Banks and Endogenous Firm Entry

William Swanson
Department of Economics
University of California, Davis

May 14, 2019

Abstract

The Global Financial Crisis brought much needed attention to the hazards of highly leveraged bank balance sheets. In the wake of the crisis, the rate of consolidation of non-financial firms has increased, and more large firms are turning inward to finance new investments. This begs the question, what are the consequences of firm leverage as compared to bank leverage? This paper starts with a standard RBC model augmented with balance-sheet constrained financial intermediaries. To address the question, we then introduce financially constrained firms that own the country's capital stock as they enter and exit endogenously. We show principally from the estimated model that much of the variation in U.S. consumption and investment originates from constraints on firm balance sheets, rather than constraints on bank balance sheets. Modeling the two agents independently with separate but related sources of financing allows the model to parse the two sources of volatility, and to take a stand on which is empirically a greater source of variance. The paper's results are increasingly relevant in a world with large firms that depend more on internal sources of finance, instead of borrowing from banks or government entities.

1 Introduction

Access to financing is well documented in the literature as being important for firm creation and business cycle dynamics. Empirical papers on the topic are unequivocal that barriers to financing can arise both internally from a firm, and externally from lending institutions or banks. Bricongne, Fontagné, Gaulier, Taglioni, and Vicard (2012) and Manova, Wei, and Zhang (2015) are well cited examples of papers on the topic, showing that much of the idiosyncratic firm behavior exhibited during the Global Financial Crisis can be linked to constraints in the banking sector and a firms balance sheet. A sophisticated theoretical literature has evolved to put structure on some of the complex relationships at work between firms and their needs for financing. Models of endogenous firm entry and financially constrained firms ((Jermann & Quadrini, 2012), (Bergin, Feng, & Lin, 2017)) have proven useful in replicating stylized facts, particularly when sunk and fixed costs are included. Models of this class, however, often feature a single financial constraint on either a firm or a bank. In this paper we argue that, it may be inappropriate to represent the role of 'financial constraints' with only one constraint, particularly as firms are increasingly their own source of financing.

While vulnerability to country-level banking-sector shocks diminish as firms grow, but risks from over-leveraging may grow too. In this paper we expand on the existing models of firm entry to feature banks and firms as subject to separate financial constraints and shocks. While firms and banks are clearly related, we show through the estimated model that most of the volatility in U.S. consumption and investment can be explained by shocks to firm balance sheets, rather than banks.

The model yields simple expressions for the degree that banks and firms are financially constrained. The expressions can be readily taken to the data, so that after presenting the model in this paper, we do just that. First, we calibrate the model using standard parameters for a small open, and compare the models predicted business cycle properties to the actual. Our focus is particularly on the behavior of the Lagrange multipliers for the firm's and bank's financial constraints. Then, we estimate the model using a standard Bayesian estimation toolkit to parse out the distinct role of financial constraints for firms and banks.

This paper is related to several literature, but is closest to the research in endogenous firm entry and business cycle dynamics following Bilbiie, Ghironi, and Melitz (2012), Bilbiie, Ghironi, Melitz, Midrigan, and Rotemberg (2007). This strand of papers feature models with fixed or sunk costs of entry, monopolistic pricing and often a focus on monetary policy. In particular, this paper is related to models that incorporate financial constraints on firm investment such as Jermann and Quadrini (2012) and Bergin et al. (2017). Bond, Tybout, and Utar (2008) also approaches a related question to ours, but their approach is strictly empirical and is focused on Colombia in the 1990s.

A key contribution of this paper is that we introduce an incentive compatibility constraint on firm entry and investment, rather than a collateral constraint. A key feature of our constraint, borrowed from (Gertler & Karadi, 2011), is that it allows us to generate high degrees of persistence through an intuitive mechanism. In the model, new firms devote a smaller share of their net assets to capital because they are more constrained. In effect, sunk costs hinder new entrants from taking on as much leverage as existing entrants. This is because a sunk cost is an investment that does not serve as collateral. It therefore causes the ICC constraint to bind more tightly for newly entering firms.

