

Lending Efficiency Shocks

Kaiji Chen*
Emory University

Tao Zha
FRB Atlanta, Emory Univ. and NBER

June, 2013 (preliminary and incomplete)

Abstract

This paper develops a theory in which shocks to the efficiency of financial intermediation translate into business cycle fluctuations via capital reallocation. In our theory, under costly verification, the bank chooses to only monitor the returns of those entrepreneurs with insufficient net worth. This distorts the existing capital allocation among entrepreneurs of different sizes. A crucial ingredient of the model is that the outcome of monitoring is random and depends on both the efficiency of monitoring and the resources devoted to policing the returns of a project. As a consequence, a negative shock to monitoring efficiency forces bank to increase monitoring intensity and reduce the loan toward small entrepreneurs. This results in an increase in productivity dispersion and a recession. Using the COMPUSTAT dataset, we find a significant countercyclical pattern for the relative capital productivity of financially constrained to unconstrained firms, and a procyclical capital allocation between these two types of firms. Such an empirical observation distinguishes the lending efficiency shocks from other aggregate shocks as the sources of business cycles.

JEL Classification: E32, G34

Keywords: Lending efficiency shocks, financial frictions, capital reallocation, TFP fluctuations.

*Department of Economics, Emory University Atlanta, GA 30322. Phone: (404) 727-2944. Email: kaiji.chen@emory.edu.

1 Introduction

The recent financial crisis suggests that shocks to financial intermediation might be an important driver of economic fluctuations. In particular, a disturbance in the efficiency of information acquisition by banks, say under a volatile economic environment, may force them to increase their lending standards toward firms. The mechanism through which such disturbances are transmitted into business cycles is, however, an open issue. One potential channel is resource reallocation across firms of different sizes. Intuitively, firms with different sizes might be subject to different degrees of information asymmetry and, thus, financial constraints. Such an asymmetry in financing constraints, moreover, is found to be crucial in driving productivity dispersion at the firm level and its fluctuations over business cycles.¹ These observations indicate that variations in financial frictions, by reallocating capital at the disaggregate level, might be important for shocks to financial intermediation to translate into aggregate TFP fluctuations and, thus, business cycles.

This paper develops a theory for disturbances in the efficiency of financial intermediation to translate into aggregate TFP fluctuations via capital reallocation. In our theory, due to costly verification, the bank chooses to only monitor those entrepreneurs with insufficient net worth. A key element is that the outcome of monitoring is random and depends on both the efficiency of monitoring and the resources devoted to policing the returns of a project. As a consequence, a negative shock to monitoring efficiency (“lending efficiency shocks” hereafter) forces bank to increase monitoring intensity and reduce the loan toward small entrepreneurs. This results in an increase in productivity dispersion between entrepreneurs of different sizes and a recession. Such countercyclical productivity dispersion is strongly supported by our evidence using COMPUSTAT data, and helps to distinguish the lending efficiency shocks from other aggregate shocks as the sources of business cycles.

Our main results are driven by two key model ingredients. The first is heterogeneity of entrepreneurs in terms of net worth. Only entrepreneurs with insufficient net worth, who rely heavily on bank loans, are subject to bank monitoring. Accordingly, small entrepreneurs are financially constrained, while large ones are not. The second element in our model is shocks to the efficiency of bank monitoring. Such shocks alter the probability of a bank to detect malfeasance by entrepreneurs and, therefore, the monitoring intensity of a bank. This, in turns, varies the tightness of financial constraints of the constrained entrepreneurs.

¹See Chen and Song (2013).

A combination of these two elements delivers the following mechanism for lending efficiency shocks to translate into TFP fluctuations. A reduction in the efficiency of monitoring triggers bank to monitor more intensively. This incurs a higher monitoring cost and, thus, a larger wedge between the marginal product and the user cost of capital. Accordingly, the amount of bank loan advanced to the constrained entrepreneurs is reduced. Capital, therefore, is reallocated from constrained to unconstrained projects, leading to a widening of productivity dispersion between these two types of projects. This increase in the productivity dispersion, which represents a loss in allocative efficiency, shows up as a fall in aggregate TFP.

Our theory deliver a key testable prediction that distinguishes the lending efficiency shocks from other shocks. That is, the dispersion of marginal product of capital between financially constrained and unconstrained firms is countercyclical. While other shocks, such as aggregate technology shocks and shocks to marginal efficiency of investments, may trigger similar boom-bust cycles in aggregate variables, in our model they all exhibit *procyclical* variations in the dispersion of marginal product of capital. Intuitively, the marginal cost for financially constrained entrepreneurs involves not only the replacement cost, but also the monitoring cost. Hence, given monitoring intensity, unconstrained entrepreneurs' input demand is more sensitive than constrained ones to those shocks that influence the price of inputs. Accordingly, those shocks imply a procyclical ratio of capital deployed in financially constrained firms to that in unconstrained firms, contrast to the predictions by lending efficiency shocks.

To test the above theoretical implication, we use the COMPUSTAT dataset to estimate the ratio of capital productivity between constrained to unconstrained firms. Firms are classified into constrained and unconstrained groups by asset sizes. We find that, on average, the constrained firms are more productive than the unconstrained in terms of revenue-based capital productivity. Moreover, consistent with the model prediction, the relative capital productivity between the two groups has a correlation coefficient with GDP of -0.44 . The key factor underlying this countercyclical capital productivity dispersion is the procyclical capital allocation between small and large firms: the ratio of capital of constrained to unconstrained firms is positively correlated with GDP with a correlation coefficient of 0.55 . Therefore, our empirical evidence supports lending efficiency shocks as important drivers of U.S. business cycles.

Our model is closely related to Greenwood, Sanchez and Wang (2010). In both paper, the probability of detecting malfeasance depends on the efficiency and intensity of bank monitoring. Their paper, however, differs from ours along several dimensions: First, in their paper, entrepreneurs are shut down from savings and their asymmetric access to bank loans stems from idiosyncratic productivity uncertainty. By contrast, entrepreneurs in our model are al-

lowed to save and subject to the same degree of idiosyncratic uncertainty. This results in a completely different mechanism for the heterogeneous access to bank credit by entrepreneurs, i.e. via their net worth. Second, in their paper, resource reallocation and aggregate productive efficiency gain occur along the extensive margins, i.e. via entry and exit, a margin empirically more relevant for aggregate productivity growth over the medium or long run.² By contrast, capital reallocation in our model happens along the intensive margin, i.e. between firms of different sizes. Such a capital reallocation pattern and implied variations in productivity dispersion are in line with our evidence. Finally, Greenwood et. al (2010) explore the role of information production in the financial sector in the context of economic development. Our paper, instead, focus on how disturbances in the efficiency of financial intermediation affect economic fluctuations.

Our work is closely related to the emerging literature on uncertainty-driven business cycles. In particular, following the recent financial crisis, researchers point to financial frictions as a transmission mechanism for uncertainty shocks to trigger the business cycles. The first attempt to link uncertainty and financial frictions is Christiano, Motto and Rostagno (2010). They explore the roles of uncertainty shocks in the framework of Bernanke, Gertler and Gilchrist (1999, “BGG” henceforth). More recently, Gilchrist, Sim and Zakrajšek (2010) extend the BGG framework to allow uncertainty shocks to affect the capital allocation among investment-good production firms, as an amplification mechanism for investment dynamics over business cycles.³ Alternative frictions for uncertainty shocks to transmit into economic fluctuations include firm-level factor adjustment costs (see Bloom, Floetotto and Jiamovich, 2009 and Bachmann and Bayer, 2009) or a combination of entry costs and investment irreversibility (Sim, 2008). A common feature of the above studies is that variations in the dispersion of idiosyncratic productivity or time-series volatility of productivity shocks are treated as exogenous shocks to the economy. By contrast, in our model, we show that variations in dispersion of capital productivity at firm levels can be itself driven by shocks to financial intermediation sector. Hence, our work contributes to the understanding of the source of uncertainty shocks.

Our empirical evidence on capital allocation echoes previous empirical findings on resource reallocation and productivity dispersion over business cycles. For example, Maksimovic and Philips (2001) find that less productive firms tend to be sold as prospects of the aggregate economy improve. Correspondingly, aggregate output and the productivity dispersion across

²Foster, Haltiwanger and Krizan (2000, 2002) find that new entrants tend to be less productive than surviving incumbents, but exhibit substantial productivity growth over five or ten year horizons.

³See also, Chugh (2010), which explores the roles of uncertainty shocks in the framework of Carlstorm and Furest (1997).

firms are found to be negatively correlated (Eisfeldt and Rampini, 2006). Using COMPUSTAT dataset, Chen and Song (2013) further find that the dispersion of marginal (revenue) productivity of capital between constrained and unconstrained firms are significantly countercyclical. Our evidence suggests that such counter-cyclicality in productivity dispersion originates from procyclical capital allocation between small and large firms.

The paper is organized as follows. Section 2 describes the evidence to motivate our theory on lending efficiency shocks. Section 3 presents the model environment. Section 4 introduces the primitive shock and calibrates the benchmark model. In section 5, we report the simulation results. Section 6 concludes. The Appendix provides proof of various propositions.

