tag{62}$$

Otherwise, $\kappa(A) = 0$.

Under an untransparent bailout policy, the banker's consumption also increases by $\underline{\kappa}$ for the realization of aggregate shock lower than the banker's default threshold plus this additional transfer, $\theta_{BO}^D + \underline{\kappa} < A$. The transfer is simply shifted upwards by the additional transfer,

$$\kappa(A) = \frac{2\mu}{1-\mu}\left((\rho^D + \delta)D - B(A) - (1-\lambda)(1-\delta)k_0^B + \underline{\kappa}\right), \quad \text{if} \quad A \in [\underline{A} + \underline{\kappa}, \theta_{BO}^D]. \tag{63}$$

Otherwise, $\kappa(A) = \underline{\kappa}$.

Proposition 9. *A transparent bank bailout is welfare improving.*

Proof. Suppose the repayment function R^D and R^L were not reoptimized. With a transparent bailout policy, bankers do not gain or lose. The depositors and borrowers would share the cost of bank defaults while without it the borrowers would not share. The depositors would be better off but the borrowers would be worse off. However, from ex ante point of view, an entrepreneur could reduce the expected consumption volatility stemming from uncertainty for talent shocks that makes an entrepreneur either a borrower or a depositor. Therefore, better sharing the risk between a depositor and a borrower for a large negative aggregate shock is welfare improving for an entrepreneur before knowing his talent shock. By the occupational arbitrage, the banker's utility must be higher, too. In addition to this gain, there would be a gain from reoptimized repayment functions. *Q.E.D.*

The debt contract with limited liability implies that the borrowers and bankers are perfectly insured for a very low realization of aggregate shock while the depositors face the consequences. If there is a way to redistribute the borrowers' retained assets to the depositors, the ex ante overall welfare improves. Given the limited liability laws, one of a few is to use the tax system. If there is no limited liability and there is nothing left to the borrowers, then the bailout policy does not do anything. All tax collection is done from depositors to pay themselves because firms and bankers do not possess anything when bankers default.[12] But, if there is some still left in the hands of borrowers and banker who defaulted, depositors (and everyone in ex ante) would be better off to tax those assets.

B. Optimal Tax-Transfer System

If a government can tax on the relative loan repayment for each borrower at rate conditional on even idiosyncratic shocks and vary tax rates between depositors, borrowers, and bankers, the economy could move toward even better allocation by effectively nullifying the limited liability constraint in the social planner's problem.

In this optimal allocation under nullified limited liability, a borrower internalizes the expected tax and the loan amount become smaller. The banker's demand for deposits lowers and credible demand contract offer curve tilts flatter before the general equilibrium

[12] In this case, however, taxing depositors and then transfer some to bankers and firms would improve the overall welfare from ex ante point of view.

consideration—the effect of lower s^* in (47) is a higher intercept and a flatter slope, asking for a higher spread with a lower threshold. There will also be a general equilibrium effect. Due to a higher entrepreneur's utility, a banker population μ would be smaller than in the economy without any bailout policy and even than in the economy with bailouts under conventional tax $\kappa(A)$. With this general equilibrium effect, the banker's credible demand contract offer curve $s^d(\theta^D, \pi)$ shifts downward than in the case of the conventional tax. The equilibrium spread is now smaller and the default threshold should be lower than in the case of the conventional tax $\kappa(A)$.

If the government raises (consumption) tax from bailed-out bankers, the allocation become more similar to the first best. However, tax-transfer system based on idiosyncratic shocks are usually difficult to implement.

Under the conventional tax policy $\kappa(A)$, which is conditional only on the aggregate shocks, there can be several ways to correct the number of bankers and associated capital ratio. I discuss pros and cons of several policy implications below.

C. Deposit Insurance and Double Liability

A deposit insurance scheme can achieve the similar welfare improvement by a transparent bailout. A deposit insurance can be defined as a protection for depositors' income in case that a banker would default. The government is assumed to fully finance the transfer by taxing the bankers, ex ante. This "taxing bankers ex ante" is the difference from the bailout policy which is "taxing everyone ex post." I further assume that if the collected insurance premiums are not used, the funds will be paid back to bankers, so that the insurance fees are determined ex post and funded by tax.

I assume that the depositors will not lose the face value of the deposit—that is, a full coverage deposit insurance. Also, consider an unconventional system that are funded ex post.