This feature of the model, that new firms start with a lower ratio of capital to net assets, is key to generating high degrees of persistence as the economy responds to shocks. With a large jump in entry, the average ratio of capital to assets falls, and can even cause the aggregate stock of capital to fall, depending on the calibration of the model. The fall in capital per-firm mutes the full impact of the technology shock, and only after many periods, during which the capital ratio is improving, will the result of the shock start to dissipate. Imposing sunk costs on entering firms allows this model to easily account for the hump-shaped response of firm-entry often seen in the data following technology shocks.

With this feature comes a cost: all firms are not identical. As we show in the model, we need only track new firms for one period. After the initial period of entry, all firms allocate identical shares of their net assets to new investments. As long as the investment shares are identical, tracking the average level of firm net assets is sufficient to solve the model.

Empirical papers that are related to this paper include Manova (2012), who documents patterns in international trade that are likely caused by credit constraints. The authors model constrained firms by having an upfront fixed cost that needs to be financed externally. Hölzl (2005) is an empirical study that focuses on entry and exit in small open economies. They find that sunk costs are vital for explaining the asymmetry in entry and exit rates across countries–in particular, high sunk costs tend to decrease the exit rate of firms, just as they increase rents for surviving firms. Uusküla (2016) builds on the literature to show that financial constraints help endogenous entry models mimic the impact of a monetary expansion. Lewis (2009) uses a small open economy model based on Bilbiie et al. (2012), but they add in components such as sticky wages to get achieve the hump shaped response of firm entry to expansionary monetary policy.

2 A Model of Constrained Banks and Firms

We introduce banking frictions into a standard closed economy Real Business Cycle model with endogenous firm entry. The unique feature of this model is that both firms and banks are constrained in the degree that they can invest. While the firm is constrained in the amount of capital they can purchase for next period, the bank is constrained in the quantity of loans they can supply to the firm. The addition of a financial intermediary, along with endogenous firm entry, generates significant persistence and an amplification mechanism for technology shocks.

There are two main agents in the model. First, a continuum of monopolistically competitive firms produce a differentiated good using capital and labor. Along with normal operation profits, we allow these firms to purchase and own capital, sell claims on their profits, and have loans from a financial intermediary. As the second principal agent, the bank follows roughly from (Gertler & Karadi, 2011). In particular, these banks supply loans and purchase the equity claims from the firm, while holding deposits from a representative consumer. As is standard in the RBC literature, time is discrete and infinite. Therer are exogenous technology shocks z_t, bank financial shocks κ_t, and firm financial shocks, γ_t that follow exogenous processes as they arrive unexpectedly in the economy at the start of each period. The consumer exhibits rational expectations when making their investment decisions for next period, and since we assume the bank and firms are owned by the consumer, they too exhibit rational expectations.

In the remainder of this section, we outline the details of the model. In addition to the bank and differentiated goods firms, a representative consumer supplies labor and holds deposits with the bank to purchase consumption of a final good. Final goods are a composite of the differentiated intermediate good and a non-differentiated good that requires only labor to produce.

2.1 Market Structure

Final goods, G_t, are produced by perfectly competitive firms using inputs from two sources: a non-differentiate intermediate good G_t^N, and a composite of the differentiated goods, G_t^D. Intermediate goods are combined using a standard Constant Elasticity of Substitution (CES) aggregator with elasticity of substitution ϕ. The parameter ϱ determines the degree of preference in production for good N versus D.

$$G_t = \left(\varrho^{1/\phi} \left(G_t^H \right)^{\frac{\phi-1}{\phi}} + (1-\varrho)^{1/\phi} \left(G_t^D \right)^{\frac{\phi-1}{\phi}} \right)^{\frac{\phi}{\phi-1}} \tag{1}$$