2 Evidence

This section establishes the empirical evidence that motivates our theory in the next section. We first describe our data and the measurement of capital productivity. We then introduce our empirical strategy to estimate the dispersion of capital productivity. After that, we report the empirical results.

2.1 Data and Measurement

Our dataset consists of quarterly COMPUSTAT data from 1975Q2 to 2012Q3 for publicly listed firms, excluding foreign firms (those with a foreign incorporation code), financial firms (SIC code 6000-6999) and utilities (SIC codes 4000-4949). The full sample includes 216187 observations, for an average of 1511 observations per quarter. The details of data sources and construction are in the Appendix.

2.1.1 Constructing Firm Groups

The finance literature provides various proxies for the severity of financial constraints a firm is subject to. What is mostly relevant to this paper is firm size. Intuitively, small firms rely more heavily on bank credit than large firms. Moreover, small firms are more vulnerable to bank monitoring. As a result, they are more likely to be financially constrained.

Therefore, firms in our sample are classified into two groups in each year, with size as proxy for being financially constrained. The fraction of potentially/likely financially constrained firms in COMPUSTAT, accordingly to Hadlock and Pierce (2010, Table 1), is 26 percent, on average. Therefore, in our sample, the group of financially constrained firms includes those in the bottom size quartile, while the group of unconstrained firms is composed of those in the remaining size quartiles.

2.1.2 Measuring Capital Productivity Dispersion

We now turn to the firm-level productivity measure using COMPUSTAT data. The literature provides various approaches to estimate plant-level productive efficiency (e.g., Olley and Pakes, 1996 and Levinsohn and Petrin, 2003). These estimations are difficult to apply here since COMPUSTAT does not report firm-specific wage compensation, nor does COMPUSTAT have information on value-added. However, COMPUSTAT contains information on operating income.⁴ Then, capital productivity (KP henceforth) can be measured by the ratio of Operating Income before Depreciation (OIBDP) to one-quarter-lag net Plant, Property & Equipment (PPENT). We focus on all firm-quarter observations with positive operating income before depreciation and a non-missing value for capital stock.

We next compute the ratio of capital productivity between the two groups (KP ratio henceforth) as a proxy for the corresponding productivity dispersion caused by financial frictions.⁵

2.1.3 Estimating Capital Productivity Dispersion

We then address the empirical strategy of estimating the capital productivity dispersion between financially constrained and unconstrained firms or, more precisely, the relative capital productivity of constrained to unconstrained firms. For each time t , the KP ratio is estimated by regressing log of capital productivity, denoted as $\log KP_{it}$, on a dummy variable, d_{it} , where d_{it} equals one for the constrained firms and zero for the unconstrained.

$$\log KP_{it} = a_t + b_t d_{it} + \varepsilon_{it}. \tag{1}$$

The key coefficient of b_t in (1) corresponds to the log of the ratio of average capital productivity of financial constrained firms to unconstrained firms. To reduce the influence of outliers, we Winsorize $\log KP_{it}$ at the first and ninety-ninth percentiles. Our results hold qualitatively without Winsorization.

2.1.4 Results

Table 1 reports the summary statistics of estimated $\exp(b_t)$, the estimated relative capital productivity of constrained to unconstrained firms. The first four columns report the time-series mean, median, minimum and maximum of $\exp(b_t)$ between 1975Q2 and 2012Q3. The

⁴In COMPUSTAT, operating income (before depreciation) is equal to sales minus the cost of goods sold and selling, general and administrative expenses.

⁵Ideally, we should use the ratio of marginal product of capital, which is not directly observable. However, the ratio of marginal product of capital and the ratio of average product of capital are equal in our model presented in the next Section.

estimated b_t is statistically significant at one percent throughout the sample years, suggesting that financially constrained firms are more productive than unconstrained firms. As shown by the first two columns, the estimated capital productivity of constrained firms is, on average, more than 45-percent higher than that of unconstrained firms.

Table 1. Estimated KP Ratio

	<i>mean</i>	<i>median</i>	max	min	<i>STD</i>	mean t-stat
Estimated b	1.474	1.484	1.916	1.088	0.097	7.00

Figure 1 plots the H-P filtered estimated b_t . The NBER recessions are highlighted with the shaded bars. It is clear that the dispersion of capital productivity is countercyclical. The correlation coefficient between the H-P filtered real GDP and the estimated b_t is equal to -0.44 . The p -value for testing the hypothesis of no correlation is virtually zero.

[Insert figure 1 here]

We go a step further to explore the source of countercyclical variations in capital productivity dispersion. Note that the variation in the dispersion of capital productivity between the two groups can be decomposed into two terms: changes in the ratio of operating income and changes in the ratio of capital stock between two groups.

$$\Delta \log \frac{KP_t^c}{KP_t^u} = \Delta \log \frac{OI_t^c}{OI_t^u} - \Delta \log \frac{K_t^c}{K_t^u}$$

We adopt similar empirical strategy to estimate the ratio of operating income and the ratio of capital stock between constrained and unconstrained firms.

Figure 2 plot the cyclical component of the estimated ratio of capital between the two groups. Obviously, the ratio of capital stock of constrained to unconstrained group is highly procyclical. Furthermore, Table 2 shows the contemporaneous correlation of the ratio of capital and GDP is 0.55. By contrast, the ratio of operating income is essentially acyclical and, if any, correlates positively with GDP. In other words, the countercyclicity of capital productivity dispersion between small and large firms is entirely explained by the procyclical capital allocation between the two groups.

[Insert figure 2 here]

Table 2. Cross Correlation of Capital Productivity and its Components with GDP

	GDP_{t-2}	GDP_{t-1}	GDP_t	GDP_{t+1}	GDP_{t+2}
Ratio KP	-0.34 (0.0000)	-0.41 (0.0000)	-0.44 (0.0000)	-0.42 (0.0000)	-0.35 (0.0000)
Ratio of K	0.33 (0.0000)	0.49 (0.0000)	0.55 (0.0000)	0.55 (0.0000)	0.50 (0.0000)
Ratio of OI	-0.03 (0.7339)	0.05 (0.5560)	0.08 (0.3106)	0.11 (0.1994)	0.13 (0.1162)

In summary, our exercise uncovers three empirical observations that motivate our theory below: first, the capital productivity of financially constrained firms is, on average, higher than that of unconstrained firm. Second, the dispersion of capital productivity between financially constrained and unconstrained firms is significantly countercyclical. Finally, this countercyclicity is essentially driven by procyclical capital allocation between financially constrained and unconstrained firms.

3 The Benchmark Model

In this section, we describe the environment for the model in a general equilibrium economy. The economy is inhabited by an infinitely-lived representative household, a continuum of entrepreneurs and intermediate good producers with unit mass and a representative capital good producer.

Each period, entrepreneurs purchase intermediate goods to produce the final output. There are two types of entrepreneurs (type- c and type- u), differing in their utility discount factors, which governs entrepreneurs' net worth at the steady state. A competitive bank exists to provide working capital loan to entrepreneurs to finance the purchase of intermediate goods. All entrepreneurs are subject to idiosyncratic shocks to production technology, which becomes private information after they are realized. To shut down the extensive margin for resource allocation, we assume that the share of each type of entrepreneurs is constant, with a fraction η of type- entrepreneurs.

Intermediate goods producers for each type of project choose capital and labor to minimize the production cost, given the demand for intermediate goods.

3.1 The Household

We now turn to the household problem. The household has no access to production technology, but provides physical capital and labor to intermediate good producers each period. In addition, the household is entitled to the profits of the capital good producer. After the production takes place, the household makes optimal decisions on consumption, hours to work, and investment in physical capital. The representative household solves the following problem:

$$E_0 \sum_{t=0}^{\infty} \beta^t u(C_t^h - \varpi C_{t-1}^h, H_t) \text{ with } 0 < \beta < 1, \quad (2)$$

subject to

$$C_t^h + q_t a_{t+1}^h = [q_t (1 - \delta) + r_t] a_t^h + w_t H_t + \Pi_t^k$$

where a_{t+1}^h is end-of-period t physical capital purchased by the household, c_t^h and H_t denote the household's consumption and the total hours supplied, q_t is the price of physical capital. Π_t^k is the profits of the representative capital producer. The Euler equation is therefore

$$q_t MU_t^h = \beta E_t [q_{t+1} (1 - \delta) + r_{t+1}] MU_t^h$$

where $MU_t^h = u_{c_t^h} (c_t^h - \varpi c_{t-1}^h, H_t) - \beta \varpi u_{c_t^h} (c_{t+1}^h - \varpi c_t^h, H_{t+1})$.

3.2 Entrepreneurs

Each type of entrepreneur is endowed with the following technology to produce the final goods:

$$Y_t^j = A_t^j (y_t^j)^\mu \quad \text{for } j = c \text{ or } u$$

where A_t^j is production technology to a type- j project, y_t^j is the intermediate goods, Y_t^j is the final output. $0 < \mu < 1$.

