Econ Theory (2013) 52:729–754
DOI 10.1007/s00199-011-0662-2

RESEARCH ARTICLE

A theory of the non-neutrality of money with banking frictions and bank recapitalization

Zhixiong Zeng

Received: 29 January 2010 / Accepted: 19 August 2011 / Published online: 7 September 2011

Abstract The unconventional monetary policy actions of the Federal Reserve during the recent Global Financial Crisis often involve implicit subsidies to banks. This paper offers a theory of the non-neutrality of money associated with capital injection into banks via nominal transfers, in an environment where banking frictions are present in the sense that there exists an agency problem between banks and their private-sector creditors. The analysis is conducted within a general equilibrium setting with two-sided financial contracting. We first show that even with perfect nominal flexibility, the recapitalization policy has real effects on the economy. We then introduce banking riskiness shocks and study optimal policy responses to such shocks.

Keywords Bankruptcy of banks · Banking riskiness shocks · Two-sided debt contract · Unconventional monetary policy · Financial crisis

JEL Classification E44 · E52 · D82 · D86

1 Introduction

The Federal Reserve took a variety of unconventional policy actions to cope with the recent Global Financial Crisis (GFC). As traditional interest rate policy that adjusts the federal funds rate was perceived to be ineffective (Cecchetti 2009), the Fed adopted

I thank an anonymous referee whose comments have led to great improvement of the paper. Participants at the Econometric Society World Congress 2010, the 16th Australasian Macroeconomics Workshop 2011, the Conference "The Role of Finance in Stabilizing the Past, Present, and Future Real Economy" held in DIW Berlin, and Monash University Department of Economics Workshop also provided helpful comments. I am especially indebted to Larry Christiano and Yi Jin for their encouragement.

Z. Zeng (✉)
Department of Economics, Monash University, Caulfield East, VIC 3145, Australia
e-mail: zhixiong.zeng@monash.edu

various measures of the so-called unconventional monetary policy.[1] In addition to injecting liquidity into the financial system (Brunnermeier 2009), some of the Fed's policy measures also had the flavor of providing capital subsidy to banks, a point forcefully made by Cecchetti (2009). This is certainly true when the Fed directly purchased assets previously held by banks, such as mortgage-backed securities, at above market prices. It can also be argued that lending by the Fed during the crisis almost always involved subsidies. By accepting collaterals at prices that were almost surely above their actual market prices (Tett 2008) and charging lower interest rates (relative to the federal funds rate) when banks were actually perceived by the market to be exposed to greater risks, Fed lending in effect recapitalized the borrowing banks through nominal transfers: on one hand, reserves and monetary base were created. On the other hand, banks were getting more funds than they could borrow from the market for the same interest rates and the same collateral assets. During the early phase of the crisis, the Fed attempted to stimulate discount borrowing, which is collateralized, by reducing substantially the premium charged on primary discount lending over the federal funds rate target and raising the term of lending from overnight to as long as 3 months. In addition, to remove the stigma attached to discount borrowing,[2] the Fed created the Term Auction Facility (TAF) in December 2007 and enlarged it later on in order to better provide funds to banks that need them most. The rules of the TAF allowed banks to pledge collaterals that might otherwise have little market value. With few exceptions, the interest rates paid on TAF loans were near or below the expected primary discount lending rate.[3]

To be sure, the unconventional monetary policy is multi-faceted. This paper focuses on one particular aspect of the policy, namely, implicit capital subsidy to banks financed by money creation. In light of the celebrated Modigliani–Miller theorem (Modigliani and Miller 1958), such recapitalization policy would be ineffective in stimulating employment and output in a world where banks can frictionlessly raise funds to finance the loans they make, as the capital structure of banks would be irrelevant for their lending activities and the real market value of their loan portfolios. In that kind of world, the classical dichotomy holds and recapitalization of banks by the monetary authority is neutral, despite that it does involve a real transfer that enlarges banks' net worth relative to debt (because other sectors of the economy are not getting the same nominal transfer). However, as will be demonstrated in this paper, once an agency problem is introduced to the relationship between banks and their private-sector creditors (henceforth "depositors" for ease of exposition),[4] the Modigliani–Miller theorem fails for

[1] See Reis (2009) and Goodfriend (2011) for reviews. Recent models of unconventional monetary policy include Gertler and Karadi (2011) and Cúrdia and Woodford (2011), among others.

[2] Traditionally, banks that borrowed from the discount window might be seen by other banks and institutions as having financial stress.

[3] For details, see Cecchetti (2009). The quantity of TAF lending turned out to be large. In January 2009 they constituted more than one fifth of the Fed's total assets. Similar programs established by the Fed include the Term Securities Lending Facility, the Primary Dealer Credit Facility, and the Term Asset-Backed Securities Loan Facility, etc.

[4] It should be clarified here that 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 into a single category of agents called "depositors." We also lump all types of banks into a single banking sector.

banks, the classical dichotomy breaks down, and money is no longer neutral when central bank policy takes the form of injecting money to the banking system to increase bank capital. In particular, a bank recapitalization effort by the monetary authority triggers a redistribution of wealth in favor of the banks, lowers their debt–equity ratio and costs of external finance, hence stimulates bank lending, and raises employment and output. Importantly, this non-neutrality result obtains even without any kind of nominal rigidities.[5] The potency of the bank recapitalization policy allows it to be used as a stabilization tool when the economy is subject to shocks to the riskiness of banking.[6]

Needless to say, understanding the mechanism through which policy works is crucial for assessing the effectiveness of central bank reactions to the crisis. Impotent policy is clearly not interesting. The main thrust of the paper is that to make sense of the bank recapitalization policy, one has to take seriously frictions on the liability side of the bank balance sheet, i.e., frictions in the relationship between banks and depositors. The reason is that it is precisely frictions on this side, rather than frictions on the asset side, that are responsible for the real effects of bank recapitalization policy. As is already well known, on the asset side of the bank balance sheet, there might exist informational asymmetry regarding the ability of (non-financial) firms to repay their loans, giving rise to an agency problem between banks and firms as emphasized in the seminal work of Bernanke and Gertler (1989) and a large literature that follows.[7] Frictions of this kind are the literature's main focus thus far. We shall refer to them as "credit frictions," for the sake of distinguishing it from the informational asymmetry and agency problem on the liability side of the bank balance sheet, which we shall call "banking frictions." To introduce the latter kind of frictions, we apply the costly state verification (CSV) framework of Townsend (1979); Gale and Hellwig (1985), and Williamson (1986) to the bank–depositor relationship. In our model, banks face idiosyncratic risks and depositors have to expend monitoring costs in order to verify banks' capacities to repay. We emphasize that bank recapitalization by the monetary authority is neutral when banking frictions are absent, even if the conventionally studied credit frictions are present. This implies that what credit frictions do is at best to amplify and propagate the policy's real effects, which are brought forth solely by the existence of banking frictions. We are, thus, compelled to give special attention to the roles banking frictions play. Modeling banking frictions and studying their implications for the effects of bank recapitalization policy is precisely the goal of this paper.

In a model that allows for perfect nominal flexibility, some other sort of frictions must be employed to generate the non-neutrality of money. In Lucas (1972) misperceptions theory, it is the imperfect information about the overall price level that temporarily misleads suppliers and generates real effects of money supply shocks. It seems that information on money supply and other policy instruments is available to

[5] Diamond and Rajan (2006) analyze how changes in money supply affect real activities through a liquidity version of the bank lending channel, without relying on sticky prices, reserve requirements, or deposit insurance.

