MONASH University
Business and Economics

DEPARTMENT OF ECONOMICS

ISSN 1441-5429

DISCUSSION PAPER 14/11

The Financial and Macroeconomic Implications of Banking Frictions and Banking Riskiness

Yi Jin[*] and Zhixiong Zeng[†]

Abstract:

This paper develops a model of banking frictions and banking riskiness, the importance of which is highlighted by the recent Global Financial Crisis (GFC). We propose a model-based approach to decompose the effect of a banking riskiness shock into a pure default effect and a risk effect when risk sharing among the depositors is imperfect. Although the default effect is quantitatively more important, the risk effect is not to be neglected. When the shock generates a bank spread similar in value to the peak during the GFC, the overall effect is a decline in employment by 6:57 percent. The pure default effect leads to a 4:76 percent employment decline by a "within-model" measure, and a 5:05 decline by a "between-model" measure. The remaining is attributed to the risk effect.

JEL Classifcation: E44, E32, D82, D86.
Keywords: Banking riskiness shocks; two-sided debt contract; default effects; risk effects; financial crisis.

[*] Corresponding author. Department of Economics, Monash University, Caulfield East, VIC 3145, Australia. Phone: +61-3-99032042, Fax: +61-3-99031128, Email: yi.jin@monash.edu
Homepage: http://users.monash.edu.au/~yij/.
[†] Department of Economics, Monash University, Caulfield East, VIC 3145, Australia.
Phone: +61-3-99034045, Fax: +61-3-99031128, Email: zhixiong.zeng@monash.edu

1 Introduction

Banking is risky business, and the bankruptcy of banks is a real possibility. When banks fail they default on at least part of their liabilities. Although there has been deposit insurance and the insurance coverage was raised from $100,000 to $250,000 in October 2008 in the United States, a large amount of bank liabilities remain uninsured. For example, large-denomination certificates of deposits (CDs) are normally issued in $1 million pieces, well exceeding any deposit insurance limit. It is the interest rates on these uninsured liabilities that determine the marginal cost of external finance for banks. To compensate for the possibility of default, banks' liability holders (depositors[1] henceforth for ease of exposition) require a premium on their funds over default-free securities, giving rise to interest spreads between, say, CDs, and Treasury bills (T-bills). The recent Global Financial Crisis (GFC) calls attention to the importance of banking risks and the frictions present in the bank-depositor relationship. From 2001Q1 to 2007Q2, the spread between 3-month CD rate and the 3-month T-bill rate was as low as 27 basis points per annum on average. For the second half of 2007 and the year of 2008 this spread rose to as high as 153 basis points per annum on average, with a spike at 252 basis points in the last quarter of 2008 (Figure 1a).[2] The rise in this spread partly reflected the rising likelihood of bankruptcy of banks. In 2005 and 2006, the number of failed FDIC insured financial institutions was simply zero. The number, in contrast, was 3 in 2007, 30 in 2008, and 148 in 2009 (Figure 1b).[3]

---

[1] It should be clarified that in this paper we use the term "deposits" in the broadest sense, referring to all liabilities of banks that are held by the private sector. Meanwhile, we lump all the private-sector creditors of banks, including consumers, nonfinancial businesses, and nonbank financial firms, into a single category of agents called "depositors".

[2] Data source: the Board of Governors of the Federal Reserve System.

[3] Source: FDIC, http://www2.fdic.gov/hsob/HSOBSummaryRpt.asp?BegYear=1934&EndYear=2010&State=1.

[Insert Figure 1 about here.]

In this paper we develop a model of banking frictions, where stochastic changes in the riskiness of banking affect the economy's employment and output. By banking frictions we mean the asymmetric information and agency problem on the liability side of the bank balance sheet, that is, between banks and their lenders (depositors). The literature on financial market imperfections has so far focused on what we call "credit frictions"— the agency problem on the asset side of the bank balance sheet, that is, between banks and their borrowers, eg., entrepreneurs. See the seminal work of Bernanke and Gertler (1989) and a large literature that follows[4]. To introduce banking frictions we extend the costly-state-verification (CSV) framework of Towsend (1979), Gale and Hellwig (1985), and Williamson (1986, 1987) to a two-sided financial contracting framework. In our model banks face idiosyncratic risks and depositors have to expend monitoring costs in order to verify banks' capacities to repay, just like banks themselves have to incur such costs in order to verify entrepreneurs' revenues. If the banks are subjected to risks that cannot be fully diversified, then the kind of agency problem between banks and entrepreneurs applies equally well to the relationship between banks and depositors. In that case there are needs to "monitor the monitor," in the terminology of Krasa and Villamil (1992a). In our environment the optimal financial contract is a two-sided debt contract, which features equilibrium default by both the entrepreneurs and banks. The overall financial frictions that are relevant for the determination of equilibrium employment and output are summarized by a financial friction indicator, which itself is a function of the entrepreneurs and banks' default thresholds as specified by the contract.

---

[4]Examples include Carlstrom and Fuerst (1997), Bernanke, Gertler and Gilchrist (1999), Fisher (1999), and Christiano, Motto and Rostagno (2005, 2009), etc.

We capture the extent of banking riskiness by a dispersion parameter in the distribution of banks' idiosyncratic risks and allow this parameter to be subjected to stochastic disturbances. This formulation parallels the formulation of entrepreneurial riskiness shocks in the earlier work of Williamson (1987) and the recent work of Christiano, Motto, and Rostagno (2003, 2009). These authors consider the CSV problem between banks and entrepreneurs but not the problem between banks and depositors. In the inspiring work of Williamson (1987), savers delegate monitoring of entrepreneurs to a large financial intermediary, which perfectly diversifies away all of the credit risks and is able to guarantee the depositors a risk-free return. There are thus agency costs between the entrepreneurs and the financial intermediary, but no such costs between the financial intermediary and the depositors. In that environment stochastic changes in the riskiness of the entrepreneurs' projects generate aggregate fluctuations, fluctuations that would not obtain were there no costly monitoring on the outcomes of the entrepreneurial projects. In his model, however, banking frictions (i.e., agency costs between financial intermediaries and depositors) are absent, leaving no role for stochastic changes in banking riskiness to play. This abstraction might be innocuous for episodes where banking frictions are not severe enough to deserve attention. Yet the experience of the GFC pinpoints the importance of banking frictions and stochastic changes in the riskiness of banking. As Figure 1a indicates, during the GFC there was a sharp increase in the spread between the bank CD rate and the T-bill rate, but not in the spread between the bank lending rate and the CD rate, suggesting that banking riskiness shocks might be more important than entrepreneurial riskiness shocks for the turmoil the financial and real sectors of the economy recently went through.

What are the financial and macroeconomic consequences of shocks to banking riskiness? The answer to this question hinges on the way the financial friction indicator—a sufficient statistic for the determination of employment—responds to the shocks. The way the indicator responds, in turn, depends on the prevailing risk sharing arrangements among the depositors. In our model two polar risk sharing models are considered. In the first model there is no risk sharing so that the depositors, endowed with a logarithmic von Neumann-Morgenstern utility function, remain risk averse with respect to banking risks (the log form simplifies the analysis by a great deal). In the second model there is perfect risk sharing so that the depositors are in effect risk neutral with respect to banking risks. An important point to notice is that with imperfect risk sharing the response of the financial friction indicator and hence employment includes both a pure default effect and a risk effect, while with perfect risk sharing only the pure default effect is present. The pure default effect stems from the fact that the optimal contract dictates equilibrium bankruptcy of banks, the rate of which depends on the extent of banking riskiness. This effect will be present as long as the banks' default rate changes with the shock, even when the depositors are risk neutral. The risk effect results from the fact that the payment streams under the optimal contract are uncertain for the depositors. This effect obtains when risk sharing among the depositors is imperfect so that they remain risk averse with respect to banking risks.

Taking the imperfect risk sharing model to be the "true" model (perfect risk sharing seems to be less realistic), we propose a model-based approach to decompose the overall effect of a banking riskiness shock into the pure default effect and the risk effect. Specifically, two measures are developed to assess these effects: a within-model measure and a

between-model measure. The within-model measure relies on using the equilibrium default thresholds prevailing in the "true" model, while the between-model measure takes into account the fact that with the same realization of banking riskiness shock the equilibrium default thresholds will vary from one model to the other. A quantitative exercise is carried out to evaluate the default effect versus the risk effect. We find that benchmarked to the mean level of riskiness, the pure default effect of a banking riskiness shock contributes about 73% to the overall effect according to the within measure, and about 77% according to the between measure. When the shock generates a bank spread similar in value to the peak during the GFC, the overall effect is a decline in employment by 6.57 percent. According to our decomposition, the pure default effect leads to a 4.76 percent employment decline by the within measure, and a 5.05 decline by the between measure. On the other hand, the risk effect produces a 1.81 percent employment decline by the within measure, and a 1.52 percent decline by the between measure. Although the pure default effects are quantitatively more important, the risk effects are not to be neglected.

The rest of the paper is organized as follows. Section 2 describes the model—the economic environment, the two-sided financial contracting framework, and the general equilibrium. Section 3 introduces banking riskiness shocks and proposes our model-based approach to decompose their financial and macroeconomic impacts into pure default effects and risk effects. Quantitative evaluations of these effects are also presented. The last section offers some concluding remarks.

2 The Model

2.1 The Environment

There are four types of agents in the economy—saver/depositors, bankers, entrepreneurs, and workers. Entrepreneurs own the production technologies and operate the firms. They need to hire labor from the workers but are short of funds in paying the wage bills if they do not borrow from the banks in advance. Banks, which are run by the bankers, in turn secure funds from the saver/depositors to finance their lending activities. The financial contracting problem is thus two-sided: Banks sign loan contracts with the firms and deposit contracts with the depositors.

To simplify the analysis, we a consider a two-period setup.[5] Production uses capital and labor and takes place only in period 1. We assume that each firm owns the same fixed amount of physical capital K^f, and that each bank owns the same fixed amount K^b. There is a competitive rental market with rental rate R^k. And the rental income of capital constitutes the firms and banks' internal funds.[6] Since the firms' internal funds are generated entirely from the current rental value of the capital stock they own, in a market clearing equilibrium the firms must borrow additional funds to finance their purchase of labor inputs supplied by the workers plus the rental services provided by

---

[5]An infinite-horizon version of the model is presented in Zeng (2010), who assumes perfect risk sharing among the depositors so that they are effectively risk-neutral with respect to banking risks. This assumption allows for the usage of a representative-household setup when characterizing the saving behavior. In the present paper we allow for the possibility of imperfect risk sharing. With imperfect risk sharing, agents receiving different shocks will end up with different levels of wealth. Here we choose to work with the two-period, rather than infinite-horizon, setup in order to avoid the difficulty of keeping track of the distribution of money balances across the risk-averse depositors, which would complicate the analysis without adding much more insight.

[6]As there is no production activity in period 2, the act of confiscating the borrowers' physical capital in the event of default is of no value to the lenders. Hence the changes in the price of capital are not potential sources of changes in the net worth of the borrowers in our model. Note that the fixity of capital stock does not prevent it from generating variable internal funds as the rental rate reacts to shocks to the economy.

the stock of physical capital owned by the banks. The paper thus emphasizes working capital financing as in Christiano and Eichenbaum (1992). Our model differs from theirs in that financial frictions are inflicted on the firms' purchases of labor inputs in the present setting, giving rise to a *financially distorted labor market.*

At the beginning of period 1, the agents are endowed with initial purchasing powers, i.e., money balances of given amounts. Let the money balance be M^d for each depositor, M^b for each bank, and M^w for each worker (for simplicity assume that the firms do not possess any initial money balances). The total amount of money balances is then $M \equiv M^d + M^b + M^w$. There is a risk-free government bond of zero supply. The gross interest rate on this bond, i.e., the risk-free rate, is pegged by the government and normalized to be one.