From the cost minimization problem of final-goods firms, the final price, P_t, is a weighted composite of prices of N and D, where the price for good D is itself a composite over the prices set by n_t differentiated firms: $P_t = \left(\varrho \left(P_t^N \right)^{1-\phi} + (1-\varrho) \left(P_t^D \right)^{1-\phi} \right)^{\frac{1}{1-\phi}}$. The corresponding demand for each input is standard.

$$G_t^N = \left(\frac{P_t^N}{P_t}\right)^{-\phi} G_t \varrho \tag{2}$$

$$G_t^D = \left(\frac{P_t^D}{P_t}\right)^{-\phi} G_t \left(1-\varrho\right) \tag{3}$$

Demand for differentiated goods is itself a CES composite over the many varieties of intermediates, each indexed by i and given an elasticity of substitution σ.

$$G_t^D = \left(\int_0^{n_t} \left(g_t(i)\right)^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}} \tag{4}$$

In equilibrium we show that the prices will be identical across varieties, following from the assumption that the market for productive resources (capital and labor) are perfectly competitive With n_t firms, we can express the demand for each variety as a function of the number of firms in the market producing at time t. The demand for each variety is:

$$g_t(i) = \left(n_t\right)^{\frac{-\sigma}{\sigma-1}} G_t^D \tag{5}$$

The price index for all varieties is, likewise, a function of the individual firm's prices and the number of firms, $P_t^H = p_t(i)\left(n_t\right)^{\frac{1}{1-\sigma}}$. In the next section we will provide more details about the components that determine the price and the number of firms.

2.2 Consumers

The representative consumer will purchase consumption C_t using their earnings from labor supplied L_t, and their returns from holding domestic deposits with the Bank in the form a one-period non-contingent bond, B_t. Providing labor will earn the consumer a wage of W_t, and savings will earn the consumer a return of R_{t-1}.

Changing the quantity of savings above or below the steady state level of savings, $\bar{B}$, will cost the consumer a quadratic adjustment cost of AC_t. It is necessary to impose these costs to ensure the solution is stable after borrowing and lending is introduced between firms and banks. Lump sum payment also arrive to the consumer from the profits and net assets of exiting banks and firms. We group these payments into a single term, Π_t.

$$C_t \leq W_t L_t + \Pi_t + B_{t-1} R_{t-1} - B_t - AC_t \tag{6}$$

Consumers choose C_t and L_t to maximize their lifetime utility, given each period by the function $U(C_t, L_t)$. Future utility is discounted at a constant rate β.

$$\max_{C_t, L_t} E_t \left\{ \sum_{s=t}^{\infty} \beta^{s-t} U(C_s, L_s) \right\} \tag{7}$$

where we use a standard form of utility for these models that has the properties of constant relative risk aversion, controlled by the parameter γ. Dis-utility from labor is controlled by the parameters ξ and the inverse of the Frische Elasticity of labor supply, $1/\psi$.

$$U(C_t, L_t) = \frac{C_t^{1-\gamma}}{1-\gamma} - \frac{L_t^{1+\psi}}{1+\psi} \tag{8}$$

Letting λ_t represent the consumer's marginal utility, $\lambda_t = C_t^{-\gamma}$, the first order conditions from the consumer problem yield the condition for labor supply,

$$L_t = (W_t \lambda_t)^{1/\psi}. \tag{9}$$

The interest rate on bonds gives a price on the future expected marginal utility. Consumers hold a non-zero balance of bonds in equilibrium as they lend to the Bank, and we let B_t represent the net balance of consumer-Bank borrowing and lending, so that B_t may be positive, negative or zero. To match U.S. data, we calibrate the model so that B_t is positive and represents a savings rate of around 5%.

$$\left(1 + \left[B_t - \bar{B}\right] \psi_b\right) \lambda_t = R_t \beta E_t (\lambda_{t+1}) \tag{10}$$

We use an ad-hoc but commonly used form for bond adjustment costs.

$$AC_t = \frac{\psi_b}{2} \left(B_t - \bar{B}\right)^2 \tag{11}$$

The parameter $\psi_b \geq 0$ governs the size of the cost for adjustment. In the calibrations, we keep this parameter small (on the order of 1E-3) to prevent distortion but provide just enough of a first order presence of B_t to ensure stability of the second order-approximated solution.