[6] The extent of banking riskiness is represented by a dispersion parameter in the distribution of idiosyncratic bank risks and is assumed to stochastic in the paper.

[7] Examples include Carlstrom and Fuerst (1997); Bernanke et al. (1999); Fisher (1999), and Christiano et al. (2003, 2009), etc.

the public with little delay, so there is no serious signal extraction problem to solve. Hence, the misperceptions story might not be particularly relevant for studying the effects of unconventional monetary policy. In contrast, this paper assumes full information on all aggregate variables, but uses a different kind of information problem to generate the non-neutrality of money. The problem here concerns costly revelation of banks' information to depositors, which leads to the breakdown of the Modigliani–Miller theorem and gives rise to a non-trivial role for banks' capital structure. The basic framework of banking frictions this paper builds on is laid out in Zeng (2002). The current paper presents a two-period model with risk-averse depositors. An infinite-horizon version with risk-neutral depositors is analyzed in Zeng (2010).

The rest of the paper is organized as follows. Section 2 outlines the economic environment, the agents, and their financial relationships, as well as the production and information structure. Section 3 analyzes two-sided financial contracting with idiosyncratic bank risks. The general equilibrium effects of the bank recapitalization policy and the optimal stabilization policy are studied in Sect. 4. The last section concludes. All proofs are relegated to the Appendix.

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.[8] 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.[9] Since the firms' internal funds are generated entirely from the current rental value of the capital stock they own, in a market-clearing equilibrium, they must borrow additional funds to finance their purchase of labor inputs supplied by the workers plus rental services provided by the stock of physical capital owned by the banks. Our model thus emphasizes

[8] An infinite-horizon version of the model is presented in Zeng (2010), who assumes perfect insurance among the depositors so that they are effectively risk neutral with respect to bank risks. This assumption allows for the usage of a representative-household setup when characterizing the saving behavior. In the present paper there is no perfect insurance among the depositors. Agents receiving different shocks will end up with different levels of wealth. 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.

[9] As capital is productive only in period 1, its price is zero in period 2 and equals the rental rate in period 1.

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 factor inputs, giving rise to a financially distorted labor market. In essence, what we are proposing is a finance-augmented neoclassical theory of production and employment.

At the beginning of period 1, the agents are endowed with initial money balances of certain amounts. Since the government provides nominal transfers (money injections) before financial contracts are negotiated in the period, these initial balances are irrelevant. What matters is the amount of money balances held by each type of agents that results *after* the transfers are made. Let the after-transfer 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 receive any money balance). The total amount of after-transfer money balance is then $M \equiv M^d + M^b + M^w$. There is a risk-free bond of zero supply. The interest rate on this bond, i.e., the risk-free nominal interest rate, is pegged by the government at $R > 1$.[10]

The funds 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 nominal wage rate W. In fact, the loan market clears when[11]

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

The sum then becomes labor income at the hands of the workers. The workers use WL plus their after-transfer money balance 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*:

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

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 1 illustrates the flow of funds in the model.

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 always equals ν. We treat the workers as being risk neutral since they

[10] The ability of the government to fix R in addition to the money balances is derived from the assumption that the savers consume in the second period only. Having R pegged allows us to abstract away from the effect of money injections on the extent of monetary distortions and to concentrate on financial distortions. See Zeng (2010) for an alternative setup that endogenizes the risk-free nominal interest rate and introduces a potential tradeoff between monetary and financial distortions.

[11] 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 $R^k K + WL = R^k K^f + \left(R^k K^b + M^b\right) + M^d$. This simplifies to (1) since $K = K^f + K^b$.

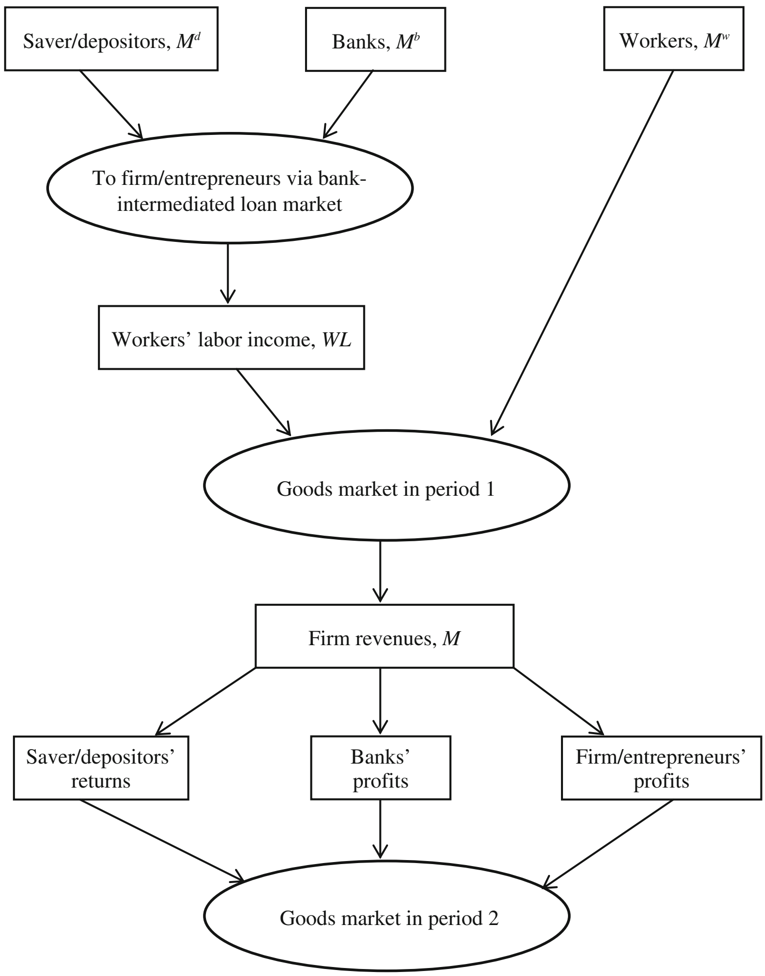

Fig. 1 Flow of funds in the model

do not face any idiosyncratic uncertainty at all: they always receive the full payment of wages since the firms must 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 utility function. In the financial relationships among these three parties, the banks face risks of default by the firms that borrow from them, and the depositors face risks of default by the banks where they made deposits.

The focus of our analysis is on how 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 affect the quantity of labor input and output produced in that period. Note that the terms of financial contracts also determine the division of firm revenues M among the contracting parties, and hence the distribution of period-2 purchasing powers (claims on consumption goods) among the depositors, bankers, and entrepreneurs. Put 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 employment and production. Before analyzing the financial contracting problem, a detailed description of the production and information structure is necessary.