The funds in this economy circulate in the following way. First, the sum of M^d and M^b is channelled by the banks and goes to the firms to purchase labor L in the competitive labor market at wage rate W. In fact, the loan market clears when[7]

$$WL = M^b + M^d. \tag{1}$$

The sum then becomes labor income at the hands of the workers. The workers use WL plus their initial money balances M^w to purchase consumption goods C_1 at price level P_1. That is, $P_1 C_1 = WL + M^w$. Substituting (1) into this budget equation, we obtain the quantity equation:

$$P_1 C_1 = M. \tag{2}$$

---

[7]The loan market clearing condition takes the form (1) because the firms' rental payment on capital is covered by the rental value of the stock of capital owned by the firms and banks. It remains that their wage bills are to be ultimately financed by the after-transfer money balances of the banks and depositors. To write the loan market clearing condition in full, we have $R^k \left(K - K^f\right) + WL = \left(R^k K^b + M^b\right) + M^d$. This simplifies to (1) since $K = K^f + K^b$.

The sum of money M is received by the firms as revenues. It is then divided among the firms, banks, and depositors according to the financial contracts and carried over by these agents to period 2 to purchase consumption goods. We assume that there are output endowments in period 2 given by $C_2 > 0$. The period-2 price level is thus $P_2 = M/C_2$. At the end of period 2, all of the money stock M retires. Figure 2 illustrates the flow of funds in the model.

[Insert Figure 2 about here.]

The workers work and consume only in period 1. They have constant marginal rate of substitution between leisure and consumption, given by $\nu > 0$. Hence the real wage rate W/P_1 simply equals ν. We treat them as being risk neutral since they do not face any idiosyncratic uncertainty at all: They always receive the full payment of wages since the firms are required to deliver this payment before labor can be provided. In contrast, the depositors, bankers, and entrepreneurs, who for simplicity only consume in period 2, all face idiosyncratic uncertainty. We assume that the bankers and entrepreneurs are risk neutral, but the depositors are risk averse, with logarithmic von Neumann-Morgenstern utility function (this functional form simplifies the analysis by a great deal when risk sharing among the depositors is imperfect). In the financial relationships among these three parties, the banks face the possibility of default by the firms that borrow from them, and the depositors face the possibility of default by the banks where they made deposits.

Note that the distribution of purchasing powers in period 1 (the relative fractions of M^d and M^b in M) affects the terms of financial contracts negotiated, which in turn affects the quantity of labor input and output produced in that period. The terms of financial contracts also determine the division of firm revenues M among the contracting

8

parties and hence the distribution of period-2 purchasing powers (claims on period-2 consumption goods) among the depositors, bankers, and entrepreneurs. Hence to put it in a different way, the division of surplus (in the form of future consumption) as dictated by the financial contracts has non-trivial implications for current (period-1) employment and production. Before analyzing the financial contracting problem a detailed description of the production and information structure is necessary.

Production in period 1 takes place in an environment with a unit-mass continuum of regions indexed by i, $i \in [0,1]$. In region i there is one bank, called bank i, and a unit-mass continuum of firms indexed by ij, $j \in [0,1]$. Each firm resides in a distinct location and is owned by an entrepreneur, who operates a stochastic production technology that transforms labor and capital services into a homogeneous final output. The technology of firm ij is represented by the production function

$$y_{ij} = \theta_i \omega_{ij} F\left(k_{ij}, l_{ij}\right), \tag{3}$$

where y_{ij}, k_{ij}, and l_{ij} denote final output, capital input, and labor input, respectively, of firm ij. The function $F\left(\cdot\right)$ is linearly homogeneous, increasing and concave in its two arguments, and satisfies the usual Inada conditions. All sources of idiosyncratic risks are captured in the productivity factor, with θ_i being the random productivity specific to region i, and ω_{ij} the random productivity specific to location ij. We assume that θ_i is identical and independently distributed across regions, with c.d.f. $\Phi^r\left(\cdot\right)$ and p.d.f. $\phi^r\left(\cdot\right)$, and that ω_{ij} is identical and independently distributed across all locations, with c.d.f. $\Phi^l\left(\cdot\right)$ and p.d.f. $\phi^l\left(\cdot\right)$. Both θ_i and ω_{ij} have non-negative support and unit mean. Furthermore, θ_i and $\omega_{\tau j}$, $i, \tau, j \in [0,1]$, are uncorrelated with each other. The distributions are known by all agents in the economy. Once the firms acquire factor inputs, production

9

takes place, and the region and location specific productivities realize. The final output is sold in a competitive goods market.

We use the CSV approach of Towsend (1979), Gale and Hellwig (1985), and Williamson (1986) to model financial frictions and financial contracting. It is assumed that there is an informational asymmetry regarding borrowers' ex post revenues. In particular, only borrowers themselves can costlessly observe their realized revenues, while lenders have to expend a verification cost in order to observe the same object. In our environment only firm ij can observe at no cost $s_{ij}^f \equiv \theta_i \omega_{ij}$, and only bank i can observe θ_i costlessly. For a bank to observe s_{ij}^f (or ω_{ij}) and for a depositor to observe θ_i, verification costs have to be incurred. Note that by lending to a continuum of firms in a particular region each bank effectively diversifies away all the firm/location specific risks. But the region specific risk is not diversifiable, giving rise to the possibility that a bank becomes insolvent when an adverse regional shock occurs. Our model thus features potential bankruptcy of banks in addition to bankruptcy of nonfinancial firms. Note that even if the working capital loans are perfectly safe for the banks (no default by the firms), the depositors still regard their claims on the banks as being risky due to the informational asymmetry about the idiosyncratic bank/region productivities.

The concept of "regions" should not be interpreted literally as reflecting geographic areas, albeit this is certainly one of the many possible interpretations.[8] Rather, it is a device designed to generate risks idiosyncratic to individual banks. If banks are subjected to risks that cannot be fully diversified, then the kind of agency problem between banks and firms applies equally well to the relationship between banks and depositors. In that

---

[8]Bank-level risks might stem from geographic confinement of an individual bank's operation to specific areas, as in the U.S. when out-of-state branching was restricted (see Williamson, 1989).

case there are needs to "monitor the monitor," in the terminology of Krasa and Villamil (1992a). It should be noted that even without branching restrictions or regulations on banks' lending and investment activities, an individual bank might optimally choose to limit its scale and/or scope of operation so that the risks associated with its lending activities are not fully diversified. See, for example, Krasa and Villamil (1992b) and Cerasi and Daltung (2000). In this paper we follow Krasa and Villamil (1992a) and Zeng (2007) to assume that an individual bank cannot contract with a sufficient variety of borrowers so that the credit risks are not perfectly diversifiable.

2.2 Financial Contracting with Banking Risks

The Two-Sided Debt Contract

The three groups of players in the financial market—firms, banks, and depositors— are connected via a two-sided contract structure. Both sides of the contract, one between the firms and banks and the other between the banks and depositors—fit into a generic framework we now describe. Here attention is restricted to deterministic monitoring.[9] Since the borrowers (firms and banks) are assumed to be risk neutral, the optimal contract between a generic borrower and a generic lender takes the form of a standard debt contract, in Gale and Hellwig (1985)'s term.

As for the depositors' attitude toward bank risks, two models are considered in this paper. In both models each depositor contracts with only one bank, which might be due to some transaction costs that prevent a depositor from dividing her deposits among

---

[9]The assumption of deterministic monitoring is actually less restrictive than it appears. Krasa and Villamil (2000) articulates a costly enforcement model that justifies deterministic monitoring when commitment is limited and enforcement is costly and imperfect. See also Mookherjee and Png (1989) and Boyd and Smith (1994) on deterministic versus stochastic monitoring.

many banks in order to diversify her portfolio. The paper is silent on the exact nature and details of these costs, as our focus is on whether there are risk sharing arrangements that individual depositors can reply on to handle the consequence of not being able to achieve full diversification through splitting deposits.[10] In the first model, which we label "Model A," there is no risk-sharing among the depositors so that they remain risk averse with respect to bank risks. In the second model, "Model B," there is perfect risk sharing so that the depositors become effectively risk neutral with respect to bank risks, despite that each of them has a strictly concave von Neumann-Morgenstern utility function. The analysis in this paper mainly concerns the financial and macroeconomic impacts of banking riskiness shocks under different risk-sharing arrangements among the depositors. The two models considered here represent two opposite extremes, which makes the contrast as stark as possible.[11]

Suppose that the borrower's revenue is given by Vs, where V is a component freely observable to the lender, and $s \geq 0$ is a unit-mean risky component that is subject to informational asymmetry, whereby the borrower can costlessly observe s while the lender has to expend a verification cost in order to do so. The verification cost is assumed to be μ times the borrower's revenue, with $\mu \in (0, 1)$. The c.d.f. of s, given by $\Phi(\cdot)$, is common knowledge. The contract specifies a set of realizations of s for which monitoring occurs, together with a payment schedule. A standard debt contract with monitoring

---

[10]For a paper that explicitly models the costs of financial transactions and endogenizes lenders' asset indivisibility, see Zeng (2007).

[11]Note that with risk averse depositors standard debt contracts are optimal not only because they minimize the need for monitoring, but also because they provide optimal risk sharing between the bank and the depositor. Problems only arise if the borrowers are more risk averse, because it is then optimal for the lenders to reduce the borrowers' exposure to risk (see, for example, Hellwig, 2000). To avoid confusion, the risk sharing arrangements we talk about in this paper refer to those among the depositors with respect to bank risks.

threshold $\bar{s} > 0$ has the following features: (i) the monitoring set is $\{s|s < \bar{s}\}$, (ii) the fixed payment is $V\bar{s}$ for $s \in \{s|s \geq \bar{s}\}$, and (iii) the payment is Vs for $s \in \{s|s < \bar{s}\}$. Resembling many financial contracts in the real world, the debt contract allows for fixed payment for non-default states and state-contingent payment when default occurs.

Under the standard debt contract, the borrower and the lender each obtains a share of the expected revenue V. The borrower receives $V\Gamma(\bar{s}; \Phi)$ where

$$\Gamma(\bar{s}; \Phi) \equiv \int_{\bar{s}}^{\infty} (s - \bar{s}) \, d\Phi(s), \tag{4}$$

reflecting the fact that with s above $\bar{s}$, the borrower gives out the fixed payment $V\bar{s}$ and keeps the remaining, while with s below $\bar{s}$, all revenues are confiscated by the lender. The lender receives $V\Psi(\bar{s}; \Phi)$ where

$$\Psi(\bar{s}; \Phi) \equiv \bar{s}[1 - \Phi(\bar{s})] + (1 - \mu) \int_{0}^{\bar{s}} s \, d\Phi(s). \tag{5}$$

When s is larger than or equal to $\bar{s}$, which occurs with probability $1 - \Phi(\bar{s})$, the lender recoups the fixed proportion $\bar{s}$ of the expected revenue V. If s falls below $\bar{s}$, the lender takes all of the realized revenue while expending a verification cost which equals a fraction μ of the revenue.[12] The following assumption on the distribution of s is imposed.

Assumption 1. (a) The p.d.f $\phi(\cdot)$ is positive, bounded, and continuously differentiable on $(0, \infty)$, and (b) $s\phi(s) / [1 - \Phi(s)]$ is an increasing function of s.[13]

It is easy to show that for $\bar{s} > 0$,

$$\Gamma'(\bar{s}; \Phi) = -[1 - \Phi(\bar{s})] < 0,$$

---

[12] Note that $\Gamma(\bar{s}; \Phi) + \Psi(\bar{s}; \Phi) = 1 - \mu \int_{0}^{\bar{s}} s \, d\Phi(s) < 1$, indicating that there is a direct deadweight loss $\mu \int_{0}^{\bar{s}} s \, d\Phi(s)$ due to costly monitoring.