2.3 Banks

Banks collect deposits from consumers and invest in the differentiated goods firms through two instruments: equity shares or loans. Shares are simply claims on a portion of next periods profits which are always positive because firms set prices. Total shares sold by each firms $(\theta_t(i))$ are sold at a competitive rate of $p_t^\theta(i)$ and will yield a return of $\theta_t^\theta(i)\left(p_{t+1}^\theta(i)+\pi_{t+1}(i)\right)$ next period. As an alternative to risky equities, banks can also extend one-period non-contingent loans to a firm $(Q_t(i))$ with a certain return of R_t^L. As we discuss below, the model is calibrated so that both loans and equity shares are positive in steady state, that is, both loans and equities appear as assets on the bank's balance sheet, in accordance with U.S. data.

A death shocks arrives at the end of each period so that with probability ξ, a bank continues into the next period and with probability $1-\xi$ they exit to become a consumer again, taking with them their stock of retained earnings. While the total net assets in the banking sector is certainly important for the model, the number of firms is indeterminate. This is a standard result in keeping with the banking literature (Gertler and Kiyotaki (2010),Gertler and Karadi (2011)) and it greatly simplifies the model. Intuitively, the reason lies in the fact that bank services are not differentiated and banks enter without fixed or sunk costs.

Firms by contrast, produce a differentiated good and have sunk costs. Even so, the problem is made less complicated because we show in the next section that every firm sets identical prices and earns identical profits. They sell the same quantity of equity shares and borrow equally. With n_{t+1} firms investing and operating at the start of each period, we can drop the i indexes and write $Q_t = n_{t+1}Q_t(i)$, $p_t^\theta = n_{t+1}p_t(i)^\theta$, $\pi_t = n_{t+1}\pi_t(i)$ and $\theta_t = n_{t+1}\theta_t(i)$. Together with loans from the consumer, these instruments compose the banks' net assets, N_t^B

$$N_t^B = Q_{t-1}R_{t-1}^L + \theta_{t-1}\left(p_t^\theta + \pi_t\right) - B_{t-1}R_{t-1} \tag{12}$$

At the start of period t, the banks then distribute their retained earnings across new investment in loans and equities, while borrowing from consumers in form of deposits.

$$N_t^B = Q_t - \theta p_t^\theta - B_t \tag{13}$$

At the end of each period, after the resolution of uncertainty, banks are confronted with maximizing the lifetime value of their investments. Letting V_t^{bank} be the value of being a bank at the end of period t, every bank solves for the optimal allocations for Q_t and θ_t.

Since banks are owned by the consumer, future earnings are discounted stochastically by the consumer's expected future change in marginal utility, $\Omega_{t+1} = \lambda_{t+1}/\lambda_t$.

$$V_t^{bank} = \max_{Q_t,\theta_t} E_t \Omega_{t+1} \left\{ (1-\xi) N_{t+1}^B + \xi V_{t+1}^{bank} \right\} \tag{14}$$

An incentive compatibility constraint (ICC) limits banks' investments each period. Shown below, the constraint says that the continuing value of the firm must exceed some share of the assets they intend to purchase. If this constraint is not satisfied, the rationale goes, then bank managers have the incentive to abscond with the assets as long as they can escape with portion Γ_t^Q of the loans, or portion Γ_t^θ of the profit shares.

$$V_t \geq Q_t \Gamma_t^Q + \theta_t p_t^\theta \Gamma_t^\theta \tag{15}$$

The bank then solves (14) subject to (15), (12) and (13). We summarize the solution in Result 1, and leave the proof for Appendix A.