We assume that $A_t^j = \bar{A}^j \xi_t$ where z_t is an aggregate technology shock and ξ_t is the idiosyncratic technology shock. Consider the following *i.i.d* shock process to idiosyncratic technology, ξ_t ,

$$\xi_t = \begin{cases} 1 - \sigma \sqrt{\frac{1-\pi}{\pi}} & \text{with probability } \pi \\ 1 + \sigma \sqrt{\frac{\pi}{1-\pi}} & \text{with probability } 1 - \pi \end{cases}$$

where the (unconditional) mean of this shock process is 1 and the unconditional variance is σ^2 . Denote

$$A_1^j = \bar{A}^j \left[1 - \sigma \sqrt{\frac{1-\pi}{\pi}} \right],$$

$$A_2^j = \bar{A}^j \left[1 + \sigma \sqrt{\frac{\pi}{1-\pi}} \right].$$

Note that $E(A_t^j) = \bar{A}^j$. In addition, $Var(A_t^j | \sigma) = \pi(1-\pi) (A_{2t}^j - A_{1t}^j)^2$.

Assuming that aggregate technology shock z_t follows AR(1) process:

$$z_{t+1} = \rho^z z_t + \epsilon_{t+1}^z$$

where $\epsilon_{t+1}^z \sim N(0, \sigma_{\epsilon^z}^2)$.

3.2.1 The Financial Contract

Before final good production takes place, entrepreneurs need to purchase the intermediate goods at a cost $\omega_t y_t^j$, where ω_t is the price of intermediate good to be determined later. Entrepreneurs can use both internal funds accumulated at the end of last period and external

borrowing to finance working capital. The gap between the working capital and entrepreneur's net worth is borrowed from a financial intermediary.

Each period, both types of entrepreneurs enter a financial contract with a financial intermediary. The contract is designed after the aggregate productivity shock, z_t , is realized, but *before* the idiosyncratic productivity shock, ξ_t^j , is realized. Debt is repaid at the end of the period.

Since the realized idiosyncratic technology is private information, payments to the financial intermediary at the end of the period will be made according to the report of the entrepreneur. The financial intermediary utilizes a costly-state-verification technology to verify the veracity of the report. Following (Greenwood, Sanchez, and Wang 2010), the probability that the entrepreneur of type- j is found cheating is specified as

$$P(m_t^j/y_t^j) = \begin{cases} 1 - \frac{1}{(\varepsilon_t m_t^j/y_t^j)^\psi} < 1, \psi > 0 \\ \quad \text{for a report } \xi_t^j \neq \xi_{2t}^j, \\ 0, \quad \text{for a report } \xi_t^j = \xi_{2t}^j. \end{cases}$$

where m_t^j denotes the monitoring input of a type- j project and is measured in consumption goods.

Note that the financial intermediary will only have to check when a bad state is reported. Also, under the assumption of $\psi > 0$, the probabilities of detecting malfeasance depends positively on the amount of monitoring per unit of intermediate goods, m_t^j/y_t^j , which we refer to as monitoring intensity. In other words, as the demand for intermediate goods and thus the size of loan increases, more monitoring inputs are needed to maintain the same probability of detecting misreport. Moreover, the above specification requires some threshold level of monitoring, $m_t^j > y_t^j/\varepsilon_t$. The assumption of the threshold level of monitoring makes sure that $P \geq 0$. ε_t affect the efficiency of monitoring technology, and follows a stochastic process. Note that ε_t is an aggregate shock to financial sector, capturing the fact that during a crisis, the efficiency of financial intermediation deteriorates due to, for example, a more volatile economic environment. We refer ε_t as the lending efficiency shocks.

The timing of financial contract at period t is as follows. Financial intermediary first decides on the amount of working capital to be lent out, and resources devoted to detect cheating if monitoring is needed. Then entrepreneurs use the working capital to purchase intermediary goods before the realization of the technology shock. At the end of period, entrepreneurs of type- j projects make a report on the production outcome. The financial intermediary decides whether to conduct a monitoring action and how much monitoring resources to put in. Finally, output is split between entrepreneurs and the financial intermediary based on the outcome of monitoring. Since this is a within-period loan, the interest payment is 0.

Given the entrepreneurial net worth a_t^j at the beginning of each period and a contract value v_t^j to be determined later, the optimal contract problem for a type- j project is

$$\max_{p_{1t}^j, p_{2t}^j, y_t^j, m_t^j} \{ \pi_1 p_{1t}^j + \pi_2 p_{2t}^j - (\omega_t y_t^j - q_t a_t^j) - \pi_1 m_t^j / \theta \} \quad (3)$$

subject to

$$p_{1t}^j \leq A_1^j (y_t^j)^\mu \quad (4)$$

$$p_{2t}^j \leq A_2^j (y_t^j)^\mu \quad (5)$$

$$[1 - P(m_t^j / y_t^j)] [A_1^j (y_t^j)^\mu - p_{1t}^j + (A_2^j - A_1^j) (y_t^j)^\mu] \leq A_2^j (y_t^j)^\mu - p_{2t}^j \quad (6)$$

$$\pi_1 (A_1^j (y_t^j)^\mu - p_{1t}^j) + \pi_2 (A_2^j (y_t^j)^\mu - p_{2t}^j) = v_t^j \quad (7)$$

where $\omega_t y_t^c - q_t a_t^c$ is the external borrowing, θ captures the unit cost of monitoring input in terms of final goods. In equation (6), the left-hand side is the payoff to the entrepreneur at state 2, when he is not caught cheating (it is optimal to set the payoff to zero when the entrepreneur report the low state and is caught cheating). $A_1^j (y_t^j)^\mu - p_{1t}^j$ is the payoff from the bank when reporting low state, while $(A_2^j - A_1^j) (y_t^j)^\mu$ is the benefit for cheating when the true state is 2.

To pin down type- c entrepreneur's contract value v_t^j , we assume the financial sector is competitive so that financial intermediary earns zero profit. The entrepreneur's contract value v^j can be solved as:

$$v_t^j \leq (\pi_1 A_1^j + \pi_2 A_2^j) (y_t^j)^\mu - (\omega_t y_t^j - q_t a_t^j) - \pi_1 m_t^j / \theta \quad (8)$$

Proposition 1 *There is no monitoring if and only if $A_1^j (y_t^{j,fb})^\mu \geq \omega_t y_t^{j,fb} - q_t a_t^j$, where $y_t^{j,fb} \equiv \arg \max_{y_t^j} (\pi_1 A_1^j + \pi_2 A_2^j) (y_t^j)^\mu - \omega_t y_t^j$.*

Proof: see Appendix.

The intuition is straight forward. At the first-best production scale, if the payoff in the worst state is still larger than the borrowing cost for the financial intermediary, then there is no incentive for intermediary to engage in costly monitoring. Proposition 1 implies that as the firm becomes sufficiently large, it becomes less relying on loan from financial intermediary to finance working capital. Accordingly, it is optimal for intermediary to have zero monitoring if the borrowing cost is less than the output in the lowest state.

In reality, this asymmetry in monitoring intensity stems from the fact that large firms rely less on indirect financing as they can resort to self-financing or direct financing (bond, stock

market). On the other hand, small firms rely largely on indirect finance (bank loan) for working capital finance. As a result, they are subject to more intensive monitoring.

Proposition 2 *Given that $A_1^j (y_t^{j,fb})^\mu < \omega_t y_t^{j,fb} - q_t a_t^j$, the limited liability constraint (4) is binding, the limited liability constraint for (5) is not binding and the Incentive Compatibility constraint is binding.*

Proof: see Appendix.

Later, we will show that for a type- c entrepreneur, the IC constraint is binding. Hence the optimal contract problem for a type- c entrepreneur can be rewritten

$$\max_{p_{2t}^c, y_t^c, m_t^c} \{(\pi_1 A_{1t}^c + \pi_2 A_{2t}^c) (y_t^c)^\mu - (\omega_t y_t^c - q_t a_t^c) - \pi_1 m_t^c / \theta - v_t^c\} \quad (9)$$

subject to

$$[1 - P(m_t^c / y_t^c)] (A_{2t}^c - A_{1t}^c) (y_t^c)^\mu = A_{2t}^c (y_t^c)^\mu - p_{2t}^c \quad (10)$$

and

$$\pi_2 (A_{2t}^c (y_t^c)^\mu - p_{2t}^c) = v_t^c \quad (11)$$

Combine (10) and (11), we can pin down the demand of intermediate goods by a type- c entrepreneur

$$y_t^c = \left[\frac{v_t^c}{\pi_2 (A_{2t}^c - A_{1t}^c)} (\varepsilon_t m_t^c / y_t^c)^\psi \right]^{\frac{1}{\mu}} \quad (12)$$

Equation (12) shows how lending efficiency shocks affect the scale and thus resources to be allocated to a type- c project. Given v_t^c , a reduction in lending efficiency shock ε_t intensifies the degree of information asymmetry, which then reduces the demand for the intermediate goods by the type- c entrepreneur. On the other hand, a reduction v_t^c reduces the incentive for the entrepreneur to tell the truth, which also reduces the working capital advanced by the intermediary. From a difference perspective, we see that given the demand for the intermediate good, a fall in the lending efficiency increase monitoring intensity. Besides, a fall in the contract value for type- c entrepreneur increases the monitoring intensity, as the entrepreneur's incentive to cheat increases.