2.2 The production and information structure

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 $F(\cdot)$ 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. $\Phi^r(\cdot)$ and p.d.f. $\phi^r(\cdot)$, and that, ω_{ij} is identical and independently distributed across locations, with c.d.f. $\Phi^l(\cdot)$ and p.d.f. $\phi^l(\cdot)$. 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 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 Townsend (1979); Gale and Hellwig (1985), and Williamson (1986) to model financial frictions and financial contracting. It is assumed that there is an ex post informational asymmetry regarding borrowers' 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. It is to be noted 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 non-financial firms. It is to be noted 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 on 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. Rather, it is a device designed to generate risks idiosyncratic to individual banks. If banks are subject 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 case, there are needs to “monitor the monitor,” in the terminology of Krasa and Villamil (1992a). Bank-level risks might stem from geographic confinement of an individual bank’s operation to specific areas, as in the United States when out-of-state branching was restricted (see Williamson 1989). They might also be due to the concentration of a bank’s lending activities in specific industries. Savings and loan associations in the United States, which historically concentrated on mortgage loans, were good examples. 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. An example appears in Krasa and Villamil (1992b), who consider the trade-off involved in increasing the size of a bank’s portfolio (i.e., lending to additional borrowers). In their model balancing the gains from decreased default risk with the losses from increased monitoring costs leads to an optimal scale for banks. Another example is Cerasi and Daltung (2000), who introduce considerations on the internal organization of banks that render scale economies in the banking sector rapidly exhausted.[12] 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. The model thus differs from Diamond (1984) and Williamson (1986), where the size of the financial intermediary grows without bound so that the cost of delegation vanishes in the limit.

3 Financial contracting with banking risks

3.1 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.[13] 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. It is to be noted that with risk-averse

[12] Specifically, loan officers, who are the ones actually making loans, have to be monitored by the banker.

[13] 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. Krasa and Villamil (1994) analyze optimal multilateral contracts when verification technology is either deterministic or stochastic. Strategic interaction between multiple financiers is considered by Khalil et al. (2007).

depositors standard debt contracts are optimal not only because they minimize the need for monitoring but also because they provide optimal risk sharing.[14]

Suppose that the borrower's revenue is given by Vs, where V is a component freely observable to both the borrower and 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. An incentive-compatible contract must specify a fixed payment for s in the non-monitoring set, otherwise the borrower will always report the value of s for which the payment is lowest among non-monitoring states. A standard debt contract with monitoring threshold $\bar{s}$ is an incentive-compatible contract with 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 $s \in \{s|s < \bar{s}\}$. The standard debt contract is particularly interesting because it resembles many financial contracts in the real world. It features fixed payment for non-default states and state-contingent payment when default occurs. Requiring the borrower to repay as much as possible in default states allows the fixed payment for non-default states to be minimized, thus minimizing the probability of verification and thus the expected monitoring cost.

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

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

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

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

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.[15] The following assumption on the distribution of s is imposed.

[14] 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).

[15] Note that $\Gamma(\bar{s}; \Phi) + \Psi(\bar{s}; \Phi) = 1 - \mu \int_0^{\bar{s}} s\mathrm{d}\Phi(s) < 1$, indicating that there is a direct deadweight loss $\mu \int_0^{\bar{s}} s\mathrm{d}\Phi(s)$ due to costly monitoring.

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.[16]

It can be shown that for $\bar{s}>0$,

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

and

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

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 $\Psi'(\bar{s};\Phi)>0$ on this interval.[17] It is interesting to note that changes in the monitoring threshold (and hence the default probability) generate redistributions of the expected revenue between the borrower and the lender. An increase in $\bar{s}$ reduces the share Γ received by the borrower, while raising the share Ψ received by the lender. The total effect on the returns to the two parties, however, is negative, since the marginal increase in the lender's share is less than the marginal increase in the borrower's share, reflecting the additional monitoring cost born by the lender at the margin.

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 $V^f \equiv PF(k,l)\theta$ is freely observable to the bank, and ω is the risk that can be observed by the bank only with a cost.[18] 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 $PF(k,l)\theta\Gamma^f\left(\bar{\omega};\Phi^l\right)$, and the revenue of the bank from lending to the firms in its region is $PF(k,l)\theta\Psi^b\left(\bar{\omega};\Phi^l\right)$, 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 $(\bar{s};\Phi)$ in (4) and (5).[19]

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 $V^b \equiv PF(k,l)\Psi^b\left(\bar{\omega};\Phi^l\right)$. Here, $\bar{\omega}$—the monitoring threshold in the bank–firm contract—is freely observable to both the bank and the depositors.

16 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.

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

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

19 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.

Let $\bar{\theta}$ represent the monitoring threshold for θ in the bank–depositor contract. Then the 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 $\Psi^d\left(\bar{\theta}; \Phi^r\right)$ obtained from substituting $\left(\bar{\theta}; \Phi^r\right)$ for $(\bar{s}; \Phi)$ in (4) and (5). It is to be, however, noted that with risk aversion, 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.

3.2 Optimal competitive contract

To motivate competitive banking, assume that in principle, a bank is allowed to operate beyond its region. But that entails a fixed cost. If this cost goes to zero, then the limit case is perfect competition for the banking industry, where each bank offers contracts that maximize the expected return to the firms in its region such that the bank itself at least earns the riskless return on its own funds. We focus on this limit situation, and state formally the optimal competitive contract for solving the problem below. To simplify notations, the dependence of the Γ and Ψ functions on Φ^l and Φ^r are omitted.

Problem 1

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

subject to

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

$$\begin{aligned}
&\left[1-\Phi^r\left(\bar{\theta}\right)\right] U\left(\frac{P_1}{P_2} F(k,l)\,\Psi^b(\bar{\omega})\,\bar{\theta} + \frac{R\left(M^d - N^d\right)}{P_2}\right) \\
&\quad + \int_0^{\bar{\theta}} U\left(\frac{P_1}{P_2} F(k,l)\,\Psi^b(\bar{\omega})\,\theta\,(1-\mu) + \frac{R\left(M^d - N^d\right)}{P_2}\right) \mathrm{d}\Phi^r(\theta) \\
&\geq U\left(\frac{RM^d}{P_2}\right)
\end{aligned} \tag{7}$$

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

where $U(\cdot)$ is logarithmic and $0 \leq N^d \leq M^d$. Here, $P_1 F(k,l)\,\Gamma^f(\bar{\omega})$ is the expected return to the firm, unconditional on θ, from the contract in period 1. Dividing this by the period-2 price level P_2 yields the entrepreneur'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 capital (in the financial sense) in riskless securities. 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 depositor may choose to allocate her after-transfer money balance M^d between bank deposits N^d and investment $\left(M^d - N^d\right)$ 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 $R\left(M^d - N^d\right)$ from the risk-free investment. When $\theta \geq \bar{\theta}$, which occurs with probability $1 - \Phi^r\left(\bar{\theta}\right)$, the depositor receives fixed payment $P_1 F\left(k, l\right) \Psi^b\left(\bar{\omega}\right) \bar{\theta}$ from the deposit contract and utility level $U\left(P_1 F\left(k, l\right) \Psi^b\left(\bar{\omega}\right) \bar{\theta} / P_2 + R\left(M^d - N^d\right) / P_2\right)$ from period 2 consumption. When $\theta < \bar{\theta}$, the depositor receives $P_1 F\left(k, l\right) \Psi^b\left(\bar{\omega}\right) \theta\left(1 - \mu\right)$, net of monitoring costs, from the deposit contract and utility level $U\left(P_1 F\left(k, l\right) \Psi^b\left(\bar{\omega}\right) \theta\left(1 - \mu\right) / P_2 + R\left(M^d - N^d\right) / P_2\right)$. The expected utility from the portfolio $\left(N^d, M^d - N^d\right)$ must be no less than putting all of M^d into the risk-free bond, which yields expected utility $U\left(RM^d / P_2\right)$. It should be noted that implicit in (7) is the assumption that each depositor contracts with only one bank and that there is no risk sharing among the depositors.[20]

Finally, inequality (8) is the flow-of-funds constraint for the firms. The total bill for the firms' factor inputs is $R^k k + Wl$, 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

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

As shown in the Appendix, with $U\left(\cdot\right)$ taking the log form, 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 $R\left(M^d - N^d\right) / P_2$.