Proposition 10. *The full coverage deposit insurance with ex post tax funding could create more bank defaults but improves the overall welfare as much as transparent bailouts.*

Proof. (Sketch) The same as a transparent bailout with also collecting tax from bankers. The depositors are better off because of the deposit insurance. Both borrowers and bankers bear costs. The iso-deposit supply shifts outward as before on the θ^D-π plane. The credible deposit contract offer shifts upward and flattens to ask higher spread to compensate the loss (with uncertain effect on threshold). Because the bankers are worse off than in the economy without such deposit insurance, less people becomes bankers. This general equilibrium effects (i.e., higher leverage) partially correct the partial equilibrium movements just described. *Q.E.D.*

If the fees are collected ex ante as usual, however, the welfare deteriorates.

Corollary 2. *The full coverage deposit insurance with ex ante fees does not improve welfare.*

Proof. The full coverage deposit insurance is essentially the same as restricting the bankers' offer of deposit contracts to be very safe, zero threshold $\theta^D = 0$, associated with very low deposit rate (or high spread π) to pay the insurance fee. This restriction on the bankers' offers of deposit contracts is an obvious distortion to the economy. *Q.E.D.*

Note that a partial coverage deposit insurance—cover the full amount down to the government-set threshold θ_G^D—with ex ante fees would create the similar distortion as the full coverage version. Essentially, the bankers are constrained to choose the deposit repayment schedule and thus the welfare decreases.

Note that the unlimited liability or "double liability" of bankers as in the pre-Great Depression in the U.S. would not work.[13] If bankers always have to pay deposit in full (unless their consumption becomes zero), then bankers are the ones that assume all the tail risks. This is not the optimal risk sharing among different types of agents and thus is not the optimal. The key friction is not the "limited" liability itself but rather the limited liability being noncontingent to the aggregate shocks. A transparent bailouts can fine tune the limited liability, making it contingent to the aggregate shocks.

D. Fiscal and Monetary Operations

A transparent bailout can be mimicked by fiscal and monetary operations. Recall that a transparent bailout is such that depositors receive transfer $\kappa(A)$, which is financed by $\kappa(A)$ tax on borrowers. This $\kappa(A)$ tax revenue is given to bankers and then bankers need to use the same sum to repay deposits. This is "net" tax-transfer system. In gross terms, the transfer to a depositor is $2\kappa(A)$ and $\kappa(A)$ tax is levied on everyone. In this case, a bank also pays the tax but receive the same amount, canceling out each other. Still, the tax revenue from borrowers are injected to banks and used to repay deposits in full as before.

These operations do not need to be done simultaneously. A government instead can use bonds to defer the timing of the transfer and tax revenue, as is often the case. The bonds amount of $2\kappa(A)$ is provided for free to bankers, who use them to repay deposits in full. Then, later, at the consumption stage, a government collects tax from everyone at rate $\kappa(A)$ to repay bonds. This is a simple fiscal operation that achieve the transparent bailout.

In a monetary economy, monetary policy can also be used. In this case, all contracts are assumed to be nominal. In the end of the period, instead of levying tax, inflation can be created so that walked-away borrowers and bankers have less purchasing power out of retained assets. This inflation tax should be set equal to the consumption tax case above (i.e., $\kappa(A)$). Also, instead of real bonds, money can be injected for free to bankers. Amount of money should be adjusted for future inflation so that real value is the same as the scheme using the real bonds.

[13] See a history and theory paper by Kane and Wilson (1998) and an empirical work by Grossman (2001).

Different timing of inflation might work. If inflation is created after production but before repayment, real debt obligations become smaller and reservation utility from retained assets become smaller, so that borrowers and bankers would repay in full nominally even in crisis. The real repayments to depositors become also small so that depositors remain vulnerable as in the case without any policy actions. Still, the transparent bailout outcome can be achieved if the government compensate depositors directly using seignorage. However, the direct transfer to depositors would face difficulties to implement.

Although fiscal and monetary operations can have the same implications, the implementation speed may be different in the real world. If a central bank is independent from politics, it can implement this efficient bailout quickly. This is not the case with fiscal operations. On the other hand, without political scrutiny, the bailout might involve more of the inefficient transfers.