[13] The assumption that $s\phi(s) / [1 - \Phi(s)]$ is increasing in s is weaker than the increasing hazard assumption commonly made in the incentive contract literature, which requires $\phi(s) / [1 - \Phi(s)]$ to be monotonically increasing in s. Yet the latter property is already satisfied by a fairly large class of distributions.

$$\Psi'\left(\bar{s};\Phi\right)=1-\Phi\left(\bar{s}\right)-\mu\bar{s}\phi\left(\bar{s}\right)>0,\ \text{if}\ \bar{s}<\hat{s},$$

and

$$\Gamma'\left(\bar{s};\Phi\right)+\Psi'\left(\bar{s};\Phi\right)=-\mu\bar{s}\phi\left(\bar{s}\right)<0,$$

where the primes denote derivatives with respect to $\bar{s}$ and $\hat{s}$ satisfies $1-\Phi\left(\hat{s}\right)-\mu\hat{s}\phi\left(\hat{s}\right)=0$. We rule out the possibility of credit rationing by requiring $V\Psi\left(\hat{s};\Phi\right)$ to be no less than the opportunity cost of funds for the lender (see Williamson, 1986). Thus the domain of $\bar{s}$ we are interested in is $[0,\hat{s})$ and $\Psi'\left(\bar{s};\Phi\right)>0$ on this interval.[14]

We now apply this generic debt contract framework to the bank-firm relationship. The firm's revenue can be written as $V^f\omega$, where $V^f\equiv PF\left(k,l\right)\theta$ is freely observable to the bank, and ω is the risk that can be observed by the bank only with a cost.[15] The bank-firm contract specifies a monitoring threshold, denoted by $\bar{\omega}$, for the firm/location specific productivity ω. Conditional on the region specific productivity θ, the expected return to the firm is then given by $PF\left(k,l\right)\theta\Gamma^f\left(\bar{\omega};\Phi^l\right)$ and the revenue of the bank from lending to the firms in its region is $PF\left(k,l\right)\theta\Psi^b\left(\bar{\omega};\Phi^l\right)$, where $\Gamma^f\left(\bar{\omega};\Phi^l\right)$ and $\Psi^b\left(\bar{\omega};\Phi^l\right)$ result from substituting $\left(\bar{\omega};\Phi^l\right)$ for $\left(\bar{s};\Phi\right)$ in (4) and (5).[16]

The contracting problem between the bank and its depositors specifies a monitoring threshold for the bank risk θ. To fit this into the generic setup, write the bank's revenue as $V^b\theta$, where $V^b\equiv PF\left(k,l\right)\Psi^b\left(\bar{\omega};\Phi^l\right)$. Here $\bar{\omega}$—the monitoring threshold specified explicitly in the bank-firm contract—is freely observable to both the bank and the depositors. Let $\bar{\theta}$ represent the monitoring threshold for θ in the bank-depositor contract. Then the

---

[14]If the lender has logarithmic utility then the relevant $\hat{s}$ is the one that maximizes the function $\tilde{\Psi}$ defined in (15) below.

[15]From the bank's perspective, monitoring $s^f\equiv\theta\omega$ is equivalent to monitoring ω given its information in θ.

[16]By the law of large numbers, the revenue of the bank from lending to all of the firms in its region is the same as the expected revenue from lending to one firm, the expectation taken over the distribution of ω and conditional on θ.

expected return to the bank from the contract is $V^b \Gamma^b \left(\bar{\theta}; \Phi^r \right)$ and the expected return to the depositors is $V^b \Psi^d \left(\bar{\theta}; \Phi^r \right)$, where $\Gamma^b \left(\bar{\theta}; \Phi^r \right)$ and $\Psi^d \left(\bar{\theta}; \Phi^r \right)$ obtain from substituting $\left(\bar{\theta}; \Phi^r \right)$ for $\left(\bar{s}; \Phi \right)$ in (4) and (5). Note, however, that if the depositors are risk averse with respect to bank risks (as is the case in Model A), what the depositors care is their expected utility, which obviously differs from the expected financial return offered by the contract. Details are provided in the next subsection.

Optimal Competitive Contract

To motivate competitive banking assume that although a bank can lend only to the firms within one region, there is no restriction on which region this might be (free region entry). As a result each bank offers contracts that maximize the expected return to the firms in its region of operation such that the bank itself at least earns the riskless return on its own funds. The optimal competitive contract is formally stated as solving the problem below (recall that the risk-free interest rate is normalized to be one). To simplify notations, the dependence of the Γ and Ψ functions on Φ^l and Φ^r will be omitted.

Problem 1.

$$\max_{k,l,N^d,\bar{\omega},\bar{\theta}} \frac{P_1}{P_2} F\left(k,l \right) \Gamma^f \left(\bar{\omega} \right)$$

subject to

$$\frac{P_1}{P_2} F\left(k,l \right) \Psi^b \left(\bar{\omega} \right) \Gamma^b \left(\bar{\theta} \right) \geq \frac{N^b}{P_2}, \tag{6}$$

$$\left[1 - \Phi^r \left(\bar{\theta} \right) \right] U \left(\frac{P_1}{P_2} F\left(k,l \right) \Psi^b \left(\bar{\omega} \right) \bar{\theta} + \frac{M^d - N^d}{P_2} \right) \tag{7}$$

$$+ \int_0^{\bar{\theta}} U \left(\frac{P_1}{P_2} F\left(k,l \right) \Psi^b \left(\bar{\omega} \right) \theta \left(1 - \mu \right) + \frac{M^d - N^d}{P_2} \right) \phi^r \left(\theta \right) d\theta$$

$$\geq U \left(\frac{M^d}{P_2} \right)$$

$$R^k k + W l \leq N^f + N^b + N^d, \tag{8}$$

15

where $0 \leq N^d \leq M^d$. Here $P_1 F(k, l) \Gamma^f(\bar{\omega})$ is the expected return to the firm, uncondi-tional on θ, from the contract in period 1. Dividing this by the period-2 price level P_2 yields the firm's expected consumption and hence expected utility. Inequality (6) is the individual rationality (IR) constraint for the bank, which says that the bank must obtain at least what it can earn by investing all of its own funds in the riskless security. The amount of the bank's financial capital equals the rental value of the physical capital stock it owns plus its after-transfer money balance, M^b. That is, $N^b \equiv R^k K^b + M^b$.

Inequality (7), the IR constraint for the depositors, needs some explanation. A depos-itor may choose to allocate her money balance M^d between bank deposits N^d and invest-ment $(M^d - N^d)$ in the risk-free security, though in equilibrium $M^d = N^d$ because of the zero supply of the risk-free bond. No matter what happens to bank solvency, the depositor gets back $(M^d - N^d)$ from the risk-free investment. When $\theta \geq \bar{\theta}$, which occurs with prob-ability $1 - \Phi^r(\bar{\theta})$, the depositor receives fixed payment $P_1 F(k, l) \Psi^b(\bar{\omega}) \bar{\theta}$ from the deposit contract and utility level $U\left(P_1 F(k, l) \Psi^b(\bar{\omega}) \bar{\theta} / P_2 + (M^d - N^d) / P_2\right)$ from period-2 con-sumption. When $\theta < \bar{\theta}$, the depositor receives $P_1 F(k, l) \Psi^b(\bar{\omega}) \theta (1 - \mu)$, net of monitoring costs, from the deposit contract and utility level $U\left(P_1 F(k, l) \Psi^b(\bar{\omega}) \theta (1 - \mu) / P_2 + (M^d - N^d) / P_2\right)$. The expected utility from the portfolio $(N^d, M^d - N^d)$ must be no less than putting all of M^d into the risk-free bond, which yields the certain level of utility $U(M^d / P_2)$. In Prob-lem 1 the applicable functional form for $U(\cdot)$ depends on the risk-sharing arrangement among the depositors. For Model A $U(\cdot)$ is logarithmic, while for Model B it is linear.

Finally, inequality (8) is the flow-of-funds constraint for the firms. The total bill for the firms' factor inputs is $R^k k + W l$, which has to be covered by the internal funds of the firms themselves, $N^f \equiv R^k K^f$, and bank loans that equal the sum of bank capital N^b

and deposits N^d. In Problem 1 N^f and N^b are taken as given.

Define the "debt-equity ratios" for the bank and firms, denoted by ζ^b and ζ^f respectively, as

$$\zeta^b \equiv \frac{N^d}{N^b}, \quad \zeta^f \equiv \frac{N^b + N^d}{N^f}.$$

Let the model-indicator function $\chi(I)$ equal 1 if $I = A$ and 0 if $I = B$. As shown in the Appendix, the solution to Problem 1 satisfies the conditions listed below, where we impose the equilibrium condition $M^d - N^d = 0$ to simplify notations, without neglecting the necessity to take derivatives via the term $(M^d - N^d)/P_2$.

$$F_k(k, l) = q^I(\bar\omega, \bar\theta) \frac{R^k}{P_1}, \tag{9}$$

$$F_l(k, l) = q^I(\bar\omega, \bar\theta) \frac{W}{P_1}, \tag{10}$$

$$\chi(I)\tilde\Psi^d(\bar\theta) + (1 - \chi(I))\log\Psi^d(\bar\theta) - \log\Gamma^b(\bar\theta) = \log(\zeta^b), \tag{11}$$

$$q^I(\bar\omega, \bar\theta)\Psi^b(\bar\omega)\Gamma^b(\bar\theta) = \frac{1}{1 + \zeta^b}\frac{\zeta^f}{1 + \zeta^f}, \tag{12}$$

where

$$q^A(\bar\omega, \bar\theta) \equiv \frac{-\Gamma^{f\prime}(\bar\omega)}{\Gamma^f(\bar\omega)\Psi^{b\prime}(\bar\omega) - \Gamma^{f\prime}(\bar\omega)\Psi^b(\bar\omega)}\frac{-\Gamma^{b\prime}(\bar\theta)\Delta(\bar\theta)}{\Gamma^b(\bar\theta)\tilde\Psi^{d\prime}(\bar\theta) - \Gamma^{b\prime}(\bar\theta)}, \tag{13}$$

$$q^B(\bar\omega, \bar\theta) \equiv \frac{-\Gamma^{f\prime}(\bar\omega)}{\Gamma^f(\bar\omega)\Psi^{b\prime}(\bar\omega) - \Gamma^{f\prime}(\bar\omega)\Psi^b(\bar\omega)}\frac{-\Gamma^{b\prime}(\bar\theta)}{\Gamma^b(\bar\theta)\Psi^{d\prime}(\bar\theta) - \Gamma^{b\prime}(\bar\theta)\Psi^d(\bar\theta)}, \tag{14}$$

$$\tilde\Psi^d(\bar\theta) \equiv [1 - \Phi^r(\bar\theta)]\log(\bar\theta) + \int_0^{\bar\theta}\log[\theta(1 - \mu)]\phi^r(\theta)\,d\theta, \tag{15}$$

$$\Delta(\bar\theta) \equiv [1 - \Phi^r(\bar\theta)]\frac{1}{\bar\theta} + \int_0^{\bar\theta}\frac{1}{\theta(1 - \mu)}\phi^r(\theta)\,d\theta. \tag{16}$$

Equations (9) and (10) are the first-order conditions for factor demand, where the presence of the factor q creates wedges between the marginal products of factor inputs and their real prices (q takes the value q^I in Model I, $I \in \{A, B\}$). We shall call q the

17

financial friction indicator, as it reflects the distortions caused by the agency problems in the two-sided financial contracting. If either $\bar{\omega} > 0$ or $\bar{\theta} > 0$ (or both) then $q\left(\bar{\omega}, \bar{\theta}\right)$ is strictly greater than one. Here $\bar{\omega} > 0$ indicates a positive default rate by the firms and reflects the agency cost in the bank-firm relationship. This is what the existing literature on credit market imperfections has typically focused on. On the other hand, $\bar{\theta} > 0$ corresponds to a positive rate of default by the banks (to the depositors) and reflects the agency cost in the bank-depositor relationship. The variable $q\left(\bar{\omega}, \bar{\theta}\right)$ measures the overall distortions caused by the conventionally studied credit frictions and the sort of banking frictions we introduce. All aspects of financial frictions that are relevant for the determination of employment are captured by q. In the model's general equilibrium to be described below, q is a sufficient statistic for equilibrium employment. Note that q is an increasing function of $\bar{\omega}$ and $\bar{\theta}$, with $\lim_{\bar{\omega}, \bar{\theta} \to 0} q = 1$ in both models.