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

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

$$\tilde{\Psi}^d\left(\bar{\theta}\right) - \log \Gamma^b\left(\bar{\theta}\right) = \log\left(\zeta^b\right), \tag{11}$$

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

[20] Zeng (2007) endogenizes asset indivisibility by explicitly modeling financial transaction costs that prevent the depositors from perfectly diversifying their portfolios. Kilenthong (2011) considers imperfect risk sharing with limited collateral.

where

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

$$\tilde{\Psi}^{d}\left(\bar{\theta}\right)\equiv\left[1-\Phi^{r}\left(\bar{\theta}\right)\right]\log\left(\bar{\theta}\right)+\int_{0}^{\bar{\theta}}\log\left[\theta\left(1-\mu\right)\right]\phi^{r}\left(\theta\right)\mathrm{d}\theta, \tag{14}$$

$$\Delta\left(\bar{\theta}\right)\equiv\left[1-\Phi^{r}\left(\bar{\theta}\right)\right]\frac{1}{\bar{\theta}}+\int_{0}^{\bar{\theta}}\frac{1}{\theta\left(1-\mu\right)}\phi^{r}\left(\theta\right)\mathrm{d}\theta. \tag{15}$$

Equations (9) and (10) are the first-order conditions for factor demand, where the presence of the term q creates wedges between the marginal products of factor inputs and their real prices. We shall call q the *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 costs 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 and reflects the agency costs 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 introduced.[21] It should be noted that q is an increasing function of $\bar{\omega}$ and $\bar{\theta}$, with $\lim_{\bar{\omega},\bar{\theta}\to 0} q = 1$.

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 $\tilde{\Psi}^{d}\left(\bar{\theta}\right) - \log\Gamma^{b}\left(\bar{\theta}\right)$ is increasing in $\bar{\theta}$, Eq. (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.

4 General equilibrium and the effects of bank recapitalization

In this section, we characterize the competitive equilibrium of the model economy and analyze how the bank recapitalization policy, taking the form of central bank

[21] Another type of distortions is present with $R > 1$, which creates additional wedges between the marginal products and real prices of factor inputs. This type of distortions arise from the requirement that factor market transactions must use cash and can thus be named "monetary frictions." The assumption that R is pegged by the government at a fixed value implies that the extent of such frictions is unaffected by the bank recapitalization policy, which allows us to focus on the effect of the policy on the extent of financial frictions. Arseneau (forthcoming) considers monetary distortions in a new Keynesian open economy setup.

money injection into the banking system, affects the real economy. The optimal policy responses to shocks to the riskiness of banking will also be considered.

4.1 The non-neutrality of money

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

$$\zeta^b = \frac{(1-\alpha) z^d}{\alpha K^b/K + (1-\alpha) z^b}, \quad \frac{1}{1+\zeta^b} \frac{\zeta^f}{1+\zeta^f} = \alpha \frac{K^b}{K} + (1-\alpha) z^b, \tag{16}$$

where

$$z^b \equiv \frac{M^b}{M^b + M^d} \quad \text{and} \quad z^d \equiv \frac{M^d}{M^b + M^d}$$

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

In light of (16), Eqs. (11) and (12) become

$$\tilde{\Psi}^d\left(\bar{\theta}\right) - \log \Gamma^b\left(\bar{\theta}\right) = \log\left(\frac{(1-\alpha) z^d}{\alpha K^b/K + (1-\alpha) z^b}\right), \tag{17}$$

$$q\left(\bar{\omega}, \bar{\theta}\right) \Psi^b\left(\bar{\omega}\right) \Gamma^b\left(\bar{\theta}\right) = \alpha \frac{K^b}{K} + (1-\alpha) z^b. \tag{18}$$

Thus, $\left(z^b, z^d\right)$ determines the bank default threshold $\bar{\theta}$ via (17), and in addition, the firm default threshold $\bar{\omega}$ via (18). Given $\bar{\theta}$, $\bar{\omega}$ (hence q), the real wage rate ν, and the pegged risk-free interest rate R, the equilibrium employment L is determined by the following condition

$$F_l(K, L) = q\left(\bar{\omega}, \bar{\theta}\right) R\nu. \tag{19}$$

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

$$C_1 = F(K, L)\, \varphi\left(\bar{\omega}, \bar{\theta}\right), \tag{20}$$

where

$$\varphi\left(\bar{\omega}, \bar{\theta}\right) \equiv \Gamma^f\left(\bar{\omega}\right) + \Psi^b\left(\bar{\omega}\right)\left[\Gamma^b\left(\bar{\theta}\right) + \Psi^d\left(\bar{\theta}\right)\right]. \tag{21}$$

It should be noted 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 monetary policy in our model 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 the policy that is relevant for allocations is simply the distribution of $(M-M^{w})$ 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 Eq. (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

$$M^{w} = M - \nu P_1 L \tag{22}$$

must hold for the specification of policy to be internally consistent. Equation (22) can be seen as a rule that the government uses to determine M^{w} for any given specification of $\left(z^{b}, z^{d}, M\right)$. The policy vector can, thus, be equivalently represented by $\left(z^{b}, z^{d}, M, M^{w}\right)$, where M^{w} is given by (22).[22]

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 period-1 revenues that goes to that type of agents as dictated by the 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

$$C_2^{f} = C_2 \frac{\Gamma^{f}\left(\bar{\omega}\right)}{\varphi\left(\bar{\omega}, \bar{\theta}\right)}, \quad C_2^{b} = C_2 \frac{\Psi^{b}\left(\bar{\omega}\right)\Gamma^{b}\left(\bar{\theta}\right)}{\varphi\left(\bar{\omega}, \bar{\theta}\right)}, \quad C_2^{d} = C_2 \frac{\Psi^{b}\left(\bar{\omega}\right)\Psi^{d}\left(\bar{\theta}\right)}{\varphi\left(\bar{\omega}, \bar{\theta}\right)}, \tag{23}$$

respectively.[23]

Formally, a competitive equilibrium with banking frictions and two-sided financial contracting is defined as follows.

Definition 1 A *competitive equilibrium* of the model economy is a policy $\left(z^{b}, z^{d}, M, M^{w}\right)$, an allocation $\left(L, C_1, C_2^{f}, C_2^{b}, C_2^{d}\right)$, a price system $\left(P_1, W, R^{k}, P_2\right)$, and terms of financial contract $\left(\bar{\omega}, \bar{\theta}\right)$ such that

(i) The period-1 contract terms and allocations, $\bar{\omega}$, $\bar{\theta}$, L, and C_1, are determined by (17)–(20) given the allocation-relevant policy $\left(z^{b}, z^{d}\right)$.
(ii) 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 $R^{k} = \alpha WL/(1-\alpha)$.
(iii) the period-2 consumption allocation $\left(C_2^{f}, C_2^{b}, C_2^{d}\right)$ is given by (23).
(iv) M^{w} is set in accordance with the rule (22) for any given specification of $\left(z^{b}, z^{d}, M\right)$.