E. Macro Prudential Regulations

In the current regulatory environment, the capital adequacy ratio q is subject to the minimum regulatory requirement $\hat{q}$. This can be expressed as

$$q \equiv \frac{k_0^B}{D} \geq \hat{q}. \tag{64}$$

Proposition 11. *Introduction of the capital adequacy ratio regulation as defined in (64) is either redundant or welfare decreasing in an economy without a untransparent bailout.*

Proof. Proposition 8 says the optimal capital ratio is positive in an economy with debt contracts. By Proposition 5, bankers hold strictly positive capital by themselves in an equilibrium which is the constrained social optimum. Therefore, the capital adequacy ratio requirement is either binding (i.e., welfare decreasing) or not binding (i.e., redundant). *Q.E.D.*

Note that there are two sources of inefficiency. First, with the capital ratio requirement, there will be more bankers with less customer base and a higher spread to compensate less customer base. Second, with a sizable positive spread and resulting wedge between the loan and deposit rates, the marginal product of capital would be less equated between the borrowers and depositors. When $\tau = 0$ (complete market case), with the capital adequacy ratio requirement, the economy cannot reach the limit that mimics the first best, Walrasian equilibrium.

Corollary 3. *When the capital adequacy ratio requirement is introduced to the economy with the bailout policy with the conventional tax system $\kappa(A)$, there will be more banks with a higher capital ratio and a lower spread but also with a higher probability of bank bailouts. The overall welfare becomes worse.*

Proof. A banker needs a higher spread to satisfy the same default threshold due to the lower leverage. The banker's deposit demand $s^d(\theta^D, \pi)$ shifts upward and flattens as a result. This implies that a higher spread π and uncertain movement on threshold θ^D. Moreover, the capital adequacy ratio requirement, if binds, makes bankers lower utility. *Q.E.D.*

Remark. *If too large transfer $\underline{\kappa}$ is already present, introducing the capital adequacy ratio can mitigate a unnecessarily high incentive to become a banker by lowering bankers' utility. Introducing a bank levy to lower the (present value of) transfer works as well.*

Basel III now includes the liquidity ratio regulation. It can be considered as ex ante fiscal operations as long as liquid assets are defined as government bonds. In the beginning of the period, a government gives away certain amounts of bonds to banks and ask them to hold them. If the productivity shocks are good enough, no bailouts are necessary. In this case, returns from borrowers are enough to repay deposits. Unused bonds are asked to be returned to the government for free. If shocks are very low, banks use bonds to repay deposits in full. The government repays bonds by tax from everyone. This is the same as the efficient bailout scheme.

In a more realistic case, a government does not give away bonds but ask banks to buy them. In this case, a government raises some revenues (real goods) and places them in storage. In the end, if there is no need to bail out, the government repays bonds to bankers using stored goods. If there is a need for bailout, banks use bonds to repay deposits in full. The government repays bonds to depositors using stored goods. This scheme is also similar to the one above except that only banks pay tax in the form of forced purchase of government bonds. By doing this, bankers' utility would be lower and then smaller number of agents want to become bankers, implying higher loan to capital ratio per banker than the optimal. The capital account ratio regulation can be used to mitigate this distortion to make artificial monopoly rents for banks lowering utility levels of entrepreneurs. However, this scheme seems too complex to implement and to be evaluated if this mimics the efficient bailout.

In a monetary economy, the above scheme could be improved. Now the government sells nominal bonds to banks. In case of good shocks, the government creates deflation and repays banks more than it raised. Depositors also gain and borrowers lose. In case of bad shocks, the government creates inflation to compensate the tax that banks already paid. In case of bad shocks,

In any case, CAR and liquidity ratio regulation should be relaxed when economy receives bad shocks and capital and liquidity are used to mitigate the crisis impact.

VII. Conclusion

I study optimal banking sector policy design against exogenous crisis probability based on a general-equilibrium framework with realistic financial frictions, namely, costly state verification, limited liability, and simple renegotiation. Moreover, I assume endogenous banking sector—a banker is an occupation and banker's income is equated with entrepreneur's in expectation. Entrepreneurs are further sorted to either borrowers or depositors depending on the talent shocks they draw. This occupational choice—the labor allocation—is a possible place where a policy may distort. Banker's income stems from the spread between deposit and loan rates. The loans and deposits—the capital allocation—are affected by the spreads, which a policy may distort, too.

The optimal loan and deposit contracts take a form of a standard debt contract because of costly state verification in bank lending. The optimal bank capital is positive to provide a buffer to depositors and bankers themselves. And, the banking sector is sizable. However, when a large negative shock hits, both borrowers and bankers would walk away with retained assets because of limited liability protection. The depositors would assume all the tail risk. This *tail-risk dumping* problem creates the occupational choice too risky.

A government-led bailout of banks, if transparent, can improve welfare as it acts as an insurance scheme. As it makes the limited liability constraint to be contingent on the aggregate shocks, a bailout will mitigate the *tail-risk dumping* problem. The deposit insurance, if funded ex post by tax, can mimic such a transparent bailout. Some form of liquidity ratio regulation works too.