Equations (11) and (12) reflect the fact that the optimal competitive contract entails binding IR constraints for both the bank and the depositors. Essentially, the terms of contract dictate a division of expected revenues between borrowers and lenders. Since either $\tilde{\Psi}^d\left(\bar{\theta}\right) - \log \Gamma^b\left(\bar{\theta}\right)$ or $\Psi^d\left(\bar{\theta}\right) / \Gamma^b\left(\bar{\theta}\right)$ is increasing in $\bar{\theta}$, equation (11) indicates that the bank's default probability increases along with $\bar{\theta}$ when it has a larger debt-equity ratio ζ^b. The increase in $\bar{\theta}$ implies a larger share of expected revenues received by the depositors, relative to the share received by the bank, in the bank-depositor contract. Equation (12) indicates that given ζ^b and $\bar{\theta}$, the firms' default probability increases along with $\bar{\omega}$ when their debt-equity ratio ζ^f increases. The increase in $\bar{\omega}$ implies a larger share of expected revenues that goes to the firms in the bank-firm contract.

2.3   General Equilibrium

To make the analysis tractable we further assume that the production function $F(\cdot)$ takes the standard Cobb-Douglas form, i.e., $F(K, L) = K^\alpha L^{1-\alpha}$, $\alpha \in (0, 1)$. This immediately implies, via (9) and (10), that $(1 - \alpha) R^k K = \alpha W L$. Using this relationship, together with $N^d = M^d$, $N^b = R^k K^b + M^b$, the equality version of the flow-of-funds constraint (8), and $WL = M - M^w$, we have

$$\zeta^b = \frac{(1 - \alpha) z^d}{\alpha K^b / K + (1 - \alpha) z^b}, \quad \frac{1}{1 + \zeta^b} \frac{\zeta^f}{1 + \zeta^f} = \alpha \frac{K^b}{K} + (1 - \alpha) z^b, \tag{17}$$

where

$$z^b \equiv \frac{M^b}{M - M^w} \text{ and } z^d \equiv \frac{M^d}{M - M^w}$$

are the fractions of $(M - M^w)$ possessed by the banks and depositors, respectively, with $z^b + z^d = 1$. The pair (z^b, z^d) represents the distribution of initial money balances between the banks and depositors.

With ζ^b and ζ^f given by (17), equations (11) and (12) determine the equilibrium values of the default thresholds $\bar{\theta}$ and $\bar{\omega}$. Given $\bar{\theta}$, $\bar{\omega}$ (hence q), and the real wage rate ν, the equilibrium employment L is determined by the following condition

$$(1 - \alpha) K^\alpha L^{-\alpha} = q(\bar{\omega}, \bar{\theta}) \nu. \tag{18}$$

An important implication of equation (18) is that the financial friction indicator q is a *sufficient statistic* for the determination of equilibrium employment L.

Furthermore, output and consumption in period 1 (by the workers) is given by

$$C_1 = F(K, L) \varphi(\bar{\omega}, \bar{\theta}), \tag{19}$$

where

$$\varphi(\bar{\omega}, \bar{\theta}) \equiv \Gamma^f(\bar{\omega}) + \Psi^b(\bar{\omega}) \left[\Gamma^b(\bar{\theta}) + \Psi^d(\bar{\theta}) \right]. \tag{20}$$

19

Note that the net output factor $\varphi\left(\bar{\omega}, \bar{\theta}\right) < 1$ for $\bar{\omega}, \bar{\theta} > 0$, indicating a direct deadweight loss due to costly monitoring.

If we think of the initial distribution of purchasing powers as a specification of the vector $\left(M, M^b, M^d, M^w\right)$, with $M \equiv M^b + M^d + M^w$, then the aspect of this distribution that is relevant for allocations is simply the division of $\left(M - M^w\right)$ between M^b and M^d, as represented by the pair $\left(z^b, z^d\right)$. Given $\left(z^b, z^d\right)$, the only role of M is to determine the price level $P_1 = M/C_1$ through the quantity equation (2). Since the total nominal wage bill is $WL = \nu P_1 L$ and must equal $M^b + M^d$ in a cleared loan market, the relationship

$$M^w = M - \nu P_1 L \tag{21}$$

must hold for the specification of the initial distribution to be internally consistent. Equation (21) can be seen as a model-consistent rule that the government uses to determine M^w for any given specification of $\left(z^b, z^d, M\right)$. The initial distribution of purchasing powers can thus be equivalently characterized by $\left(z^b, z^d, M, M^w\right)$.

In period 2, the price level equals $P_2 = M/C_2$ given the output endowment C_2. The terms $\left(\bar{\omega}, \bar{\theta}\right)$ of period-1 financial contract determine the division of C_2 among the entrepreneurs, bankers, and depositors, who only consume in period 2. The share of period-2 purchasing power possessed by each type of agents equals the share of revenues that goes to that type of agents as dictated by the period-1 contract. Hence total entrepreneurial consumption C_2^f, banker consumption C_2^b, and depositor consumption C_2^d in period 2 are given by

$$C_2^f = C_2 \frac{\Gamma^f\left(\bar{\omega}\right)}{\varphi\left(\bar{\omega}, \bar{\theta}\right)}, \quad C_2^b = C_2 \frac{\Psi^b\left(\bar{\omega}\right) \Gamma^b\left(\bar{\theta}\right)}{\varphi\left(\bar{\omega}, \bar{\theta}\right)}, \quad C_2^d = C_2 \frac{\Psi^b\left(\bar{\omega}\right) \Psi^d\left(\bar{\theta}\right)}{\varphi\left(\bar{\omega}, \bar{\theta}\right)}, \tag{22}$$

respectively.

Formally, a competitive equilibrium with banking frictions and two-sided financial contracting is an initial distribution of purchasing powers (z^b, z^d, M, M^w), an allocation $(L, C_1, C_2^f, C_2^b, C_2^d)$, a price system (P_1, W, R^k, P_2), and terms of financial contract $(\bar{\omega}, \bar{\theta})$ such that

(1) Given (z^b, z^d), the period-1 contract terms and allocations, $\bar{\omega}$, $\bar{\theta}$, L, and C_1, are determined by (11)-(12), with ζ^b and ζ^f given by (17), and (18)-(19).

(2) Given M, the price levels P_1 and P_2 are determined by the quantity equations, i.e., $P_1 = M/C_1$ and $P_2 = M/C_2$. In addition $W = \nu P_1$ and $R^k = \alpha W L / (1 - \alpha)$.

(3) the period-2 consumption allocation (C_2^f, C_2^b, C_2^d) is given by (22).

(4) M^w is set in accordance with the rule (21) for any given specification of (z^b, z^d, M).

3   Banking Riskiness and Aggregate Fluctuations

3.1   The Default versus Risk Effects of Banking Riskiness Shocks

In this section we introduce the concept of banking riskiness shocks and study their effects under different risk-sharing arrangements among the depositors. Our formulation of these shocks parallels the formulation of entrepreneurial riskiness shocks in the earlier work of Williamson (1987) and the recent work of Christiano, Motto, and Rostagno (2003, 2009). These authors consider the CSV problem between banks and nonfinancial firms but not the problem between banks and depositors. In the inspiring work of Williamson (1987), savers delegate monitoring of entrepreneurs to a large financial intermediary, which perfectly diversifies away all of the credit risks and is able to guarantee the depositors a risk-free return. There are thus agency costs between the entrepreneurs and the financial intermediary, but no such costs between the financial intermediary and the

depositors. In that environment stochastic changes in the riskiness of the entrepreneurs'
projects generate aggregate fluctuations, fluctuations that would not obtain were there no
costly monitoring on the outcomes of the entrepreneurial projects. In his model, however,
banking frictions (i.e., agency costs between financial intermediaries and depositors) are
absent, leaving no role for stochastic changes in banking riskiness to play. This abstrac-
tion might be innocuous for episodes where banking frictions are not severe enough to
deserve attention. Yet the experience of the GFC pinpoints the importance of banking
frictions and stochastic changes in the riskiness of banking.

In this paper we assume that the bank/region specific productivity θ follows a unit-
mean log-normal distribution on $(0, \infty)$, i.e., $\log(\theta) \sim \mathcal{N}\left(-\frac{1}{2}\sigma_\theta^2, \sigma_\theta^2\right)$, where $\mathcal{N}$ stands for
the normal distribution.[17] In our model, it is the costly verification of θ that gives rise
to equilibrium bankruptcy of banks. The default rate of banks tends to zero as σ_θ tends
to zero from the right. When σ_θ equals zero, the distribution of θ becomes degenerate,
and the informational asymmetry between the banks and the depositors disappears. The
dispersion parameter σ_θ captures the extent of the riskiness of banking. Here we allow σ_θ
to be random. Specifically, its realization is given by

$$\sigma_\theta = \bar{\sigma}_\theta + \varepsilon, \tag{23}$$

where the mean level of riskiness, $\bar{\sigma}_\theta$, is a positive constant, and ε is a random disturbance
bounded away from $-\bar{\sigma}_\theta$. We interpret ε as the banking riskiness shock.

In our view, shocks to banking riskiness are highly relevant in the light of the erratic
behavior of the interest rate spreads on banks' external finance, and especially so during
the recent GFC. The historical average of the spread between the 3-month certificate of

---

[17]The distribution is completed by assigning a zero p.d.f. for $\theta = 0$.

deposits (CD) rate and the 3-month T-bill rate was about 75 basis points (annualized), based on a sample period from 1973Q1 to 2009Q4. From 2001Q1 to 2007Q2, the spread averaged only 27 basis points. In contrast, its average in the second half of 2007 and the year of 2008 rose to as high as 153 basis points, with a spike at 252 basis points in the fourth quarter of 2008. In our model, there is a direct linkage between the extent of banking riskiness and the external finance premium faced by the banks. The gross interest rate at which the banks borrow from the depositors is simply the non-default payment specified in the bank-depositor contract divided by the amount of deposits, i.e., $R^b = P_1 F(K, L) \Psi^b(\bar{\omega}) \bar{\theta}/N^d$. Using the binding IR constraint for the bank, equation (6), in Problem 1, we obtain the model's interest rate spread on bank deposits: $R^b - 1 = \bar{\theta}/\left[\Gamma^b(\bar{\theta}) \zeta^b\right] - 1$ (recall again that the risk-free rate is normalized to be unity).[18] Other things equal, an increase in σ_θ raises R^b, and fluctuations in banking riskiness give rise to fluctuations in the spread. Shocks to banking riskiness seem to be more important than shocks to entrepreneurial riskiness during the GFC, as there was a sharp increase in the spread between the bank CD rate and the T-bill rate, but not in the spread between the bank lending rate and the CD rate (Figure 1a).

What are the financial and macroeconomic consequences of shocks to banking riskiness? The answer to this question hinges on the way the financial friction indicator q—the sufficient statistic for the determination of employment—responds to the shocks. The way q responds, in turn, depends on the prevailing risk sharing arrangements among the depositors. An important point to notice is that with imperfect risk sharing (Model A) the response of q and hence employment L includes both a pure default effect and

---

[18]Similarly, the risk spread faced by the firms in the model is given by $R^f - 1 = \bar{\omega}/\left[\Psi^b(\bar{\omega})\left(\Gamma^b(\bar{\theta}) + \Psi^d(\bar{\theta})\right)\right] - 1$.

a risk effect, while with perfect risk sharing (Model B) only the pure default effect is present. The pure *default effect* results from the fact that the optimal contract dictates equilibrium bankruptcy of banks. This effect will be present as long as the banks' default rate changes with the shock, even when the depositors are risk neutral with respect to banking risks. The *risk effect* results from the fact that the payment streams under the optimal contract are uncertain for the depositors. This effect obtains when risk sharing among the depositors is imperfect so that they remain risk averse with respect to banking risks.