[22] The need to specify the rule (22) is a special feature of the current two-period setup, and is absent in the infinite-horizon model of Zeng (2010).

[23] Note that C_2^{d} is not only the total consumption of all depositors but also the expected consumption of an individual depositor, which, of course, differs from her expected utility due to risk aversion.

For analytic purpose, it will be especially convenient to look at the behavior of the model economy around a situation where no default by either the banks or the firms occurs. We define such a situation as follows.

Definition 2 A *zero-default equilibrium* is the competitive equilibrium of the model economy obtained when the distributions for θ and ω are degenerate.

Essentially, the asymmetric information problems disappear when θ and ω are non-stochastic, giving rise to zero default in equilibrium. Proof of the existence and uniqueness of the zero-default equilibrium is trivial. Our analysis will focus on the neighborhood of this zero-default equilibrium, where default occurs with small probabilities. According to Fisher (1999), the historical average of bankruptcy rate is indeed quite small. This does not, however, mean that the distortions caused by financial frictions are negligible. The following proposition establishes the existence and uniqueness of equilibrium as well as the real effect of the bank recapitalization policy.

Proposition 1 *A competitive equilibrium of the model economy with banking frictions and two-sided financial contracting exists and is unique in the neighborhood of zero default for any* $z^b \in (0, 1)$. *In such a neighborhood, an increase in* z^b *raises equilibrium employment for any* $z^b \in (0, 1)$.

Money is, therefore, non-neutral if and only if it is associated with a change in the distribution of money balances between the banks and depositors. Taking the initial distribution of money balances as given, any monetary transfer that leads to a change in $\left(z^b, z^d\right)$ has real impact. The general mechanism is as follows. An increase in z^b (with a corresponding decrease in z^b) lowers the banks' debt–equity ratio ζ^b. This reduces the bank default threshold $\bar{\theta}$ via (17), consistent with the depositors' receiving a smaller share of revenues relative to the banks, as dictated by the bank-depositor contract. No matter what happens to the firm default threshold $\bar{\omega}$, the financial friction indicator q must take a smaller value, as shown in the Appendix. The overall reduction in the two-sided financial frictions generates an increase in equilibrium employment according to (19).

Consider the scenario when bank recapitalization, as an implicit part of the unconventional monetary policy undertaken during the GFC, is implemented. Increases in z^b are not produced by directly taxing the depositors and transferring the proceeds to the banks. Rather, they are produced by injecting newly created money to the banks. The resultant increases in the aggregate amount of money balances are non-neutral, as they are associated with increases in z^b. They stimulate employment by lowering bank leverage and the overall extent of financial distortions. This non-neutrality result depends crucially on the presence of banking frictions, i.e., frictions on the liability side of the bank balance sheet due to the informational asymmetry in the bank–depositor relationship. Without such frictions, a neutrality result will obtain regardless of the value of $\left(z^b, z^d\right)$.[24] These conclusions hold even with the presence of credit frictions,

[24] To see this we can take away banking frictions from the model simply by assuming that the distribution of the region specific productivity is degenerate. It is straightforward to show that in such an environment the recapitalization policy is irrelevant for employment, output, real factor prices, and the firms' default rate.

i.e., frictions on the asset side of the bank balance sheet due to the informational asymmetry in the bank–firm relationship. It is therefore precisely the presence of banking frictions (and the fact that banks are the institutions being recapitalized) that is responsible for the potency of the recapitalization policy.

4.2 Banking riskiness and optimal stabilization policy

Our analysis thus far has treated the bank recapitalization policy as exogenous. In this section, we investigate how the policy can be used as a stabilization tool when there are shocks to the "riskiness" of banking. To introduce the concept of banking riskiness, 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. The distribution is completed by assigning a zero p.d.f. for $\theta = 0$. In our model, it is the costly verification of θ that gives rise to the bankruptcy of banks. The default rate of banks tends to zero, as σ_θ tends to zero from the right. Therefore, 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

$$\sigma_\theta = \bar{\sigma}_\theta + \varepsilon, \tag{24}$$

where $\bar{\sigma}_\theta > 0$, representing the mean level of riskiness, is a positive constant, and ε is a random disturbance bounded away from $-\bar{\sigma}_\theta$. We interpret ε as the banking riskiness shock.[25]

In our view, shocks to banking riskiness are highly relevant in the light of the erratic behavior of the risk spreads for banks' external finance. The historical average of the spread between the 3 month certificate of deposits (CD) rate and the 3 month T-bill rate is about 75 basis points (per annum), based on a sample from 1973Q1 to 2009Q4. From 2001Q1 to 2007Q2, the spread averages only 27 basis points. In contrast, its average in the second half of 2007 and the year of 2008 rises 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 level 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, Eq. (6), in Problem 1, we obtain the model's bank risk spread: $R^b - R = R\left\{\bar{\theta}/\left[\Gamma^b\left(\bar{\theta}\right)\zeta^b\right] - 1\right\}$. Other things equal, an increase in σ_θ raises the bank default rate, and hence the bank risk spread. Fluctuations in banking riskiness, thus, give rise to fluctuations in the bank risk spread.

Let the bank recapitalization policy takes the form $z^b = \eta + x$, where η is a positive constant and x is a random component bounded away from $-\eta$. It is easy to see that when the policy is specified as such and the banking riskiness σ_θ is as specified in (24),

[25] Our formulation of banking riskiness shocks parallels the formulation of entrepreneurial riskiness shocks in Williamson (1987) and Christiano et al. (2003, 2009), who consider the costly state verification problem between financial intermediaries and nonfinancial firms.

the existence and uniqueness results for the competitive equilibrium, as established in Proposition 1, remain valid. The efficacy of the recapitalization policy applies as well. This enables the policy to serve as a stabilization tool in the face of banking riskiness shocks. Taking the mean recapitalization η and the mean riskiness $\bar{\sigma}_\theta$ as given, we aim to analyze how the recapitalization policy can be used to buffer the economy from the disturbance ε to banking riskiness.[26] We shall see that stabilization considerations give rise to a particular kind of policy reaction function, or policy rule, which dictates how x should respond to ε in a systematic fashion.

Suppose that the goal of the stabilization policy is to insulate employment L and net output C_1 from the banking riskiness shock. This would require both the financial friction indicator q and the net output factor φ to be completely stabilized, which is impossible, since we would have two targets and only one policy instrument. However, as shown in Zeng (2010), compared to q, the variable φ is only of second-order importance in the neighborhood of zero default. Hence, an approximately optimal policy needs only seek to stabilize the financial friction indicator q. Our numerical result, to be presented momentarily, shows that targeting q alone actually achieves near-complete stabilization of both q and φ, and hence near-complete stabilization of C_1. Stabilization of q also turns out to insulate the total period-2 consumption of the depositors, the risk-averse agents in the economy, from the banking riskiness shock almost perfectly. This is because the consumption share of these agents, as in (23), depends on $\bar{\omega}$ and $\bar{\theta}$ in a way similar to the dependence of q on these default thresholds.[27]

Denote the value of q that would prevail without any shock by q^*. For the (approximately) optimal stabilization policy, q^* serves as the target. In order to derive the optimal policy reaction function, denote the mapping of (ε, x) to q by $\mathbf{q}(\varepsilon, x)$: the realization of ε gives the value of σ_θ, which, together with x, determines $\left(\bar{\omega}, \bar{\theta}\right)$ and hence q through (17) and (18). Given ε, targeting q at q^* amounts to setting x at x^*, where x^* satisfies $\mathbf{q}(\varepsilon, x^*) = q^*$. To keep the value of q at q^*, an increase in ε calls for a larger value of x^* to offset the effect of the increased banking riskiness. Hence, x^* varies positively with ε, with $x^* = 0$ when $\varepsilon = 0$. Such a reaction function entails recapitalization efforts that counteract banking riskiness: there is more (less) nominal capital transfer to the banks when banking becomes more (less) risky.