Substituting (12) and the expression of $P(m_t^c / y_t^c)$ into the objective function and maximizing with respect to m_t^c / y_t^c , we can get the first-order condition with respect to m_t^c / y_t^c

$$\mu \bar{A}^c (y_t^c)^{\mu-1} - \omega_t = \frac{\pi_1}{\theta} m_t^c / y_t^c \left(1 + \frac{\mu}{\psi} \right) \quad (13)$$

Equation (13) suggests that the presence of monitoring cost drives a wedge between marginal product and marginal cost of intermediate goods. Obviously, countercyclical variation in m_t^c/y_t^c will cause the wedge be countercyclical. A combination of equation (12), (13) and (8) solves y_t^c , m_t^c/y_t^c and v_t^c simultaneously.

As in the literature of financial frictions, the net worth of type- c entrepreneurs serves as financial accelerator. It is easy to see from (8) that $q_t a_t^c$ has a one-to-one direct impact on v_t^c . In addition to this direct effect, the presence of financial friction will amplify the impact of a_t^c on v_t^c . This is because as v_t^c increases, the financial constraint is relaxed. As a result, y_t^c will increase, which further increases the contract value, v_t^c . Note that the interest rate for the within-period loan is 1. Hence $\frac{v_t^c(a_t^c)}{q_t}$ is the internal rate of returns.

$$\frac{v_t^c(a_t^c)}{q_t} = \frac{1}{1 - \frac{\pi_1 m_t^c}{\theta \psi v_t^c}}. \quad (14)$$

(14) shows that the more the intermediary monitors (relative to the project value), the higher is the internal rate of return. The external financing premium can be defined as

$$\frac{A_{2t}^c (y_t^c)^\mu - \frac{v_t^c}{\pi_2}}{\omega_t y_t^c - q_t a_t^c} - 1,$$

which is the gap between the lending rate of the intermediary, measured as the ratio of payoff to the bank in good state to the amount lent to the entrepreneur, and the interest rate that the intermediary borrows from the household.

3.2.2 The Consumption-Saving Problem of Entrepreneurs

Now we switch to the consumption-saving problem of a type- j entrepreneur. Each period, after the production takes place, a type- c entrepreneur decides how much to consume and to invest in physical capital a_t^j . To simplify our problem, we assume perfect consumption insurance within the group of type- c entrepreneurs after they receive the idiosyncratic productivity shock. As a result, the consumption-saving problem of a type- j entrepreneur can be aggregated and written as a problem faced by a representative type- j entrepreneur

$$E_0 \sum_{t=0}^{\infty} (\beta^j)^t u(c_t^j - \varpi c_{t-1}^j) \text{ with } 0 < \beta < 1 \quad (15)$$

subject to

$$c_t^j + q_t a_{t+1}^j = [q_t (1 - \delta) + r_t] a_t^j + v_t^j - q_t a_t^j$$

The Euler equation becomes

$$q_t MU_t^j = \beta^j E_t \left[MU_{t+1}^j \left(r_{t+1} - q_{t+1} \delta + v_{t+1}^{j'} \left(a_{t+1}^j \right) \right) \right]$$

We assume that $\beta^u = \beta$, $\beta^c = \beta\gamma$, where $\gamma < 1$. Note that $\gamma < 1$ implies that the type- c entrepreneur to discount future consumption more than the representative household. This is to make sure that a type- c entrepreneur is borrowing constrained at the steady state. On the other hand, a type- u entrepreneur is more patient than a type- c entrepreneur. At the steady state, $v^{u'}(a_t^u)/q_t = 1$, which means that the type- u entrepreneur is not financially constrained. Then, according to Proposition (1) it must be that at steady state, type- u entrepreneurs accumulate sufficient assets such that

$$A_1^u (y^u)^\mu \geq \omega y^u - q a^u$$

Lemma 1 *At steady state, for a type- u entrepreneur, neither (4) nor (6) is binding. The demand of intermediate good by a type- u entrepreneur is determined by*

$$\mu \bar{A}^u (y^u)^{\mu-1} = \omega \tag{16}$$

Proof: see Appendix.

Equation (13) and (16) deliver the price elasticity of intermediate goods for the two types of entrepreneurs.

$$\frac{\partial \log y_t^c}{\partial \log \omega_t} = - \frac{1}{1 - \mu} \frac{\omega_t}{\omega_t + \frac{\pi_1}{\theta} m_t^c / y_t^c \left(1 + \frac{\mu}{\psi} \right)} \tag{17}$$

$$\frac{\partial \log y_t^u}{\partial \log \omega_t} = - \frac{1}{1 - \mu} \tag{18}$$

Since $\frac{\omega_t}{\omega_t + \frac{\pi_1}{\theta} m_t^c / y_t^c \left(1 + \frac{\mu}{\psi} \right)} < 1$, (17) and (18) implies that the price elasticity of intermediate goods by type- u entrepreneur is larger than that for type- c entrepreneurs. Intuitively, the marginal cost for type- c entrepreneurs involves not only replacement cost, ω , but also expected monitoring cost, captured by $\frac{\pi_1}{\theta} m_t^c / y_t^c \left(1 + \frac{\mu}{\psi} \right)$. Hence, with given monitoring intensity, unconstrained entrepreneurs' input demand is more sensitive to changes in the price of intermediate goods. As will be shown later, this feature is crucial to explain why others shocks that we examine below, such as aggregate technological shocks, implies counterfactual predictions regarding the cyclicity of the productivity dispersion.

3.3 The Intermediate Good Producer's Problem

We assume that each intermediate good producer face the same Cobb-Douglas technology

$$e^{\frac{z_t}{\mu}} \left(k_t^j \right)^\alpha \left(h_t^j \right)^{1-\alpha} \geq y_t^j \tag{19}$$

The intermediary good producers for both type of projects rent capital and labor to minimize the cost of producing intermediate goods, given its demand y_t^j .

$$\min_{k_t^j, h_t^j} \left\{ r_t k_t^j + w_t h_t^j \right\} \quad (20)$$

subject to (19).

The first order conditions imply the capital and labor demand for each type of project

$$\begin{aligned} k_t^j &= \frac{y_t^j}{e^{\frac{z_t}{\mu}}} \left(\frac{w_t \alpha}{r_t (1 - \alpha)} \right)^{1-\alpha}, \\ h_t^j &= \frac{y_t^j}{e^{\frac{z_t}{\mu}}} \left(\frac{r_t (1 - \alpha)}{w_t \alpha} \right)^\alpha. \end{aligned}$$

Accordingly we have

$$\frac{k_t^c}{h_t^c} = \frac{k_t^u}{h_t^u} = \frac{\alpha w_t}{(1 - \alpha) r_t} \quad (21)$$

Equation (21) indicates that the capital-labor ratio between the two types of projects is the same, though resources might be misallocated across projects. In other words, we shut down within-in project resource misallocation as a potential source of misallocation.

Assume that the intermediate good market is competitive. The price of intermediate goods is thus equal to the marginal cost.

$$\omega_t = e^{-\frac{z_t}{\mu}} \left(\frac{r_t}{\alpha} \right)^\alpha \left(\frac{w_t}{1 - \alpha} \right)^{1-\alpha} \quad (22)$$

To see how resources are allocated between the two types of projects, note that the final output of a type- j project can be expressed as $Y_t^j = A_t^j \left(k_t^j \right)^{\alpha\mu} \left(h_t^j \right)^{(1-\alpha)\mu}$. Accordingly, the average marginal product of capital of a type- j project is $\overline{MPK}_t^j = \mu \overline{A}^j e^{z_t} \left(k_t^j \right)^{\alpha\mu-1} \left(h_t^j \right)^{(1-\alpha)\mu}$, $j = c$ or u . We measure of the degree of capital misallocation with ratio of average marginal product of capital between Type- c and Type- u project

$$MPKR_t = \frac{\overline{MPK}_t^c}{\overline{MPK}_t^u} = \frac{\overline{A}^c}{\overline{A}^u} \left(\frac{k_t^c}{k_t^u} \right)^{\mu-1} > 1$$

Here we use the feature of 21. In the first best where there is no asymmetric information, $MPKR_t = 1$ even if $\overline{A}^c \neq \overline{A}^u$.

3.4 The Capital Producer's Problem

To pin down the price of physical capital, we assume that in this economy, there is a representative capital producer. The capital producer purchases I_t units of consumption goods from the final good producer (and $(1 - \delta) K_t$ units of physical capital from the household and type- c

entrepreneur), and produces the new capital stock to be sold to the household and type- c entrepreneur at the end of the period.

The technology to transform new investment into installed capital involves installation cost, $S(I_t/I_{t-1})$, which increases in the rate of investment growth. Since the marginal rate of transformation from previously installed capital (after it has depreciated) to new capital is unity, the price of new and used capital are the same.