In this paper we propose a model-based approach to decompose the overall effect of a banking riskiness shock into the pure default effect and the risk effect. Specifically, two measures are developed to assess these effects: a within-model measure and a between-model measure. We take the imperfect risk sharing model—Model A—to be the "true" model, as perfect risk sharing seems to be less realistic. When the realization of the banking riskiness shock is ε, the equilibrium value of the financial friction indicator is given by $q\left(\varepsilon\right) = q^{A}\left(\bar{\omega}^{A}\left(\varepsilon\right), \bar{\theta}^{A}\left(\varepsilon\right); \varepsilon\right)$, where $\left(\bar{\omega}^{A}\left(\varepsilon\right), \bar{\theta}^{A}\left(\varepsilon\right)\right)$ denotes the default thresholds in the equilibrium of Model A with banking riskiness shock ε, and the functional form of $q^{A}\left(\cdot\right)$ is given by (13). Define the within-model pure default component of $q\left(\varepsilon\right)$ as

$$q^{D}\left(\varepsilon|A\right) = q^{B}\left(\bar{\omega}^{A}\left(\varepsilon\right), \bar{\theta}^{A}\left(\varepsilon\right); \varepsilon\right),$$

where the functional form of $q^{B}\left(\cdot\right)$ is given by (14). The residual of $q\left(\varepsilon\right)$ over $q^{D}\left(\varepsilon|A\right)$ is

24

then regarded as the risk component of $q(\varepsilon)$:

$$
\begin{aligned}
q^R(\varepsilon|A) &= \frac{q^A\left(\bar{\omega}^A(\varepsilon), \bar{\theta}^A(\varepsilon); \varepsilon\right)}{q^B\left(\bar{\omega}^A(\varepsilon), \bar{\theta}^A(\varepsilon); \varepsilon\right)} \\
&= \frac{\Delta\left(\bar{\theta}^A(\varepsilon)\right)\left[\Gamma^b\left(\bar{\theta}^A(\varepsilon)\right)\Psi^{d\prime}\left(\bar{\theta}^A(\varepsilon)\right) - \Gamma^{b\prime}\left(\bar{\theta}^A(\varepsilon)\right)\Psi^d\left(\bar{\theta}^A(\varepsilon)\right)\right]}{\Gamma^b\left(\bar{\theta}^A(\varepsilon)\right)\tilde{\Psi}^{d\prime}\left(\bar{\theta}^A(\varepsilon)\right) - \Gamma^{b\prime}\left(\bar{\theta}^A(\varepsilon)\right)},
\end{aligned}
$$

where the specific forms for the functions in the second line are understood to depend on ε. Note that the risk component $q^R(\varepsilon|A)$ depends only on the bank default threshold and not on the firm default threshold.

When the banking riskiness shock changes from ε to ε', our within-model measures of the pure default effect and the risk effect of the shock on q are

$$
DE(\varepsilon, \varepsilon'|A) = \log q^D(\varepsilon'|A) - \log q^D(\varepsilon|A),
$$

and

$$
RE(\varepsilon, \varepsilon'|A) = \log q^R(\varepsilon'|A) - \log q^R(\varepsilon|A),
$$

respectively. Obviously the effect of the shock on $\log q$ is exactly the sum of the above two effects. In the light of condition (18), the within-model measures of the pure default and risk effects of the banking riskiness shock on equilibrium employment L are naturally $-1/\alpha$ times the corresponding effects on q.

The within-model measures described above rely on using the equilibrium default thresholds prevailing in Model A. Clearly, with the same realization of banking riskiness shock the value of $(\bar{\omega}, \bar{\theta})$ will vary from one model to the other. Our between-model measures take this into account. Let $\left(\bar{\omega}^B(\varepsilon), \bar{\theta}^B(\varepsilon)\right)$ denote the default thresholds in the equilibrium of Model B (perfect risk sharing) with banking riskiness shock ε. The

25

between-model pure default component of $q(\varepsilon) = q^A\left(\bar{\omega}^A(\varepsilon), \bar{\theta}^A(\varepsilon); \varepsilon\right)$ is defined as

$$q^D(\varepsilon|B) = q^B\left(\bar{\omega}^B(\varepsilon), \bar{\theta}^B(\varepsilon); \varepsilon\right).$$

The corresponding risk component is then the residual of $q(\varepsilon)$ over $q^D(\varepsilon|B)$:

$$q^R(\varepsilon|A, B) = \frac{q^A\left(\bar{\omega}^A(\varepsilon), \bar{\theta}^A(\varepsilon); \varepsilon\right)}{q^B\left(\bar{\omega}^B(\varepsilon), \bar{\theta}^B(\varepsilon); \varepsilon\right)}.$$

This between-model measure of the risk component of q reflects not only that imperfect risk sharing creates severer financial distortions for the same values of the default thresholds, but also that the equilibrium default thresholds themselves will change once we switch from perfect risk sharing to imperfect risk sharing.

When the banking riskiness shock changes from ε to ε', our between-model measures of the pure default effect and the risk effect of the shock on q are

$$DE(\varepsilon, \varepsilon'|B) = \log q^D(\varepsilon'|B) - \log q^D(\varepsilon|B),$$

and

$$RE(\varepsilon, \varepsilon'|A, B) = \log q^R(\varepsilon'|A, B) - \log q^R(\varepsilon|A, B),$$

respectively. The effect of the shock on $\log q$ is exactly the sum of these two effects. Again, the pure default and risk effects of the banking riskiness shock on equilibrium employment are $-1/\alpha$ times the corresponding effects on q.

3.2 Quantitative Evaluations

To evaluate quantitatively the default and risk effects of banking riskiness shocks on the financial friction indicator and the level of employment, we calibrate Model A as follows. Let a time period correspond to a quarter. The weight of leisure relative to consumption

in worker utility, ν, is chosen to deliver $L = 1/3$ absent shocks and frictions. The elasticity parameter in the production function, α, is set to be $1/2$, implying an asset-net worth ratio of about 2 for the firms (see Bernanke, Gertler, and Gilchrist, 1999).[19] Normalizing $K = 1$ and $K^b = 0$, the value of z^b is set to be 0.076, which matches the historical average of an asset-net worth ratio of 13.18 for U.S. commercial banks.[20] The monitoring cost parameter, μ, is set to be 0.36.[21] Similar to the bank/region specific productivity, we assume that the firm/location specific productivity ω follows a unit-mean log-normal distribution: $\log(\omega) \sim \mathcal{N}\left(-\frac{1}{2}\sigma_\omega^2, \sigma_\omega^2\right)$. To isolate the effects of banking riskiness shocks, we assume that σ_ω is fixed, while σ_θ follows the specification in (23) and is therefore subjected to stochastic disturbances. The value of σ_ω and the mean value of σ_θ, $\bar{\sigma}_\theta$, are chosen to match (1) an annualized spread between the firms' borrowing rate and the risk-free rate of 293 basis points, and (2) an annualized spread between the banks' borrowing rate and the risk-free rate of 75 basis points.[22]

Figures 3 and 4 depict the effects of banking riskiness shocks on the financial friction indicator q and employment L. Figure 3 shows the effects of the shocks relative to the frictionless economy, where $q = 1$ (hence $\log q = 0$) identically. The top-left part shows the value of $q(\varepsilon)$ (solid line), the within-model measure of its pure default component $q^D(\varepsilon|A)$ (dashed line), and the between-model measure of that component $q^D(\varepsilon|B)$ (dash-dot line), all expressed in log and multiplied by one hundred. The gap between the solid

---

[19]If the variable K in the production function were interpreted literally as "physical capital", then $1/2$ would be too large a value for α. Nevertheless, a broader interpretation may be adopted: the variable may be thought to include bank and firm managers' human capital, e.g., managerial skills, as well.

[20]This calculation is based on "Assets and Liabilities of Commercial Banks in the United States" of the Federal Reserve. The sample period is 1973Q1-2009Q4.

[21]By comparing the value of a firm as a going concern with its liquidation value, Alderson and Betker (1995) estimate that liquidation costs are equal to approximately 36 percent of firms assets.

[22]The empirical measures of the risk-free rate, the banks' borrowing rate, and the firms' borrowing rate are the 3-month T-bill rate, the 3-month CD rate, and the prime lending rate, respectively. The data are from the Board of Governors of the Federal Reserve System. The sample period is again 1973Q1-2009Q4.

line and the dashed (resp. dash-dot) line represents the within (resp. between)-model measure of the risk component. The bottom-left part plots the corresponding effects on log employment relative to the value that would prevail with $q = 1$, also expressed in percentage points. The right parts of the figure illustrate the percent contributions by the pure default components to the overall effects of banking riskiness shocks.

[Insert Figures 3 & 4 about here.]

As is apparent from the figure, the effect of a positive (resp. negative) shock to banking riskiness is to raise (resp. lower) q and reduce (resp. increase) employment L. The effects are asymmetric around $\varepsilon = 0$ in that the effects of positive shocks are stronger. This is because negative shocks drive the economy toward the situation without banking frictions, which provides the limit for the strength of the effects. When a negative shock is sufficiently large, virtually 100% of q is made of the pure default component. As the shock gets larger in algebraic value, the risk component gains importance at the expense of the pure default component, with the between-model measure of the contribution by the pure default component somewhat larger than the corresponding within-model measure. At the mean level of riskiness ($\varepsilon = 0$), the contribution by the pure default component is 88% by the within measure and 90% by the between measure. When the shock reaches $\varepsilon = 0.07$, which is particularly interesting since it generates a bank spread of about 250 basis points per annum—the highest point during the GFC, the contribution by the pure default component is 82% by the within measure and 85% by the between measure.

Figure 4 shows the effects of banking riskiness shocks ($\varepsilon \neq 0$) relative to the mean level of riskiness ($\varepsilon = 0$, or $\sigma_\theta = \bar{\sigma}_\theta$). The effects on q plotted in the figure include the overall effect $\log q(\varepsilon) - \log(q(0))$ and the within (resp. between)-model measure of the pure

default effect, i.e., $DE(0, \varepsilon|A)$ (resp. $DE(0, \varepsilon|B)$), all expressed in percentage points. The gap between the overall effect and a particular measure of the pure default effect is the associated risk effect. As discussed earlier, the effects on log employment are simply $-1/\alpha$ times the effects on q, the percent contributions by the pure default effect versus the risk effect being the same for these two variables. For the range of banking riskiness shocks shown in the figure, the pure default effect contributes about 73% to the overall effect according to the within measure, and about 77% according to the between measure. When $\varepsilon = 0.07$, which generates a bank spread similar in value to the one present in the fourth quarter of 2008, the overall effect of the banking riskiness shock, relative to the mean level of riskiness, is a decline in employment by 6.57 percent. According to our decomposition, the pure default effect leads to a 4.76 percent employment decline by the within measure, and a 5.05 decline by the between measure. On the other hand, the risk effect produces a 1.81 percent employment decline by the within measure, and a 1.52 percent decline by the between measure. Although the pure default effects are quantitatively more important, the risk effects are not to be neglected.

4   Conclusions

This paper develops a model of banking frictions and banking riskiness, the importance of which is highlighted by the recent Global Financial Crisis (GFC). A model-based approach is proposed to decompose the effect of a banking riskiness shock into a pure default effect and a risk effect. Although the default effect is quantitatively more important, the risk effect is not to be neglected. When the shock generates a bank spread similar in value to the peak during the GFC, the overall effect is a decline in employment by 6.57 percent.

The pure default effect leads to a 4.76 percent employment decline by a within-model measure, and a 5.05 decline by a between-model measure. The remaining is attributed to the risk effect.

We conclude by suggesting two directions for future research. First our analysis can be extended to include "deposit rationing" as a possible equilibrium outcome so that another dimension in which banking riskiness shocks exert influence on the economy can be explored. Credit rationing, whereby entrepreneurs are unable to obtain the bank loans they desire, has been extensively studied in the literature by, for example, Stiglitz and Weiss (1981) and Williamson (1986). This type of rationing happens on the asset side of the bank balance sheet. A different type of rationing can happen on the liability side of the bank balance sheet, whereby banks are unable to take in the amount of deposits they desire. The latter type of rationing will be an interesting topic to explore in future research. Second, entrepreneurial riskiness shocks, as analyzed in Williamson (1987) and Christiano, Motto, and Rostagno (2003, 2009), can be considered in tandem with banking riskiness shocks. Both kinds of shocks are likely to be relevant for financial and macroeconomic fluctuations, but their relative importance might vary from one episode to another. Furthermore, it is possible that these two kinds of shocks are correlated with each other. Investigating the role they play jointly is an important direction for business cycle studies.

References

[1] Alderson, M. and B. Betker (2001), "Liquidation costs and capital structure," *Empirical Corporate Finance*, **3**, 92–116.

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

[3] Bernanke, B., M. Gertler, and S. Gilchrist (1999), "The financial accelerator in a quantitative business cycle framework," in John Taylor and Michael Woodford (ed.), *Handbook of Macroeconomics*, Vol. 1, Elsevier.