To demonstrate numerically the optimal reaction function, we calibrate the model economy as follows. Let a time period correspond to a quarter. We set $R = 1.014$, consistent with the historical average of the 3 month T-bill rate.[28] The weight of leisure

[26] Zeng (2010) discusses the optimal choice of η.

[27] There is an important distinction between individual depositor consumption and total depositor consumption, the former subject to idiosyncratic bank risks (the θ shocks), the latter immune to idiosyncratic risks but subject to system risks (the $\bar{\theta}$ shocks) and hence the banking riskiness shock. The banking riskiness shock is an economy-wide shock since all individual banks face the common level of riskiness. The effect of this shock is offset by the stabilization policy.

[28] We do not pursue the possibility of reducing R to near the zero lower bound for the risk-free nominal interest rate.

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 et al. 1999).[29] Normalizing $K = 1$ and $K^b = 0$, the value of η is set to be 0.076, which matches the historical average of an asset–net worth ratio of 13.18 for US commercial banks.[30] The monitoring cost parameter, μ, is set to be 0.36.[31] Similar to the bank/region-specific productivity, we assume that the firm/location-specific productivity ω follows a unit mean log-normal distribution on $(0, \infty)$, completed with the assignment of a zero p.d.f. for $\omega = 0$. For $\omega > 0$, $\log(\omega) \sim \mathcal{N}\left(-\frac{1}{2}\sigma_\omega^2, \sigma_\omega^2\right)$. We assume that σ_ω is fixed, while σ_θ follows the specification in (24). The value of σ_ω and the mean value of σ_θ, i.e., $\bar{\sigma}_\theta$, are chosen to match (1) a spread between the firms' borrowing rate and the risk-free rate of 293 basis points per annum, and (2) a spread between the banks' borrowing rate and the risk-free rate of 75 basis points per annum.[32]

Figure 2 depicts the optimal recapitalization policy in relation to the level of banking riskiness. The middle and bottom panels show the effects of the riskiness shock ε on the financial friction indicator q and employment L. The dashed lines correspond to the case where there is no policy reaction to the shocks, i.e., x equals zero identically. The solid lines correspond to the case where the policy reacts in the optimal fashion described above. As can be seen from the no-reaction lines, the effect of a positive (resp. negative) shock to banking riskiness is to raise (resp. lower) q and reduce (resp. increase) L. The effects are asymmetric 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. The non-linearity is also evident from the larger marginal effects of positive shocks (the dashed lines are steeper to the right of $\varepsilon = 0$).[33] By reacting to the banking riskiness shocks in the optimal fashion, the bank recapitalization policy stabilizes employment, as shown by the solid lines. The optimal reaction function is plotted on the top panel, where x^* turns out be an increasing, approximately linear function of ε.[34]

[29] 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.

[30] 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.

[31] 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.

[32] 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 Federal Reserve. The sample period is again 1973Q1–2009Q4.

[33] Krasa et al. (2008) analyze agents' incentives to default and show that the enforcement parameters in their model can generate a sharply nonlinear effect on firm finance.

[34] The approximate linearity obtains since the marginal employment effect of x is also weaker when the marginal employment effect of ε is weaker, i.e., when banking is less risky.

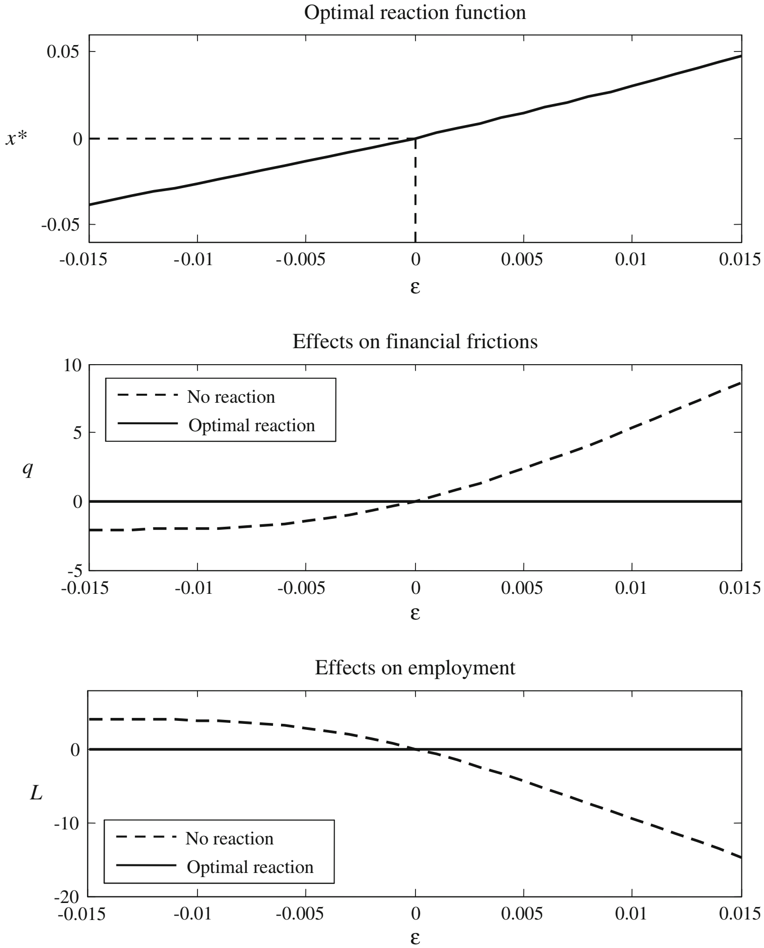

Fig. 2 Banking riskiness shock and the optimal stabilization policy

5 Conclusions

This paper develops a general equilibrium framework with banking frictions and two-sided financial contracting. The framework is used to analyze the effects of bank recapitalization, taking the form of nominal capital transfers to the banking system. The design of optimal stabilization policy, in relation to the riskiness of banking, is also investigated. The paper contributes to understanding the transmission mechanisms of the unconventional monetary policy measures adopted during the GFC and to understanding how policy should be designed to mitigate the adverse effects of financial shocks.