The capital producer's period- t profit can be expressed as

$$\Pi_t^k = q_t [(1 - \delta) K_t + \chi_t (1 - S(I_t/I_{t-1})) I_t] - q_t (1 - \delta) K_t - I_t$$

where χ_t is the marginal efficiency shocks to investment (MEI shocks).

Dynamically, the capital producer solves the following optimization problem

$$\max_{I_{t+j}} E_t \left\{ \sum_{j=0}^{\infty} \beta^j \lambda_{t+j} \Pi_{t+j}^k \right\} \quad (23)$$

where λ_t is the multiplier on the household's budget constraint. Note that if $S(I_t/I_{t-1}) = 0$, $q_t = 1$ and the model goes back to our previous case. The first order condition delivers

$$q_t = \frac{1 - E_t \beta \frac{\lambda_{t+1}}{\lambda_t} \left[q_{t+1} \chi_{t+1} S'(I_{t+1}/I_t) \left(\frac{I_{t+1}}{I_t} \right)^2 \right]}{\chi_t \left[1 - S'(I_t/I_{t-1}) \frac{I_t}{I_{t-1}} - S(I_t/I_{t-1}) \right]}$$

In order for q_t to be equal to 1 at the steady state and procyclical we assume $S(1) = S'(1) = 0$. More importantly, $S''(1) > 0$. As a result, when I_t/I_{t-1} falls relative to the steady state, $S'(I_t/I_{t-1})$ also falls. This leads to a decline in q_t . It is easy to see that at the steady state, $\Pi_t^k = 0$.

3.5 Competitive Equilibrium

A competitive equilibrium consists of financial contract $\{p_{1t}^j, p_{2t}^j, y_t^j, m_t^j, v_t^j\}_{t=0}^{\infty}$ for each type- j project, a set of factor inputs, $\{k_t^j, h_t^j\}$ for each intermediate good producer producing y_t^j , a set of allocation $\{c_t^j, a_{t+1}^j\}_{t=0}^{\infty}$ for each type- j entrepreneur, a set of allocation $\{c_t^h, a_{t+1}^h, H_t\}_{t=0}^{\infty}$ for the representative household, and the prices $\{q_t, r_t, w_t, \omega_t\}$, such that:

1. Given $\{\omega_t, v^j\}$, the financial intermediary of type- j project offers a contract $\{p_{1t}^j, p_{2t}^j, y_t^j, m_t^j\}$ by solving (3) (??).
2. v^j is such that the intermediary earns zero profit in a competitive environment.

3. Given $\{y_t^j, r_t, w_t\}$, the intermediate good producer minimizes its production cost, by solving the problem (20).
4. Given prices $\{r_t, w_t\}$, the representative household solves the problem .
5. Given r_t , a type- c entrepreneur solves the problem (15).
6. Given q_t , the capital producer solves the problem (23)
7. All markets clear:
 - (a) Wage rate w_t clears labor market: $\eta h_t^c + (1 - \eta) h_t^u = H_t$.
 - (b) Rental rate r_t clears capital market: $\eta a_t^c + a_t^h = K_t = \eta k_t^c + (1 - \eta) k_t^u$.
 - (c) The price of physical capital, q_t , clears the capital markets: $K_{t+1} = (1 - \delta) K_t + (1 - S(I_t/I_{t-1})) I_t$.
 - (d) Good market clears:

$$\begin{aligned}
Y_t &= (1 - \eta) \bar{A}^h e^{z_t} (k_t^h)^{\alpha\mu} (h_t^h)^{(1-\alpha)\mu} + \eta \bar{A}^l e^{z_t} (k_t^l)^{\alpha\mu} (h_t^l)^{(1-\alpha)\mu} \\
&= C_t + I_t + \eta \pi_1 m_t^c / \theta
\end{aligned}$$

where $C_t \equiv c_t^h + (1 - \eta) c_t^u + \eta c_t^c$ denote the aggregate consumption.

4 Numerical Results

To explore how financial frictions and capital allocation responds to various shocks, in this section, we provide numerical examples to illustrates the impulse responses of different variables to various shocks in our model: lending efficiency shocks, aggregate technology shocks and other shocks commonly used in the literature.

Table 3 summarizes the parameter values used in our numerical simulation.

Table 3: Parameter Values

Parameter	Target	Value
Standard Parameters		
α	Labor income share of type- u project	0.294
δ	Investment-capital ratio of 0.074	0.074
β	Interest rate of 4%	0.979
ϕ	Hours worked of 1/3	1.8
ν	Frisch elasticity of labor 2.5	0.4
μ	Akteson and Kehoe (1995)	0.85
Non-standard Parameters		
\bar{A}^l	Normalization	1.0
\bar{A}^h	Normalization	1.88
ψ	Inter-quartile range of $TFPR$, 1.56	0.20
ε	Aggregate capital output of 2.5	1.5
σ	Bloom Floetotto and Jiamovich (2010)	0.44
ρ^z	Quarterly persistence of 0.987	0.0
σ_{ε^z}	SD of H-P filtered aggregate TFP	0.016
η	Kaplan and Zingales (1997)	0.25
γ	External Financing Premium	0.975
θ	Normalization	1
κ	Estimation	1
ϖ	Estimation	0.5
$\bar{\chi}$	Normalization	1
π	Two-state approx. to normal distribution	0.5

4.1 Lending Efficiency Shocks

Figure 3 plots the responses of various macro variables to a one percentage decrease in lending efficiency, ε_t . As shown by the top panel, such a shock generates a recession, as aggregate output, consumption, investment and hours worked all fall on impact. Moreover, aggregate TFP decreases, though the aggregate technology, z , is unchanged. Such a fall in aggregate TFP results from an increase in the $MRPK$ gap between financially constrained and unconstrained projects, as evident by the middle right panel.

The intuition is as follows. A reduction in lending efficiency reduces the probability for the bank to identify the misreport of the entrepreneurs. This tightens the incentive compatibility constraints. As a result, the bank increases the monitoring intensity, as shown by the bottom left panel of figure 3. Since monitoring is costly, an increase in monitoring intensity tightens the financial constraint of type- c projects, forcing them to reduce the demand of intermediate goods. This widens the gap of marginal product of capital between the two types of projects. As a consequence, aggregate TFP falls on impact.

The above mechanism is further amplified by the financial accelerator channel. The lower demand for intermediate goods causes the production scale and the profit of the type- c project

to be shrunk. Accordingly, the contract value for the type- c entrepreneur falls, which further tightens the bank loan and shrinks the production scale to a type- c project. Associated with a tightening of the financial constraint is an increase in the external financing premium and a fall in the price of capital, shown in the bottom middle and right panels of figure 3.

In summary, a fall in lending efficiency not only triggers a recession for aggregate variables, but also generates countercyclical dispersion of marginal product of capital between financially constrained and unconstrained projects. As a consequence, aggregate TFP falls.

[Insert figure 3 here]

4.2 Aggregate Technology Shocks

We now turn to the impulse response to a reduction in aggregate technology, z_t . Figure 4 plots the impulse responses of various variables to such a shock. As in a standard RBC model, a fall in aggregate technology leads to a fall in aggregate output, investment, consumption and hours worked. However, it also generates a counterfactual increase in aggregate TFP. The reason is that the dispersion of $MRPK$ between constrained and unconstrained projects shrinks, contrast to what we have observed in the data. Moreover, the bottom panels show that both monitoring intensity and external financing premium falls in recessions. Again, these results are counterfactual.

The key underlying the counterfactual predictions of aggregate technological shocks on variations in the dispersion of MPK and aggregate TFP is the asymmetric price elasticity of input demand for the two types of entrepreneurs. A fall in aggregate technology, by construction, discourages the input demand of entrepreneurs via a higher replacement cost. Since the input demand by unconstrained firms is more sensitive to an increase in price, such a shock reallocates resources from the unconstrained to constrained entrepreneurs. Financially constrained projects, therefore, expands initially, despite a fall in aggregate technology. This results in a narrowing of the the gap of MPK and an increase in aggregate TFP . In addition, the initial expansion of financially constrained project increases the project value, which helps to relax the incentive compatibility constraint and reduce the monitoring intensity. These predictions run counter to the countercyclical dispersion of MPK and procyclical capital allocation between small and large firms as we found in Section II.

[Insert figure 4 here]

4.3 Other Aggregate Shocks

In this section, we further explore the predictions of other aggregate shocks regarding the cyclical behavior of MPK dispersion and aggregate TFP. These shocks include marginal efficiency shocks to investment or so called “MEI shocks” (e.g. Justiano, Primiceri and Tambalotti, 2011), shocks to labor disutility, shocks to utility discount factors and shocks to capital depreciation rates.

The impulse responses to each of the above shocks are plotted in Figure 5, 6, 7, and 8 respectively. It is clear that though these shocks can generate a recession in terms of aggregate variables such as aggregate output, hours worked, investment and/or consumption, they all predict that the dispersion of MPK narrows in recessions, contrast to what is observed in the data. Accordingly, each of these shocks predicts a counterfactual increase in aggregate TFP.