[4] Boyd, J. and B. Smith (1994), "How good are standard debt contracts?: Stochastic versus nonstochastic monitoring in a costly state verification model," *Journal of Business*, **67**, 539–562.

[5] Carlstrom, C. and T. Fuerst (1997), "Agency costs, net worth, and business fluctuations," *American Economic Review*, **87**(5), 893–910.

[6] Cerasi, V. and S. Daltung (2000), "The optimal size of a bank: Costs and benefits of diversification," *European Economic Review*, **44**, 1701–26.

[7] Christiano, L. and M. Eichenbaum (1992), "Liquidity effects and the monetary transmission mechanism," *American Economic Review Papers and Proceedings*, 346–353.

[8] Christiano, L., R. Motto, and M. Rostagno (2003), "The Great Depression and the Friedman-Schwartz Hypothesis," *Journal of Money, Credit and Banking*, **35**(6), 1119-97.

[9] Christiano, L., R. Motto, and M. Rostagno (2009), "Financial factors in economic fluctuations," Working paper.

[10] Fisher, J. (1999), "Credit market imperfection and the heterogeneous response of firms to monetary shocks," *Journal of Money, Credit and Banking*, **31**, 187–211.

[11] Gale, D. and M. Hellwig (1985), "Incentive compatible debt contracts I: the one period problem," *Review of Economic Studies*, **52**, 647–664.

[12] Hellwig, M. (2000), "Financial intermediation with risk aversion," *Review of Economic Studies*, **67**, 719–742.

[13] Krasa, S. and A. Villamil (1992a), "Monitoring the monitor: an incentive structure for a financial intermediary," *Journal of Economic Theory*, **57**, 197–221.

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

[15] Krasa, S. and A. Villamil (2000), "Optimal contract when enforcement is a decision variable," *Econometrica*, **68**, 119–134.

[16] Modigliani, F. and M. Miller (1958), "The cost of capital, corporation finance and the theory of investment," *American Economic Review*, **48**, 261–297.

[17] Mookherjee, D. and I. Png (1989), "Optimal auditing, insurance, and redistribution," *Quarterly Journal of Economics*, **104**, 399–415.

[18] Stiglitz, J. and A. Weiss (1981), "Credit rationing in markets with imperfect information," *American Economic Review*, **70**, 393–410.

[19] Townsend, R. (1979), "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, **21**, 265–293.

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

[21] Williamson, S. (1987), "Financial intermediation, business failures, and real business cycles," *Journal of Political Economy*, **95**(6), 1196–1216.

[22] Williamson, S. (1989), "Restrictions on financial intermediaries and implications for aggregate fluctuations: Canada and the United States, 1870-1913," in Blanchard, O., Fischer, S. (ed.), *NBER Macroeconomics Annual*.

[23] Zeng, Z. (2007), "The price of size and financial market allocations," *Economic Theory*, **30**, 21–48.

[24] Zeng, Z. (2010), "A Theory of the non-neutrality of money with banking frictions and bank recapitalization," Working paper.

Appendix. Derivation of the Optimality Conditions for Problem 1.

Model A

We first show that conditions (9)-(12) hold. In the derivation below we impose the fact that $M^d - N^d = 0$ in equilibrium to simplify notations, without neglecting the necessity to take derivatives via the term $R\left(M^d - N^d\right)/P_2$. Let λ^b and λ^d be the Lagrangian multipliers for (6) and (7), respectively. With $U\left(\cdot\right)$ taking the log form, the first-order conditions with respect to $\bar{\omega}$ and $\bar{\theta}$ are

$$0 = \left[\Gamma^{f\prime}\left(\bar{\omega}\right) + \lambda^b \Psi^{b\prime}\left(\bar{\omega}\right)\Gamma^b\left(\bar{\theta}\right)\right] + \frac{\lambda^d}{\frac{P_1}{P_2}F\left(k,l\right)}\frac{\Psi^{b\prime}\left(\bar{\omega}\right)}{\Psi^b\left(\bar{\omega}\right)}, \tag{A.1}$$

$$0 = \lambda^b \Psi^b\left(\bar{\omega}\right)\Gamma^{b\prime}\left(\bar{\theta}\right) + \frac{\lambda^d}{\frac{P_1}{P_2}F\left(k,l\right)}\tilde{\Psi}^{d\prime}\left(\bar{\theta}\right). \tag{A.2}$$

Equations (A.1) and (A.2) imply

$$\lambda^b = \frac{-\Gamma^{f\prime}\left(\bar{\omega}\right)\tilde{\Psi}^{d\prime}\left(\bar{\theta}\right)}{\Psi^{b\prime}\left(\bar{\omega}\right)\left[\Gamma^b\left(\bar{\theta}\right)\tilde{\Psi}^{d\prime}\left(\bar{\theta}\right) - \Gamma^{b\prime}\left(\bar{\theta}\right)\right]}, \tag{A.3}$$

$$\lambda^d = \frac{P_1}{P_2}F\left(k,l\right)\frac{\Gamma^{f\prime}\left(\bar{\omega}\right)\Psi^b\left(\bar{\omega}\right)\Gamma^{b\prime}\left(\bar{\theta}\right)}{\Psi^{b\prime}\left(\bar{\omega}\right)\left[\Gamma^b\left(\bar{\theta}\right)\tilde{\Psi}^{d\prime}\left(\bar{\theta}\right) - \Gamma^{b\prime}\left(\bar{\theta}\right)\right]}. \tag{A.4}$$

The first-order conditions with respect to k and l are given by (9) and (10), where

$$q\left(\bar{\omega},\bar{\theta}\right) \equiv \frac{\frac{\lambda^d}{\frac{P_1}{P_2}F(k,l)}\frac{\Delta\left(\bar{\theta}\right)}{\Psi^b\left(\bar{\omega}\right)}}{\Gamma^f\left(\bar{\omega}\right) + \lambda^b \Psi^b\left(\bar{\omega}\right)\Gamma^b\left(\bar{\theta}\right) + \frac{\lambda^d}{\frac{P_1}{P_2}F(k,l)}}.$$

Substitution of (A.3) and (A.4) into the above definition gives the expression of q in terms of $\bar{\omega}$ and $\bar{\theta}$ as in (13), with $I = A$.

At the optimum constraints (6) and (7) bind, implying

$$P_1 F\left(k,l\right)\Psi^b\left(\bar{\omega}\right)\Gamma^b\left(\bar{\theta}\right) = N^b, \tag{A.5}$$

$$\log\left(\frac{P_1}{P_2}F\left(k,l\right)\Psi^b\left(\bar{\omega}\right)\right) + \tilde{\Psi}^d\left(\bar{\theta}\right) = \log\left(\frac{M^d}{P_2}\right). \tag{A.6}$$

Substituting (A.5) into (A.6) yields (11) with $I = A$.

To derive (12), note that the linear homogeneity of $F\left(\cdot\right)$ together with (9) and (10) imply

$$P_1 F\left(k,l\right) = q\left(R^k k + Wl\right). \tag{A.7}$$

Substituting (A.7) and the equality version of (8) into (A.5) yields (12).

We then show that $\lim_{\bar{\omega},\bar{\theta}\to 0} q = 1$ and that q increases with $\bar{\omega}$ and $\bar{\theta}$, hence $q > 1$ for all $\bar{\omega}, \bar{\theta} > 0$ in the neighborhood of $\bar{\omega}, \bar{\theta} = 0$. Rewrite $q\left(\bar{\omega},\bar{\theta}\right) \equiv \varrho\left(\bar{\omega}\right) \varkappa\left(\bar{\theta}\right)$, where

$$\left[\varrho\left(\bar{\omega}\right)\right]^{-1} \equiv \Psi^b\left(\bar{\omega}\right) - \Gamma^f\left(\bar{\omega}\right) \frac{\Psi^{b\prime}\left(\bar{\omega}\right)}{\Gamma^{f\prime}\left(\bar{\omega}\right)}, \qquad \left[\varkappa\left(\bar{\theta}\right)\right]^{-1} \equiv \frac{1}{\Delta\left(\bar{\theta}\right)}\left[1 - \Gamma^b\left(\bar{\theta}\right) \frac{\tilde{\Psi}^{d\prime}\left(\bar{\theta}\right)}{\Gamma^{b\prime}\left(\bar{\theta}\right)}\right].$$

Look at the term $\varrho\left(\bar{\omega}\right)$. We have $\left[\varrho\left(\bar{\omega}\right)\right]^{-1} < 1$ or $\varrho\left(\bar{\omega}\right) > 1$ for all $\bar{\omega} > 0$ since $-\Psi^{b\prime}\left(\bar{\omega}\right)/\Gamma^{f\prime}\left(\bar{\omega}\right) < 1$ and $\Gamma^f\left(\bar{\omega}\right) + \Psi^b\left(\bar{\omega}\right) < 1$. Also, $\lim_{\bar{\omega}\to 0}\left[\varrho\left(\bar{\omega}\right)\right]^{-1} = 1$ since $\lim_{\bar{\omega}\to 0}\left[-\Psi^{b\prime}\left(\bar{\omega}\right)/\Gamma^{f\prime}\left(\bar{\omega}\right)\right]$ $= 1$ and $\lim_{\bar{\omega}\to 0}\left[\Gamma^f\left(\bar{\omega}\right) + \Psi^b\left(\bar{\omega}\right)\right] = 1$. By differentiation,

$$\frac{\partial \varrho^{-1}}{\partial \bar{\omega}} = \frac{\Gamma^f\left(\bar{\omega}\right)}{\left[\Gamma^{f\prime}\left(\bar{\omega}\right)\right]^2}\left[\Psi^{b\prime}\left(\bar{\omega}\right)\Gamma^{f\prime\prime}\left(\bar{\omega}\right) - \Psi^{b\prime\prime}\left(\bar{\omega}\right)\Gamma^{f\prime}\left(\bar{\omega}\right)\right].$$

But

$$\Psi^{b\prime}\left(\bar{\omega}\right)\Gamma^{f\prime\prime}\left(\bar{\omega}\right) - \Psi^{b\prime\prime}\left(\bar{\omega}\right)\Gamma^{f\prime}\left(\bar{\omega}\right) = -\mu\phi^l\left(\bar{\omega}\right)\left[1 - \Phi^l\left(\bar{\omega}\right)\right]\left[1 + \frac{\bar{\omega}\phi^l\left(\bar{\omega}\right)}{1 - \Phi^l\left(\bar{\omega}\right)} + \frac{\bar{\omega}\phi^{l\prime}\left(\bar{\omega}\right)}{\phi^l\left(\bar{\omega}\right)}\right].$$

To sign the above expression we consider two cases. Case 1: $\lim_{\bar{\omega}\to 0}\phi^l\left(\bar{\omega}\right) > 0$. In this case $\lim_{\bar{\omega}\to 0}\left[\Psi^{b\prime}\left(\bar{\omega}\right)\Gamma^{f\prime\prime}\left(\bar{\omega}\right) - \Psi^{b\prime\prime}\left(\bar{\omega}\right)\Gamma^{f\prime}\left(\bar{\omega}\right)\right] = -\mu\lim_{\bar{\omega}\to 0}\phi^l\left(\bar{\omega}\right) < 0$. Case 2: $\lim_{\bar{\omega}\to 0}\phi^l\left(\bar{\omega}\right) = 0$. But Assumption 1(a) requires $\phi^l\left(\cdot\right)$ to be positive, bounded, and continuously differentiable on $(0,\infty)$. Hence in this case we must have $\lim_{\bar{\omega}\to 0}\phi^{l\prime}\left(\bar{\omega}\right) > 0$. This means that for $\bar{\omega}$ positive and sufficiently close to 0, we have $\phi^l\left(\bar{\omega}\right) > 0$ and $\phi^{l\prime}\left(\bar{\omega}\right) > 0$ and hence $\left[\Psi^{b\prime}\left(\bar{\omega}\right)\Gamma^{f\prime\prime}\left(\bar{\omega}\right) - \Psi^{b\prime\prime}\left(\bar{\omega}\right)\Gamma^{f\prime}\left(\bar{\omega}\right)\right] < 0$. In both cases when $\bar{\omega}$ is positive and sufficiently close to 0, we have $\partial \varrho^{-1}/\partial \bar{\omega} < 0$ and hence $\partial q/\partial \bar{\omega} > 0$.