Although our study has mainly concerned the effects of bank recapitalization by the monetary authority and the analysis has been carried out in a highly stylized model, the theoretical framework can be extended to study a wide spectrum of issues related to policy and regulation, as well as the monetary transmission mechanism, in perhaps more realistic ways. First, nominal rigidities and richer dynamics, such as capital accumulation, can be introduced to allow for a quantitative assessment of the effects of policy. Second, deposit insurance can be incorporated in order to study the effects of raising the limit of deposit insurance, as was implemented in the United States in 2008.[35] Third, one can consider situations where some sort of capital adequacy requirements bind. In those situations, bank recapitalization policy may work through relaxing these constraints. Fourth, the model can be extended to allow changes in asset prices to affect the net worth of banks (and firms), as in Bernanke et al. (1999) and Christiano et al. (2003, 2009). Finally, our analysis can be extended to include credit rationing as a possible equilibrium outcome as in Williamson (1986), so that another dimension in which policy exerts influence on the economy can be explored.[36] We conclude that thorough analysis of frictions in the banking sector should be an integral part of future research on the interaction of money, finance, and the macroeconomy.

Appendix

Derivation of the optimality conditions for Problem 1

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 $U(\cdot)$ taking the log-form, the first-order conditions with respect to $\bar{\omega}$ and $\bar{\theta}$ are

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

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

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

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

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

[35] Introducing deposit insurance will not invalidate the analysis in the present paper since a large fraction of bank liabilities remain uninsured. See Kashyap and Stein (2000).

[36] In our setup one can imagine two possible types of credit rationing. The first is rationing on the banks' asset side, where firms are unable to obtain the bank loans they desire. This type of credit rationing has been extensively studied in the literature (e.g., Stiglitz and Weiss 1981 and Williamson 1986. The second type is rationing on the banks' liability side, where banks are unable to raise the loanable funds they desire. The latter type of credit rationing is an interesting aspect to explore in future research.

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

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

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).

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

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

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

Substituting (A.5) into (A.6) yields (11).

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

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

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 $q\left(\bar{\omega},\bar{\theta}\right) \equiv \varrho(\bar{\omega})\,\varkappa\left(\bar{\theta}\right)$, where

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

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

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

But

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

To sign the above expression, we consider two cases. Case 1: $\lim_{\bar{\omega}\to 0}\phi^{l}(\bar{\omega}) > 0$. In this case, $\lim_{\bar{\omega}\to 0}\left[\Psi^{b'}(\bar{\omega})\Gamma^{f''}(\bar{\omega}) - \Psi^{b''}(\bar{\omega})\Gamma^{f'}(\bar{\omega})\right] = -\mu\lim_{\bar{\omega}\to 0}\phi^{l}(\bar{\omega}) < 0$. Case 2: $\lim_{\bar{\omega}\to 0}\phi^{l}(\bar{\omega}) = 0$. But Assumption 1(a) requires $\phi^{l}(\cdot)$ to be positive, bounded, and continuously differentiable on $(0,\infty)$. Hence, in this case, we must have $\lim_{\bar{\omega}\to 0}\phi^{l'}(\bar{\omega}) > 0$. This means that for $\bar{\omega}$ positive and sufficiently close to 0, we have $\phi^{l}(\bar{\omega}) > 0$ and $\phi^{l'}(\bar{\omega}) > 0$, and hence $\left[\Psi^{b'}(\bar{\omega})\Gamma^{f''}(\bar{\omega}) - \Psi^{b''}(\bar{\omega})\Gamma^{f'}(\bar{\omega})\right] < 0$. 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 (14) and (15), we have

$$\left[\varkappa\left(\bar{\theta}\right)\right]^{-1} = \frac{A\left(\bar{\theta}\right) - B\left(\bar{\theta}\right)}{D\left(\bar{\theta}\right)},$$

where

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

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

$$D\left(\bar{\theta}\right) \equiv \left[1 - \Phi^{r}\left(\bar{\theta}\right)\right] + \bar{\theta}\int_{0}^{\bar{\theta}}\frac{1}{\theta\left(1-\mu\right)}\phi^{r}\left(\theta\right)\mathrm{d}\theta.$$

By differentiation,

$$\begin{aligned}\frac{\mathrm{d}A\left(\bar{\theta}\right)}{\mathrm{d}\bar{\theta}} &= \frac{d}{\mathrm{d}\bar{\theta}}\left\{\left[\int_{\bar{\theta}}^{\infty}\left(\theta-\bar{\theta}\right)\phi^{r}\left(\theta\right)\mathrm{d}\theta\right]\left[1 - \mu\frac{\bar{\theta}\phi^{r}\left(\bar{\theta}\right)}{1-\Phi^{r}\left(\bar{\theta}\right)}\right] + \bar{\theta}\right\}\\
&= -\mu\left[E\left(\theta|\theta\geq\bar{\theta}\right) - 2\bar{\theta}\right]\phi^{r}\left(\bar{\theta}\right)\\
&\quad -\mu\left[E\left(\theta|\theta\geq\bar{\theta}\right) - \bar{\theta}\right]\left[\bar{\theta}\phi^{r'}\left(\bar{\theta}\right) + \frac{\bar{\theta}\left[\phi^{r}\left(\bar{\theta}\right)\right]^{2}}{1-\Phi^{r}\left(\bar{\theta}\right)}\right] + \Phi^{r}\left(\bar{\theta}\right),\end{aligned}$$

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, $\lim_{\bar{\theta}\to 0}\mathrm{d}A\left(\bar{\theta}\right)/\mathrm{d}\theta < 0$. Case 2. $\lim_{\bar{\theta}\to 0}\phi^{r}\left(\bar{\theta}\right) = 0$. In this case, $\lim_{\bar{\theta}\to 0}\phi^{r'}\left(\bar{\theta}\right) > 0$ as implied by Assumption 1(a), which requires $\phi^{r}(\cdot)$ 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 $\mathrm{d}A\left(\bar{\theta}\right)/\mathrm{d}\theta < 0$. In sum, $\mathrm{d}A\left(\bar{\theta}\right)/\mathrm{d}\theta < 0$ in the neighborhood of

$\bar{\theta} = 0$. Also,

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

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 $\lim_{\bar{\theta}\to 0} dB\left(\bar{\theta}\right)/\mathrm{d}\bar{\theta} > 0$ (note that $-\log(1-\mu)-\mu > 0$). Case 2. $\lim_{\bar{\theta}\to 0}\phi^r\left(\bar{\theta}\right) = 0$. In this case $\lim_{\bar{\theta}\to 0}\phi^{r'}\left(\bar{\theta}\right) > 0$ 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 $\phi^{r'}\left(\bar{\theta}\right)$ are positive, hence $dB\left(\bar{\theta}\right)/\mathrm{d}\bar{\theta} > 0$. In sum, $dB\left(\bar{\theta}\right)/\mathrm{d}\bar{\theta} > 0$ in the neighborhood of $\bar{\theta} = 0$. Finally,

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

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

Proof of Proposition 1 To prove the existence and uniqueness of equilibrium, first note that a solution to (17) for $\bar{\theta}$ exists and is unique. This is because $\left[\tilde{\Psi}^d\left(\bar{\theta}\right) - \log\Gamma^b\left(\bar{\theta}\right)\right]$ is increasing in $\bar{\theta}$, with $\lim_{\bar{\theta}\to 0}\left[\tilde{\Psi}^d\left(\bar{\theta}\right) - \log\Gamma^b\left(\bar{\theta}\right)\right] = -\infty$ and $\lim_{\bar{\theta}\to\infty}\left[\tilde{\Psi}^d\left(\bar{\theta}\right) - \log\Gamma^b\left(\bar{\theta}\right)\right] = \infty$. Given $\bar{\theta}$, a solution to (18) for $\bar{\omega}$ also exists and is unique. This is because both $q\left(\bar{\omega},\bar{\theta}\right)$ and $\Psi^b\left(\bar{\omega}\right)$ are increasing in $\bar{\omega}$ in the neighborhood of $\bar{\omega} = 0$ and $\lim_{\bar{\omega}\to 0} q\left(\bar{\omega},\bar{\theta}\right)\Psi^b\left(\bar{\omega}\right) = 0$. Given $\bar{\theta}$ and $\bar{\omega}$, it remains to show that a solution to condition (19) exists and is unique. This is because $F_l(K, L)$ is monotonically decreasing in L, with $\lim_{L\to 0} F_l(K, L) = \infty$ and $\lim_{L\to\infty} F_l(K, L) = 0$ and $q\left(\bar{\omega},\bar{\theta}\right)R\nu > 0$ is independent of L.