Again, similar to aggregate TFP shocks, the key underlying the counterfactual predictions of these shocks on the cyclical behavior of MPK dispersion is the asymmetric price elasticity of input demands by the two types of entrepreneurs. Note that all these shocks influence the demand of intermediate goods by entrepreneurs via the price of intermediate goods alone. Specifically, a MEI shock affects the capital supply and therefore the rental price of capital. A labor disutility shock, on the other hand, affects the aggregate labor supply and, thus, the wage rate. A shock to utility discount factor alters the saving incentives and therefore the interest rate. Finally, a capital depreciation shock affects the supply of installed capital and again the rental price of capital. Therefore, these shocks trigger a larger response of the input demand by unconstrained entrepreneurs than that of constrained entrepreneurs. This leads to a reallocation of capital in the opposite direction to what we found in the data, and therefore, counterfactual predictions on dispersion of MPK and aggregate TFP.

[Insert figure 5, 6, 7 and 8 here]

5 Bayesian Estimation

to be written.

6 Conclusion

This paper develops a theory in which shocks to the efficiency of financial intermediation translate into business cycle fluctuations via capital reallocation. In our theory, the realized idiosyncratic productivity is subject to asymmetric information between the bank and the entrepreneurs. However, due to costly verification, the bank will monitor only the cash flows

to those entrepreneurs with insufficient net worth. This creates a gap of marginal product of capital between entrepreneurs of different size. Shocks to the efficiency of bank monitoring, by affecting the probability to detecting misreport by entrepreneurs, alter the monitoring intensity of the bank. Accordingly, capital is reallocated between entrepreneurs of different sizes, creating countercyclical variations in the dispersion of marginal productivity of capital and procyclical TFP fluctuations. By contrast, in our model other shocks commonly adopted in the literature predict a procyclical dispersion of marginal product of capital and countercyclical TFP fluctuations. Using the COMPUSTAT dataset, we find a significant countercyclical pattern for the relative capital productivity of financially constrained to unconstrained firms, which is driven by the procyclical capital allocation between these two types of firms. Therefore, both our theory and evidence suggest that shocks to bank lending efficiency might be important drivers of business cycles.

References

- [1] Bachmann, R. and C. Bayer (2009), “Firm-Specific Productivity Risk over the Business Cycle: Facts and Aggregate Implications”, working paper.
- [2] Bernanke, B. M. Gertler and S. Gilchrist (1999), “The Financial Accelerator in A Quantitative Business Cycle Framework”, in Handbook of Macroeconomics.
- [3] Bloom, N., S. Bond, and J. Van Reenen (2007), “Uncertainty and Investment Dynamics”, *Review of Economic Studies*, 74, 391–415.
- [4] Bloom, N., M. Floetotto, and N. Jaimovich (2007), “Really Uncertain Business Cycles”, *mimeo*.
- [5] Bloom, Nicholas.(2009), “The impact of uncertainty shocks”, *Econometrica*, vol.77, No.3, 623-685.
- [6] Castro, R., G. L. Clementi and Y. Lee (2010), “Cross-Sectoral Variation in Firm-Level Idiosyncratic Risk”, *mimeo*
- [7] Carlstrom, C. and T. Fuerst (1997), “Agency Costs, Net Worth, and Business Fluctuations: A Computable General equilibrium Analysis”, *American Economic Review*, December, 893-910.
- [8] Carpenter, R. and Petersen, B. (2002), “Capital Market Imperfections, High-tech Investment, and New Equity Financing”, *Economic Journal*, 102 (2), pp. F54-72.
- [9] Chen, K. and Z. Song, (2013), “Financial Frictions on Capital Allocation: A Transmission Mechanism of TFP Fluctuations”, *Journal of Monetary Economics*, Vol. 60, 2013, 683-703.
- [10] Christiano, L. R. Motto and M. Rostagno (2010), “Financial factors in Economic Fluctuations”, working paper, Northwestern University.
- [11] Chugh, S. (2010), “Firm Risk and Leverage-Based Business Cycle”, *mimeo*
- [12] Eisfeldt, A. and A. Rampini (2006), “Capital Reallocation and Liquidity”, *Journal of Monetary Economics*, 119 (3), 899-927.
- [13] Gilchrist, S. and C. Himmelberg (1995), “Evidence on the Role of Cash Flow for Investment”, *Journal of Monetary Economics*, December 1995, 36(3), 541-572.

- [14] Gilchrist, S., J. Sim and E. Zakrajsek (2010), “Uncertainty, financial frictions, and investment dynamics”, *mimeo*.
- [15] Greenwood, J., Z. Hercowitz and G. Hoffman (1988), “Investment, Capacity Utilization and the Real Business Cycle,” *American Economic Review*, 78, 402-417
- [16] Greenwood, J., Sanchez, J. M. and Wang, C. (2010), “Financing Development: The Role of Information Costs”, *American Economic Review*, 100, 1875-1891.
- [17] Sim, J.(2008), “Uncertainty driven business cycle”, *mimeo*.
- [18] Stock, J., and M. Watson (1998), “Business Cycle Fluctuations in US Macroeconomic Time Series”, in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford. Elsevier.

7 Appendix

We now prove the propositions in Section III. To simplify the notation, we drop the time subscript.

7.1 Proof for Proposition 1

We first drive the necessary condition for $m_t^j = 0$ or, equivalently, $P = 0$. From the IC constraint (6), we have

$$A_2^j (y_t^j)^\mu - p_{2t}^j \geq A_1^j (y_t^j)^\mu - p_{1t}^j + (A_2^j - A_1^j) (y_t^j)^\mu \quad (24)$$

also, combining (7) with (8), we get

$$\left(\pi_1 A_1^j + \pi_2 A_2^j \right) (y_t^j)^\mu - \left[\pi_1 (A_1^j (y_t^j)^\mu - p_{1t}^j) + \pi_2 (A_2^j (y_t^j)^\mu - p_{2t}^j) \right] - \pi_1 m_t^j / \theta \geq \omega_t y_t^j - q_t a_t^j \quad (25)$$

Plugging (24) (with equality) into (25), and given that $m_t^j = 0$, we obtain the necessary condition

$$\left(\pi_1 A_1^j + \pi_2 A_2^j \right) (y_t^j)^\mu - \pi_2 \left[A_1^j (y_t^j)^\mu - p_{1t}^j + (A_2^j - A_1^j) (y_t^j)^\mu \right] - \pi_1 (A_1^j (y_t^j)^\mu - p_{1t}^j) \geq \omega_t y_t^j - q_t a_t^j$$

or

$$\begin{aligned} A_1^j (y_t^j)^\mu &\geq A_1^j (y_t^j)^\mu - p_{1t}^j + \omega_t y_t^j - q_t a_t^j \\ &\geq \omega_t y_t^j - q_t a_t^j \end{aligned}$$

where the second inequality obtains from the limited liability condition (4). To prove the sufficiency, the financial contract can be simply designed as

$$p_{1t}^j = p_{2t}^j = \omega_t y_t^{j,fb} - q_t a_t^j \quad (26)$$

Note that, by assumption, the payoff at the low state, $A_1^j (y_t^j)^\mu - (\omega_t y_t^j - q_t a_t^j)$ is non-negative, and thus is feasible. Plugging (26) into the IC constraint (6),

$$[1 - P(m_t^j / y_t^j)] \left[A_2^j (y_t^{j,fb})^\mu - \omega_t y_t^{j,fb} - q_t a_t^j \right] \leq A_2^j (y_t^{j,fb})^\mu - (\omega_t y_t^{j,fb} - q_t a_t^j)$$

Obviously, the above IC constraint is always satisfied, even if no monitoring resource is used and $P = 0$. So IC constraint can be dropped from the intermediary's problem (3). Also, the zero profit condition of the intermediary, (25), is satisfied. Hence, it is optimal to set $m_t^j = 0$ and $P = 0$. Intuitively, since the intermediary does not monitor in either states, given that entrepreneur has incentive to lie, it is optimal to set the payoff at both states to be the borrowing cost of the intermediary.

7.2 Proof for Proposition 2

The Lagrangian of the problem (3) is

$$\begin{aligned}
L &= \pi_1 p_1^j + \pi_2 p_2^j - (\omega y^j - q a^j) - \pi_1 m^j / \theta \\
&+ \lambda_1 [\pi_1 (A_1^j (y^j)^\mu - p_1^j) + \pi_2 (A_2^j (y^c)^\mu - p_2^j) - v^j] \\
&+ \lambda_2 [A_2^j (y^j)^\mu - p_2^c - (1 - P(m^j / y^j)) (A_2^j (y^j)^\mu - p_1^j)] \\
&+ \lambda_{31} [A_1^j (y^j)^\mu - p_1^c] + \lambda_{32} [A_2^j (y^j)^\mu - p_2^j]
\end{aligned}$$

The first-order conditions are:

$$\frac{\partial L}{\partial p_1} = \pi_1 (1 - \lambda_1) + \lambda_2 (1 - P(m^j / y^j)) - \lambda_{31} = 0 \quad (27)$$

$$\frac{\partial L}{\partial p_2} = \pi_2 (1 - \lambda_1) - \lambda_2 - \lambda_{32} = 0 \quad (28)$$

$$\begin{aligned}
\frac{\partial L}{\partial y^j} &= -\omega + \lambda_1 \mu (\pi_1 A_1^j + \pi_2 A_2^j) (y^j)^{\mu-1} + \\
&\lambda_2 [\mu A_2^j (y^j)^{\mu-1} + \frac{\partial P(m^j / y^j)}{\partial y^j} (A_2^j (y^j)^\mu - p_1^j) - (1 - P(m^j / y^j)) \mu A_2^j (y^j)^{\mu-1}] \\
+ (\lambda_{31} A_1^c + \lambda_{32} A_2^c) \mu (y^c)^{\mu-1} &= 0
\end{aligned}$$

Proof. From the first-order conditions, we will have the following results:

Result 1 : $\lambda_1 \in (0, 1)$

Proof: From (28), we have $\lambda_1 \in [0, 1]$. Since promise-keeping constraint is binding, $\lambda_1 \in (0, 1]$.