Now look at the term $\varkappa\left(\bar{\theta}\right)$. Using (15) and (16), we have

$$\left[\varkappa\left(\bar{\theta}\right)\right]^{-1} = \frac{G\left(\bar{\theta}\right) - H\left(\bar{\theta}\right)}{J\left(\bar{\theta}\right)},$$

where

$$G\left(\bar{\theta}\right) \equiv \frac{\Gamma^b\left(\bar{\theta}\right)\Psi^{d\prime}\left(\bar{\theta}\right) - \Gamma^{b\prime}\left(\bar{\theta}\right)\bar{\theta}}{-\Gamma^{b\prime}\left(\bar{\theta}\right)},$$

$$H\left(\bar{\theta}\right) \equiv \frac{\Gamma^b\left(\bar{\theta}\right)\bar{\theta}\phi^r\left(\bar{\theta}\right)}{-\Gamma^{b\prime}\left(\bar{\theta}\right)}\left[-\log\left(1 - \mu\right) - \mu\right],$$

$$J\left(\bar{\theta}\right) \equiv \left[1 - \Phi^r\left(\bar{\theta}\right)\right] + \bar{\theta}\int_0^{\bar{\theta}}\frac{1}{\theta\left(1 - \mu\right)}\phi^r\left(\theta\right) d\theta.$$

By differentiation,

$$\frac{dG\left(\bar{\theta}\right)}{d\bar{\theta}} = \frac{d}{d\bar{\theta}}\left\{\left[\int_{\bar{\theta}}^{\infty}\left(\theta - \bar{\theta}\right)\phi^r\left(\theta\right) d\theta\right]\left[1 - \mu\frac{\bar{\theta}\phi^r\left(\bar{\theta}\right)}{1 - \Phi^r\left(\bar{\theta}\right)}\right] + \bar{\theta}\right\}$$

$$= -\mu\left[E\left(\theta|\theta \geq \bar{\theta}\right) - 2\bar{\theta}\right]\phi^r\left(\bar{\theta}\right) - \mu\left[E\left(\theta|\theta \geq \bar{\theta}\right) - \bar{\theta}\right]\left[\bar{\theta}\phi^{r\prime}\left(\bar{\theta}\right) + \frac{\bar{\theta}\left[\phi^r\left(\bar{\theta}\right)\right]^2}{1 - \Phi^r\left(\bar{\theta}\right)}\right] + \Phi^r\left(\bar{\theta}\right),$$

34

where $E\left(\theta | \theta \geq \bar{\theta}\right)$ denotes the truncated expectation of θ, with $\lim_{\bar{\theta} \to 0} E\left(\theta | \theta \geq \bar{\theta}\right) = 1$. To sign this derivative consider two cases. Case 1. $\lim_{\bar{\theta} \to 0} \phi^r\left(\bar{\theta}\right) > 0$. In this case $\lim_{\bar{\theta} \to 0} dG\left(\bar{\theta}\right)/d\theta < 0$. Case 2. $\lim_{\bar{\theta} \to 0} \phi^r\left(\bar{\theta}\right) = 0$. In this case $\lim_{\bar{\theta} \to 0} \phi^{r\prime}\left(\bar{\theta}\right) > 0$ as implied by Assumption 1(a), which requires $\phi^r\left(\cdot\right)$ to be positive, bounded, and continuously differentiable on $(0, \infty)$. Furthermore $\bar{\theta}$ goes to zero at a slower rate than $\Phi^r\left(\bar{\theta}\right)$ as $\lim_{\bar{\theta} \to 0}\left[\bar{\theta}/\Phi^r\left(\bar{\theta}\right)\right] = \lim_{\bar{\theta} \to 0}\left[1/\phi^r\left(\bar{\theta}\right)\right] = \infty$. Hence for $\bar{\theta}$ positive and sufficiently close to zero we have $\Phi^r\left(\bar{\theta}\right)$ dominated by the negative terms and hence $dG\left(\bar{\theta}\right)/d\theta < 0$. In sum $dG\left(\bar{\theta}\right)/d\theta < 0$ in the neighborhood of $\bar{\theta} = 0$. Also,

$$\frac{1}{-\log\left(1-\mu\right)-\mu}\frac{dH\left(\bar{\theta}\right)}{d\bar{\theta}} = \left[E\left(\theta | \theta \geq \bar{\theta}\right) - 2\bar{\theta}\right]\phi^r\left(\bar{\theta}\right) + \left[E\left(\theta | \theta \geq \bar{\theta}\right) - \bar{\theta}\right]\left\{\bar{\theta}\phi^{r\prime}\left(\bar{\theta}\right) + \frac{\bar{\theta}\left[\phi^r\left(\bar{\theta}\right)\right]^2}{1-\Phi^r\left(\bar{\theta}\right)}\right\}.$$

To sign this derivative again consider two cases. Case 1. $\lim_{\bar{\theta} \to 0} \phi^r\left(\bar{\theta}\right) > 0$. In this case $\lim_{\bar{\theta} \to 0} dH\left(\bar{\theta}\right)/d\bar{\theta} > 0$ (note that $-\log\left(1-\mu\right)-\mu > 0$). Case 2. $\lim_{\bar{\theta} \to 0} \phi^r\left(\bar{\theta}\right) = 0$. In this case $\lim_{\bar{\theta} \to 0} \phi^{r\prime}\left(\bar{\theta}\right) > 0$ as implied by Assumption 1(a). This means that for $\bar{\theta}$ positive and sufficiently close to zero, both $\phi^r\left(\bar{\theta}\right)$ and $\phi^{r\prime}\left(\bar{\theta}\right)$ are positive, hence $dH\left(\bar{\theta}\right)/d\bar{\theta} > 0$. In sum $dH\left(\bar{\theta}\right)/d\bar{\theta} > 0$ in the neighborhood of $\bar{\theta} = 0$. Finally,

$$\frac{dJ\left(\bar{\theta}\right)}{d\bar{\theta}} = \phi^r\left(\bar{\theta}\right)\left(\frac{1}{1-\mu}-1\right) + \int_0^{\bar{\theta}}\frac{1}{\theta\left(1-\mu\right)}\phi^r\left(\theta\right)d\theta \geq 0.$$

We therefore conclude that $d\varkappa^{-1}/d\bar{\theta} < 0$ or $d\varkappa/d\bar{\theta} > 0$ and hence $\partial q/\partial\bar{\theta} > 0$ in the neighborhood of $\bar{\theta} = 0$.

Model B

Again, let λ^b and λ^d be the Lagrangian multipliers for (6) and (7), respectively. With $U\left(\cdot\right)$ taking the linear form, the first-order conditions with respect to $\bar{\omega}$ and $\bar{\theta}$ are

$$\Gamma^{f\prime}\left(\bar{\omega}\right) + \Psi^{b\prime}\left(\bar{\omega}\right)\left[\lambda^b\Gamma^b\left(\bar{\theta}\right) + \lambda^d\Psi^d\left(\bar{\theta}\right)\right] = 0, \tag{B.1}$$

$$\lambda^b\Gamma^{b\prime}\left(\bar{\theta}\right) + \lambda^d\Psi^{d\prime}\left(\bar{\theta}\right) = 0. \tag{B.2}$$

Equations (B.1) and (B.2) imply

$$\lambda^b = -\frac{\Gamma^{f\prime}\left(\bar{\omega}\right)}{\Psi^{b\prime}\left(\bar{\omega}\right)}\frac{\Psi^{d\prime}\left(\bar{\theta}\right)}{\Gamma^b\left(\bar{\theta}\right)\Psi^{d\prime}\left(\bar{\theta}\right) - \Gamma^{b\prime}\left(\bar{\theta}\right)\Psi^d\left(\bar{\theta}\right)}, \tag{B.3}$$

$$\lambda^d = \frac{\Gamma^{f\prime}\left(\bar{\omega}\right)}{\Psi^{b\prime}\left(\bar{\omega}\right)}\frac{\Gamma^{b\prime}\left(\bar{\theta}\right)}{\Gamma^b\left(\bar{\theta}\right)\Psi^{d\prime}\left(\bar{\theta}\right) - \Gamma^{b\prime}\left(\bar{\theta}\right)\Psi^d\left(\bar{\theta}\right)}. \tag{B.4}$$

The first-order conditions with respect to k and l are given by (9) and (10), where

$$q\left(\bar{\omega}, \bar{\theta}\right) \equiv \frac{\lambda^d}{\Gamma^f\left(\bar{\omega}\right) + \Psi^b\left(\bar{\omega}\right)\left[\lambda^b\Gamma^b\left(\bar{\theta}\right) + \lambda^d\Psi^d\left(\bar{\theta}\right)\right]}.$$

Substitution of (B.3) and (B.4) into the above definition gives the expression of q in terms of $\bar{\omega}$ and $\bar{\theta}$ as in (14), with $I = B$.

At the optimum constraints (6) and (7) bind, implying

$$P_1 F\left(k, l\right) \Psi^b\left(\bar{\omega}\right) \Gamma^b\left(\bar{\theta}\right) = N^b, \tag{B.5}$$

$$P_1 F\left(k, l\right) \Psi^b\left(\bar{\omega}\right) \Psi^d\left(\bar{\theta}\right) = M^d. \tag{B.6}$$

Substituting (B.5) into (B.6) yields (11) with $I = B$. The derivation of (12) is the same as in Model A.

We then show that $q > 1$ for all $\bar{\omega}, \bar{\theta} > 0$ and $\lim_{\bar{\omega}, \bar{\theta} \to 0} q = 1$. Since $\left(-\Psi^{d\prime}\left(\bar{\theta}\right) / \Gamma^{b\prime}\left(\bar{\theta}\right)\right) < 1$ and $\left(-\Psi^{b\prime}\left(\bar{\omega}\right) / \Gamma^{f\prime}\left(\bar{\omega}\right)\right) < 1$ for all $\bar{\omega}, \bar{\theta} > 0$, we have

$$q^{-1} = \left[\Psi^b\left(\bar{\omega}\right) - \Gamma^f\left(\bar{\omega}\right) \frac{\Psi^{b\prime}\left(\bar{\omega}\right)}{\Gamma^{f\prime}\left(\bar{\omega}\right)}\right] \left[\Psi^d\left(\bar{\theta}\right) - \Gamma^b\left(\bar{\theta}\right) \frac{\Psi^{d\prime}\left(\bar{\theta}\right)}{\Gamma^{b\prime}\left(\bar{\theta}\right)}\right] < \left[\Gamma^f\left(\bar{\omega}\right) + \Psi^b\left(\bar{\omega}\right)\right] \left[\Gamma^b\left(\bar{\theta}\right) + \Psi^d\left(\bar{\theta}\right)\right] < 1,$$

and hence $q > 1$ for all $\bar{\omega}, \bar{\theta} > 0$. Since $\lim_{\bar{\theta} \to 0}\left(-\Psi^{d\prime}\left(\bar{\theta}\right) / \Gamma^{b\prime}\left(\bar{\theta}\right)\right) = 1$, $\lim_{\bar{\omega} \to 0}\left(-\Psi^{b\prime}\left(\bar{\omega}\right) / \Gamma^{f\prime}\left(\bar{\omega}\right)\right) = 1$, $\lim_{\bar{\theta} \to 0}\left[\Gamma^b\left(\bar{\theta}\right) + \Psi^d\left(\bar{\theta}\right)\right] = 1$, $\lim_{\bar{\omega} \to 0}\left[\Gamma^f\left(\bar{\omega}\right) + \Psi^b\left(\bar{\omega}\right)\right] = 1$, we have $\lim_{\bar{\omega}, \bar{\theta} \to 0} q^{-1} = 1$.