To see how an increase in z^b affects L, note that from (17), an increase in z^b (with a corresponding decrease in z^d) lowers $\bar{\theta}$. As for the change in $\bar{\omega}$, there are two cases. Case 1. $\bar{\omega}$ does not increase. In this case, $q\left(\bar{\omega},\bar{\theta}\right)$ obviously decreases, since q is increasing in both arguments in the neighborhood of $\bar{\omega}, \bar{\theta} = 0$. Case 2. $\bar{\omega}$ increases. In this case, condition (18) implies that q must decrease, since $\Gamma^b\left(\bar{\theta}\right)$ increases as a result of the decrease in $\bar{\theta}$, $\Psi^b\left(\bar{\omega}\right)$ increases as a result of the increase in $\bar{\omega}$, and the right hand side of this condition decreases as a result of the decrease in z^b. Thus, in both cases, q declines. Condition (19) then implies an increase in L. *Q.E.D.*

References

Alderson, M., Betker, B.: Liquidation costs and capital structure. J Financ Econ **39**, 45–69 (1995)
Arseneau, D.: Expectation traps in a new Keynesian open economy model. Econ Theory (forthcoming)

Bernanke, B., Gertler, M.: Agency costs, net worth, and business fluctuations. Am Econ Rev **79**, 14–31 (1989)
Bernanke, B., Gertler, M., Gilchrist, S.: The financial accelerator in a quantitative business cycle framework. In: Taylor, J., Woodford, M. (eds.) Handbook of Macroeconomics vol. 1, Amsterdam: Elsevier (1999)
Boyd, J., Smith, B.: How good are standard debt contracts? stochastic versus nonstochastic monitoring in a costly state verification environment. J Bus **67**, 539–561 (1994)
Brunnermeier, M.: Deciphering the liquidity and credit crunch 2007–2008. J Econ Perspect **23**, 77–100 (2009)
Carlstrom, C., Fuerst, T.: Agency costs, net worth, and business fluctuations. Am Econ Rev **87**, 893–910 (1997)
Cecchetti, S.: Crisis and responses: the federal reserve in the early stages of the financial crisis. J Econ Perspect **23**, 51–75 (2009)
Cerasi, V., Daltung, S.: The optimal size of a bank: costs and benefits of diversification. Eur Econ Rev **44**, 1701–1726 (2000)
Christiano, L., Eichenbaum, M.: Liquidity effects and the monetary transmission mechanism. Am Econ Rev Papers Proc **82**, 346–353 (1992)
Christiano, L., Motto, R., Rostagno, M.: The great depression and the Friedman-Schwartz hypothesis. J Money Credit Bank **35**, 1119–1197 (2003)
Christiano, L., Motto, R., Rostagno, M.: Financial factors in economic fluctuations. Working paper (2009)
Cúrdia, V., Woodford, M.: The central-bank balance sheet as an instrument of monetary policy. J Monet Econ **58**, 54–79 (2011)
Diamond, D.: Financial intermediation and delegated monitoring. Rev Econ Stud LI, 393–414 (1984)
Diamond, D., Rajan, R.: Money in a theory of banking. Am Econ Rev **96**, 30–53 (2006)
Fisher, J.: Credit market imperfection and the heterogeneous response of firms to monetary shocks. J Money Credit Bank **31**, 187–211 (1999)
Gale, D., Hellwig, M.: Incentive compatible debt contracts I: the one period problem. Rev Econ Stud **52**, 647–664 (1985)
Gertler, M., Karadi, P.: A model of unconventional monetary policy. J Monet Econ **58**, 17–34 (2011)
Goodfriend, M.: Central banking in the credit turmoil: an assessment of federal reserve practice. J Monet Econ **58**, 1–12 (2011)
Hellwig, M.: Financial intermediation with risk aversion. Rev Econ Stud **67**, 719–742 (2000)
Kashyap, A., Stein, J.: What do a million observations on banks say about the transmission of monetary policy? Am Econ Rev **90**, 407–428 (2000)
Khalil, F., Martimort, D., Parigi, B.: Monitoring a common agent: implications for financial contracting. J Econ Theory **135**, 35–67 (2007)
Kilenthong, W.: Collateral premia and risk sharing under limited commitment. Econ Theory **46**, 475–501 (2011)
Krasa, S., Villamil, A.: Monitoring the monitor: an incentive structure for a financial intermediary. J Econ Theory **57**, 197–221 (1992a)
Krasa, S., Villamil, A.: A theory of optimal bank size. Oxf Econ Pap **44**, 725–749 (1992b)
Krasa, S., Villamil, A.: Optimal multilateral contracts. Econ Theory **4**, 167–187 (1994)
Krasa, S., Villamil, A.: Optimal contract when enforcement is a decision variable. Econometrica **68**, 119–134 (2000)
Krasa, S., Sharma, T., Villamil, A.: Bankruptcy and firm finance. Econ Theory **36**, 239–266 (2008)
Lucas, R.: Expectations and the neutrality of money. J Econ Theory **4**, 103–124 (1972)
Modigliani, F., Miller, M.: The cost of capital, corporation finance and the theory of investment. Am Econ Rev **48**, 261–297 (1958)
Mookherjee, D., Png, I.: Optimal auditing, insurance, and redistribution. Q J Econ **104**, 399–415 (1989)
Reis, R.: Interpreting the unconventional U.S. monetary policy of 2007–09. Brookings Pap Econ Act **2**, 119–165 (2009)
Stiglitz, J., Weiss, A.: Credit rationing in markets with imperfect information. Am Econ Rev **70**, 393–410 (1981)
Tett, G.: US banks quietly borrow $50bn from Fed via new credit facility. Financial Times, February 19, p. 1 (2008)
Townsend, R.: Optimal contracts and competitive markets with costly state verification. J Econ Theory **21**, 265–293 (1979)

Williamson, S.: Costly monitoring, financial intermediation, and equilibrium credit rationing. J Monet Econ **18**, 159–179 (1986)

Williamson, S.: Financial intermediation, business failures, and real business cycles. J Poli Econ **95**, 1196–1216 (1987)

Williamson, S.: Restrictions on financial intermediaries and implications for aggregate fluctuations: Canada and the United States, 1870–1913. In: Blanchard, O., Fischer, S. (eds.) NBER Macroeconomics Annual (1989)

Zeng, Z.: Banking frictions and monetary policy. In: Financial Markets and Monetary Policy, Ph.D. Dissertation, Northwestern University (2002)

Zeng, Z.: The price of size and financial market allocations. Econ Theory **30**, 21–48 (2007)

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