Now we turn to prove $\lambda_1 \neq 1$. Suppose not, then from (27) and (28), we will have $\lambda_2 = \lambda_{31} = \lambda_{32} = 0$. Then from (29), $\pi_1 = 0$, a contradiction.

Result 2 : $\lambda_{31} > 0$, limited liability constraint for p_1^c is binding

Proof: from result 1 and (27), we can directly have this result.

Result 3 : $\lambda_2 > 0, \lambda_{32} = 0$, so IC constraint is binding and limited liability constraint for p_2^c is not binding.

Proof: suppose not, from result 1 and (28), we have

Case 1: $\lambda_2 = 0, \lambda_{32} > 0$ Then we have $p_2^j = A_2^j (y^j)^\mu$. However, result 2 and IC imply

$p_2^j < A_2^j (y^j)^\mu - (1 - P(m^j/y^j))(A_2^j (y^j)^\mu - A_1^j (y^j)^\mu) < A_2^j (y^j)^\mu$, a contradiction.

Case 2: $\lambda_2 > 0, \lambda_{32} > 0$ Then we have $p_2^j = A_2^j (y^j)^\mu$. However, result 2 and IC imply $A_2^j (y^j)^\mu - p_2^j = (1 - P(m^j/y^j))(A_2^j (y^j)^\mu - A_1^j (y^j)^\mu) > 0 \Rightarrow A_2^j (y^j)^\mu > p_2^j$, a contradiction. ■

7.3 Proof of Lemma 1

To prove Lemma 1, it is useful to show a dual problem of (3), which is to design contract $\{C_{1t}^j, C_{2t}^j, y_t^j, m_t^j\}$ to maximize the expected payoff of the entrepreneur, subject to the limited liability conditions and the IC and PC constraint for intermediary.

$$\max_{C_{1t}^j, C_{2t}^j, y_t^j, m_t^j} \{\pi_1 (1 - P_1) C_{1t}^j + \pi_2 C_{2t}^j\} \quad (30)$$

subject to

$$\begin{aligned} C_{1t}^j &\geq 0 \\ C_{2t}^j &\geq 0 \\ [1 - P(m_t^j/y_t^j)] [A_1^j (y_t^j)^\mu - p_{1t}^j + (A_2^j - A_1^j) (y_t^j)^\mu] &\leq A_2^j (y_t^j)^\mu - p_{2t}^j \\ (\pi_1 A_1^j + \pi_2 A_2^j) (y_t^c)^\mu - \pi_1 m_t^j/\theta - (\pi_1 C_{1t}^j + \pi_2 C_{2t}^j) &\geq \omega_t y_t^j - q_t a_t^j \end{aligned}$$

It is easy to show that proposition (1) still holds under this dual problem. As a result, when $A_1^u (y_t^u)^\mu \geq \omega_t y_t^u - q_t a_t^u$

$$C_{1t}^j = A_1^j (y_t^j)^\mu - (\omega_t y_t^j - q_t a_t^j) \quad (31)$$

$$C_{2t}^j = A_2^j (y_t^j)^\mu - (\omega_t y_t^j - q_t a_t^j) \quad (32)$$

Plug (31) and (32) into the objective function, the original problem for type- u becomes

$$\max_{y_t^u} (\pi_1 A_1^u + \pi_2 A_2^u) (y_t^u)^\mu - \omega_t y_t^u + q_t a_t^u$$

The first order condition is

$$\mu \bar{A}^u (y_t^u)^{\mu-1} = \omega_t$$



Figure 1: Ratio of MRPK between Financially Constrained and Unconstrained Firms

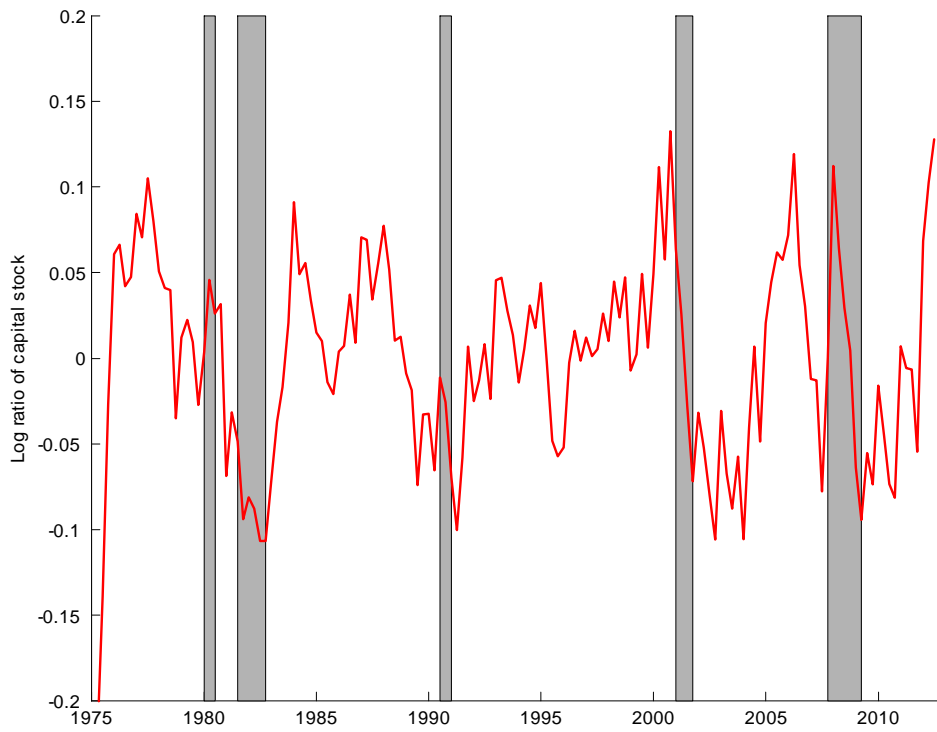


Figure 2: Ratio of Capital Stock Between Financially Constrained and Unconstrained Firms

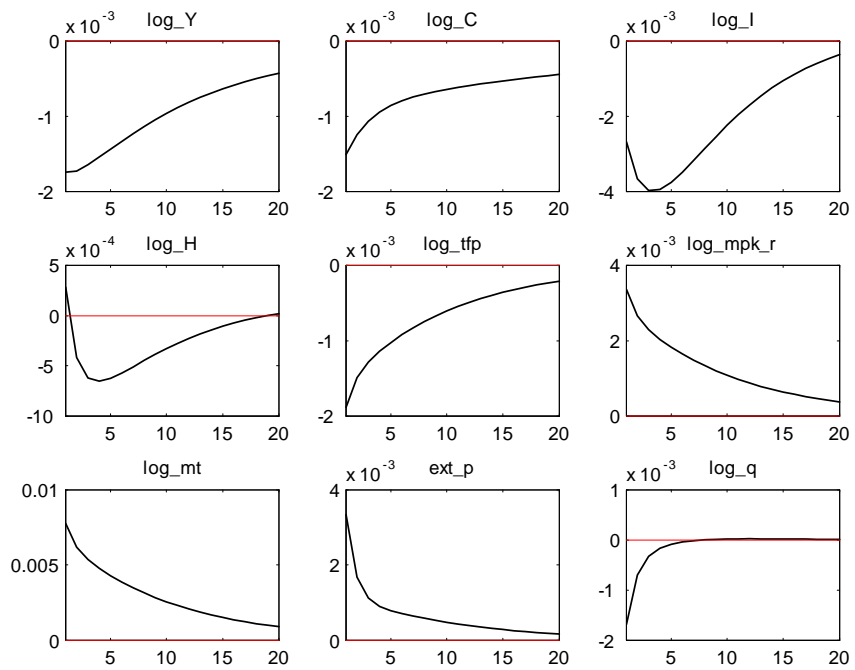


Figure 3: Response of Variables on Financial Friction to Lending Efficiency Shocks