We now show that $\partial q / \partial \bar{\omega} > 0$ and $\partial q / \partial \bar{\theta} > 0$ in the neighborhood of $\bar{\omega}, \bar{\theta} = 0$. Rewrite $q\left(\bar{\omega}, \bar{\theta}\right) \equiv \varrho\left(\bar{\omega}\right) \varkappa\left(\bar{\theta}\right)$, where $\varrho\left(\bar{\omega}\right)$ is the same as in Model A and

$$\left[\varkappa\left(\bar{\theta}\right)\right]^{-1} \equiv \Psi^d\left(\bar{\theta}\right) - \Gamma^b\left(\bar{\theta}\right) \frac{\Psi^{d\prime}\left(\bar{\theta}\right)}{\Gamma^{b\prime}\left(\bar{\theta}\right)}.$$

The proof of $\partial q / \partial \bar{\omega} > 0$ follows that in Model A. The proof of $\partial q / \partial \bar{\theta} > 0$ also follows that proof with $\bar{\theta}$ replacing $\bar{\omega}$.
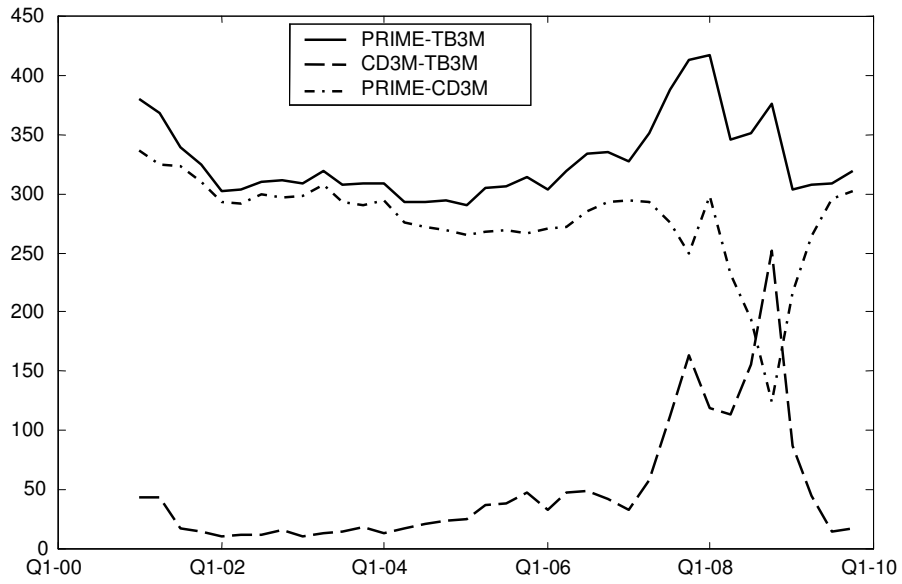
a. Interest rate spreads (in basis points per annum)
Solid: Prime lending rate minus 3-month T-bill rate; Dashed: 3-month bank CD rate minus 3-month T-bill rate; Dash-dot: prime lending rate minus 3-month bank CD rate.
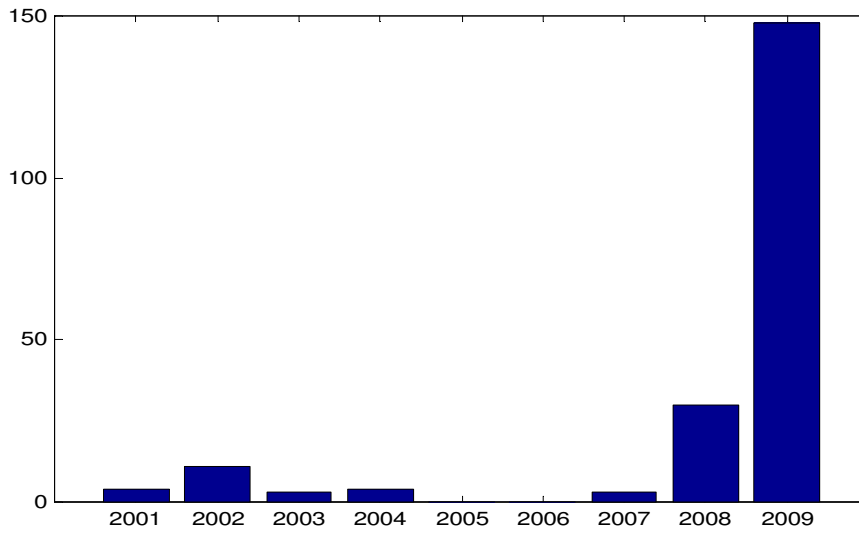
b. Number of failed FDIC-insured financial institutions
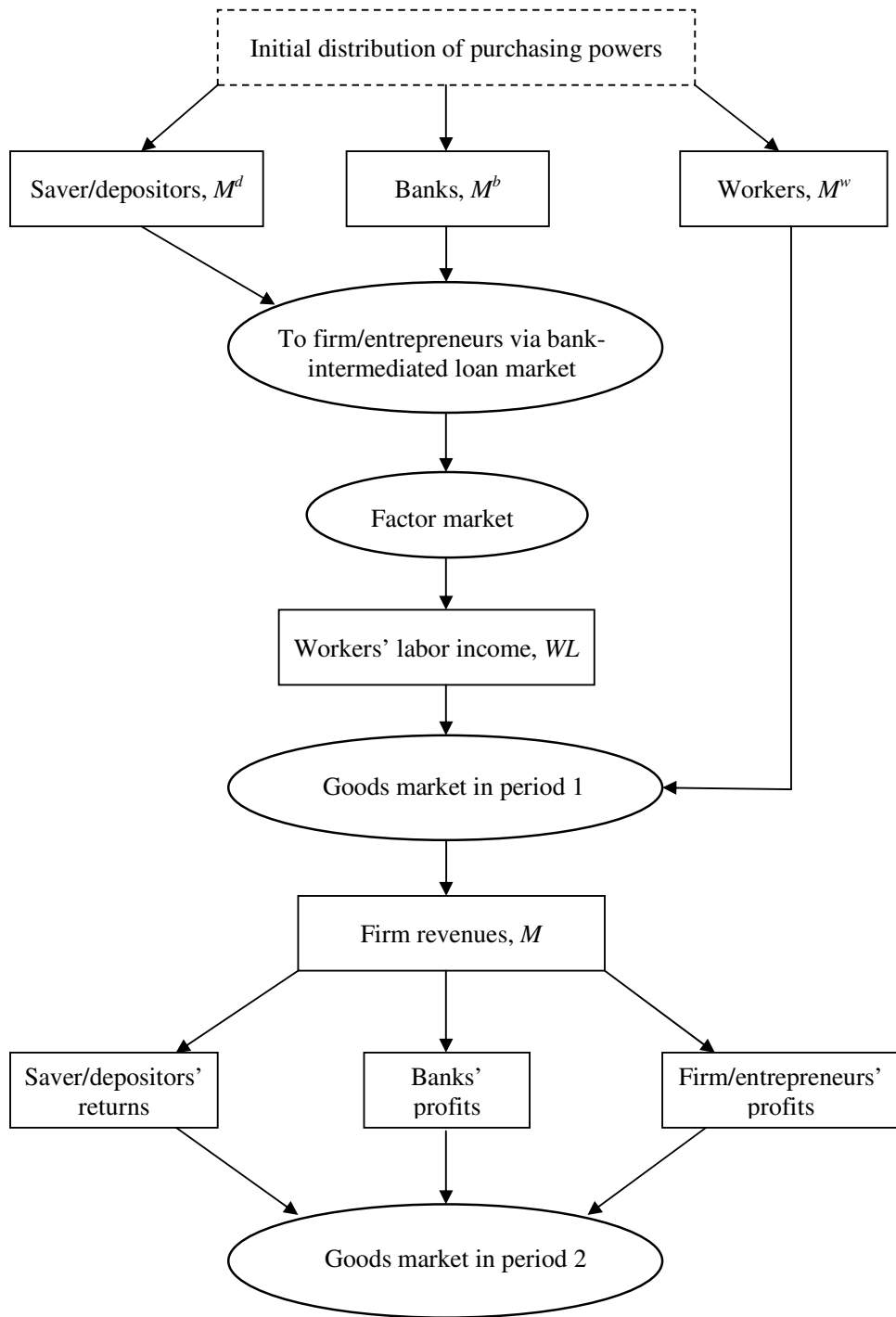
Figure 1. Interest rate spreads and bank failure

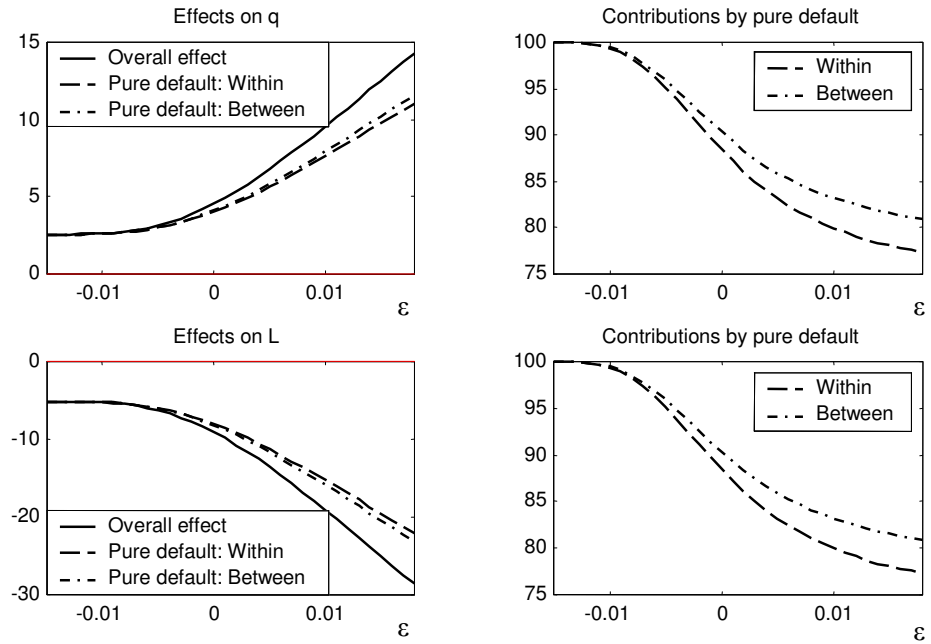**Figure 2. Flow of funds in the model**

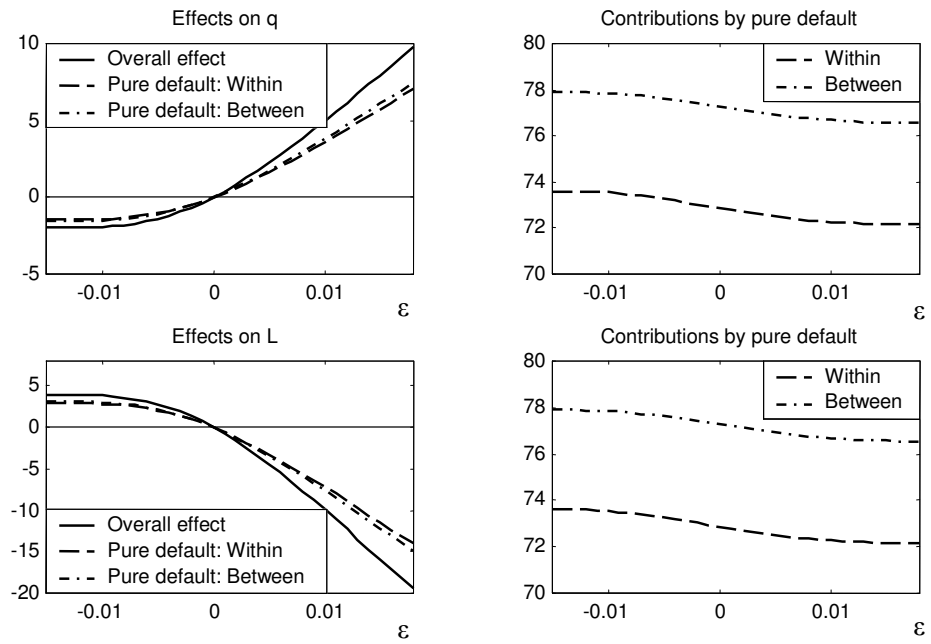**Figure 3. Effects of banking riskiness shocks (relative to frictionless)**

Figure 4. Effects of banking riskiness shocks (relative to mean)