However, note that, a bailout is welfare improving only if banks and borrowers are too much protected by limited liability. Moreover, if a bailout is not transparent and directly beneficial to banks, it distorts the factor allocation. This leads to too large a financial sector with too small outputs (*income shifting*). Capital adequacy ratio requirement or a bank levy has a role to play in this case.

References

Akerlof, George A., and Paul M. Romer, 1993, "Looting: The Economic Underworld of Bankruptcy for Profit," *Brooking Papers on Economic Activity*, Vol. 2, pp. 1–73.

Allen, Franklin, and Douglas Gale, 2000, "Financial Contagion," *Journal of Political Economy*, Vol. 108, pp. 1–33.

______, and ______, 2007, *Understanding Financial Crises* (Oxford: Oxford University Press).

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, "The Financial Accelerator in a Quantitative Business Cycle Framework," 1999, in John B. Taylor, and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1C (Amsterdam: Elsevier), chapter 21, pp. 1341–1393.

Bolton, Patrick, Tano Santos, and Jose A. Scheinkman, 2011, "Creme Skimming in Financial Markets," *NBER Working Paper*, No. 16804.

Boyd, John H., and Gianni De Nicolo, 2005, "The Theory of Bank Risk Taking and Competition Revisited," *Journal of Finance*, Vol. 60, No. 3 (June), pp. 1329–1343.

Chari, V. V., and Patrick J. Kehoe, 2009, "Baiouts, Time Inconsistency and Optimal Regulation," *Federal Reserve Bank of Minneapolis Research Department Staff Report.*

Chrsitiano, Lawrence, Roberto Motto, and Massimo Rostagno, 2010, "Financial Factors in Economic Fluctuations," *ECB Working Paper*, No. 1192.

Fulghieri, Paolo, and Eitan Goldman, "The Design of Debt Contracts," 2008, in Anjan V. Thakor, and Arnoud W.A. Boot, eds., *Handbook of Corporate Finance, Vol 3: Financial Intermediation and Banking*, (North-Holland).

Gale, Douglas, and Martin F. Hellwig, 1985, "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, Vol. 52, No. 4 (October), pp. 647–663.

Grossman, Richard S., 2001, "Double Liability and Bank Risk Taking," *Journal of Money, Credit and Banking*, Vol. 33, No. 2, pp. 143–159.

Igan, Deniz, Prachi Mishra, and Thierry Tressel, "A Fistful of Dollars: Lobbying and the Financial Crisis," 2011, in Daron Acemoglu, and Michael Woodford, eds., *NBER Macroeconomic Annual*, (MIT Press). forthcoming.

Kane, Edward J., and Berry K. Wilson, 1998, "A Contracting-Theory Interpretation of the Origins of Federal Deposit Insurance," *Journal of Money, Credit and Banking*, Vol. 30, No. 3, pp. 573–595.

Kareken, John H., and Neil Wallace, 1978, "Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition," *Journal of Business*, Vol. 51, No. 3, pp. 413–438.

Keeley, Michael C., 1990, "Deposit Insurance, Risk, and Market Power in Banking," *American Economic Review*, Vol. 80, No. 5 (December), pp. 1183–1200.

Kiyotaki, Nobuhiro, and John Moore, 1997, "Credit Cycles," *Journal of Political Economy*, Vol. 105, No. 2, pp. 211–248.

Kocherlakota, Narayana, 2010, "Taxing Risk and the Optimal Regulation of FInancial Institutions," *FRB Minneapolis Economic Policy Paper*, Vol. 10, No. 3.

Landier, Augustin, and Kenichi Ueda, 2009, "The Economics of Bank Restructuring: Understanding the Options," *IMF Staff Position Note 09/12*.

Philippon, Thomas, 2008, "The Evolution of the US Financial Industry from 1860 to 2007: Theory and Evidence." NYU: mimeo.

Reinhart, Carmen M., and Kenneth S. Rogoff, 2009, *This Time is Different* (Princeton: Princeton University Press).

Townsend, Robert M., 1978, "Intermediation with Costly Bilateral Exchange," *Review of Economic Studies*, Vol. 45, No. 3 (October), pp. 417–425.

______, 1979, "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory,* Vol. 21, No. 2 (October), pp. 265–293.

Ueda, Kenichi, 2006, "Banks as Coordinators of Economic Growth," *IMF Working Paper*, No. 06/264.

Van den Heuvel, Skander J., 2008, "The Welfare Cost of Bank Capital Requirements," *Journal of Monetary Economics*, Vol. 55 (December), pp. 298–320.