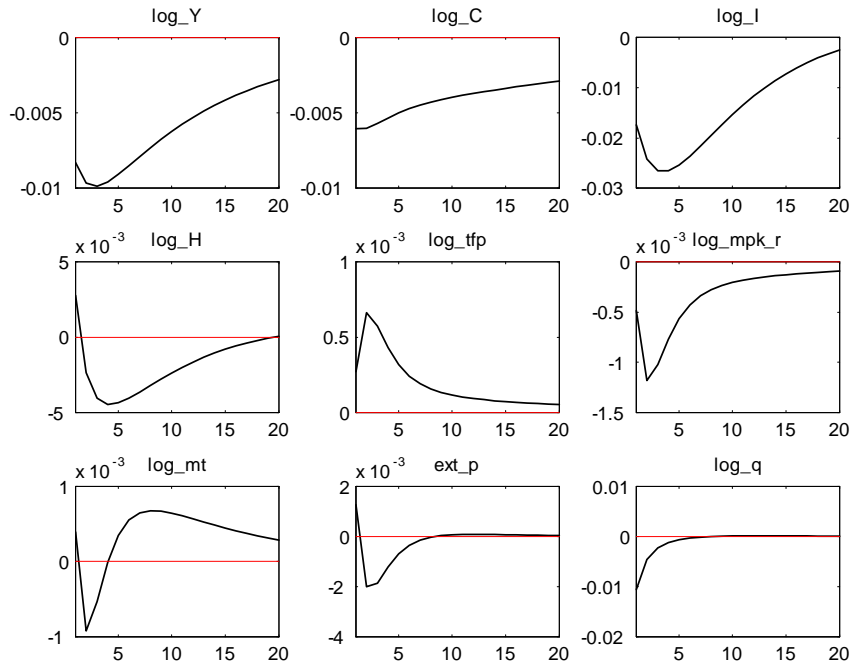


Figure 4: Response of Macro Variables to Aggregate Technology Shocks

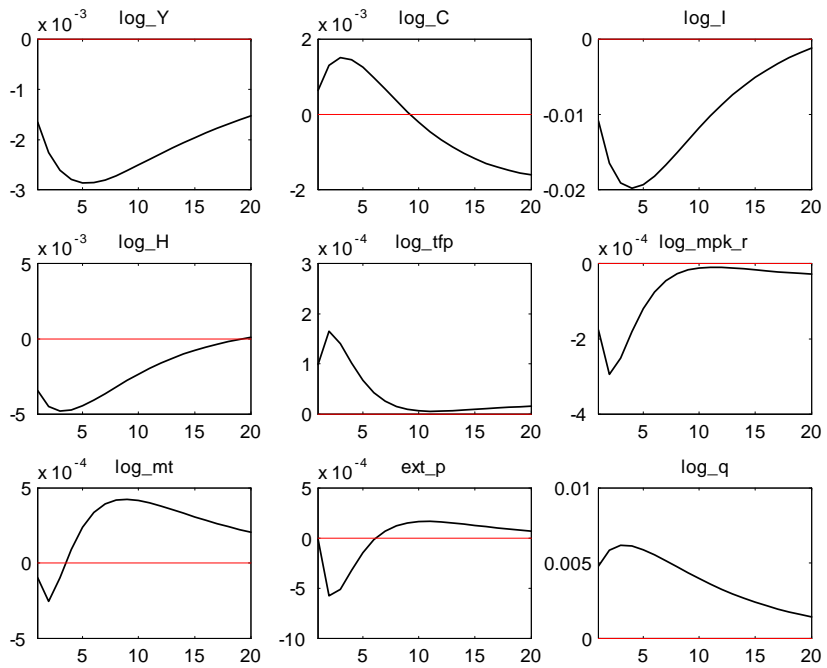


Figure 5: Responses of Macro Variables to Marginal Efficiency Shocks to Investment

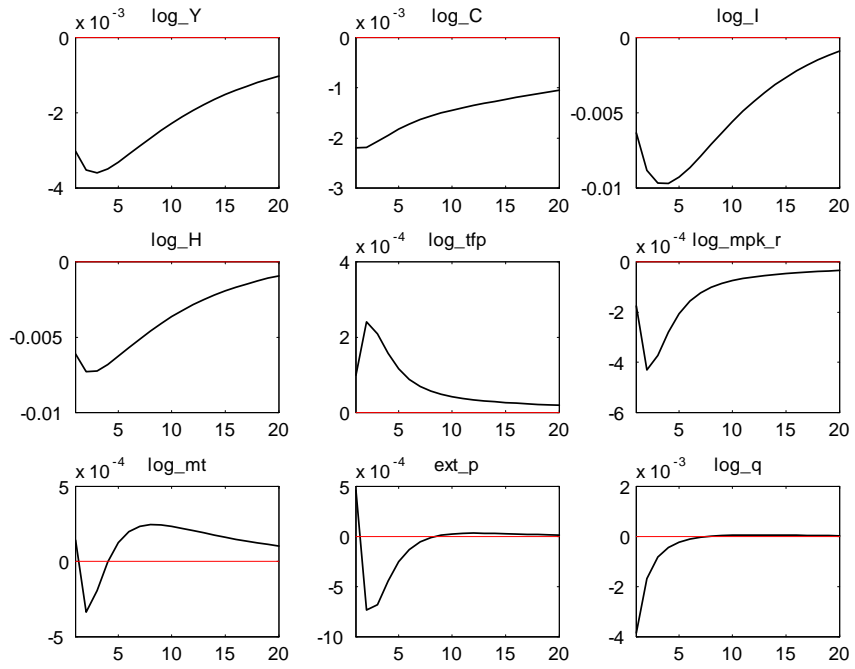


Figure 6: Response of Macro Variables to Labor Disutility Shocks

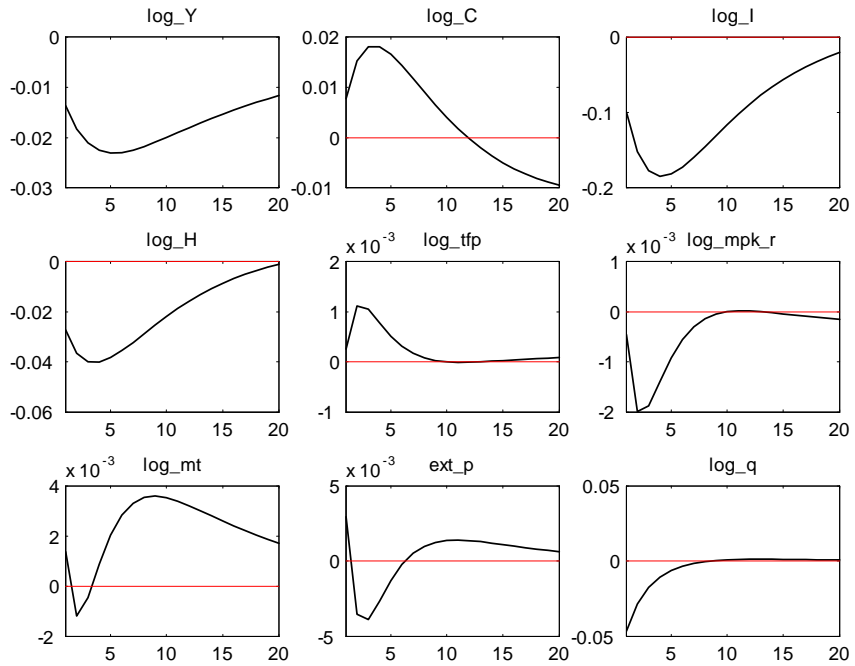


Figure 7: Response of Macro Variables to Discount Factor Shocks

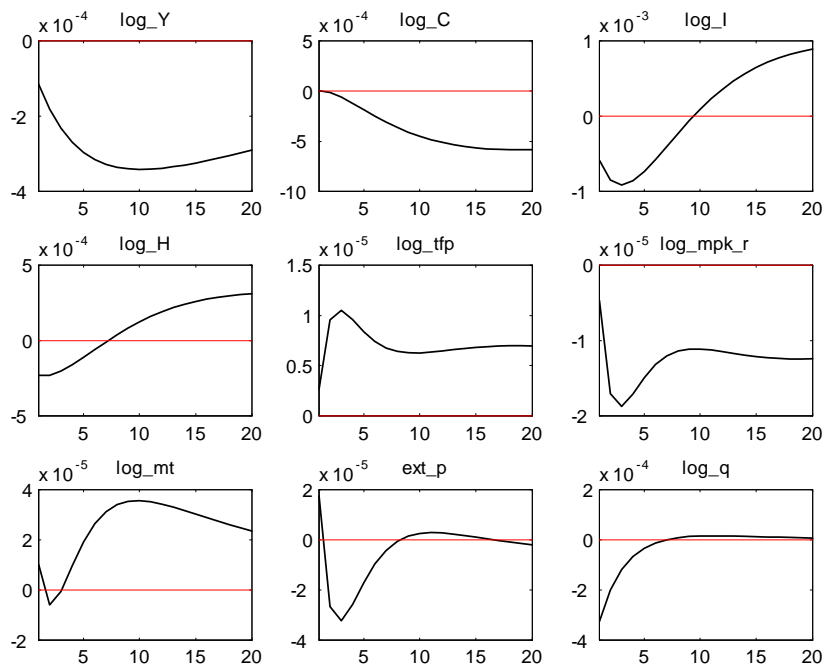


Figure 8: Response of Macro Variables to Depreciation Rate Shocks