Result 1. Solving the bank problem yields the value function as a sum of time varying coefficients v_t^Q, v_t^n, and v_t^θ multiplied by state variables Q_t, N_t, and θ_t.

$$V_t = v_t^Q Q_t + v_t^n N_t^B + v_t^\theta \theta_t p_t^\theta \quad \text{with} \tag{16}$$

$$v_t^Q = E_t \left(\hat{\Omega}_{t+1} \right) \left[R_t^L - R_t \right] \tag{17}$$

$$v_t^n = E_t \left(\hat{\Omega}_{t+1} \right) R_t \tag{18}$$

$$v_t^\theta = E_t \left(\hat{\Omega}_{t+1} \left[\frac{p_{t+1}^\theta + \pi_{t+1}}{p_t^\theta} - R_t \right] \right) \tag{19}$$

$$v_t^Q = \frac{\Gamma_t^Q}{\Gamma_t^\theta} v_t^\theta \tag{20}$$

where the stochastic discount factor for the firm in these expressions is defined as

$$\hat{\Omega}_{t+1} = \Omega_{t+1} \left(1 - \xi + \xi \frac{\Gamma_t^Q v_{t+1}^N}{\Gamma_t^Q - v_{t+1}^Q} \right) \tag{21}$$

Also, the value for the firms' Lagrange multiplier is

$$\lambda_t^{bank} = \frac{Q_t + p_t^\theta \Gamma_t^\theta / \Gamma_t^Q}{N_t^B} \left(\frac{R_t^L - R_t}{R_t} \right) \tag{22}$$

Proof. See Appendix A.

The last expression for the Lagrange multiplier on the banks' investment constraint has a simple counterpart in the data, where $(Q_t + p_t^\theta \theta_t \Gamma_t^\theta / \Gamma_t^Q)/N_t$ is the leverage ratio, and $R_t^L/R_t - 1$ is the spread between loans and deposits.

2.4 Non-Differentiated Goods: G_t^N

Non-differentiated goods are produced using a simple production function that is a linear function of labor, $f(l_t) = al_t$, with the constant technology parameter a . Since the market is competitive among produces, the price for the good is $P_t^N = w_t/a$. Total output of the product, Y_t^N must be equal to total demand, G_t^N. Therefore, the total labor used in production of good N is

$$aL_t^N = Y_t^N = G_t^N \tag{23}$$

These producers hold no assets and make no investments. From a modeling standpoint, this sector is important to pin down the wage so that differentiated goods producers can pay a wage that is different from the marginal product of labor in their production. Hence, producers in D can set a price above marginal costs, as we show below.

2.5 Differentiated Goods: $g_t(i)$

Our treatment of the differentiated goods sector is central component of the model, and differs from the existing literature (Jermann and Quadrini (2012), Bilbiie et al. (2012)) in several key respects. In particular, we impose a unique constraint on firms' investment decisions that depends on their retained earnings, a state variable, rather than on future earnings, a forward looking variable. While the literature has models with financially constrained producers, this is the first paper to impose a constraint based on retained earnings. Besides this constraint, the treatment of firms is largely similar to the existing literature on firm entry and business cycles. Firm's decide to enter when the value of doing so exceeds the benefits of staying a consumer, and firms exit at an exogenous rate.

Each period starts with a mass of firms (n_t) carried over from prior period, with each holding retained earnings from last period, $N_t(i)$.[1] Each firm produces output $y_t(i)$ and will either continue into the next period with probability ψ, or will exit upon the arrival of an idiosyncratic shock with probability $1 - \psi$. Only $\psi n_t = n_{t+1}^o$ firms survive into the

[1] We suppose that there was some n_0 mass of firms in the initial period.

next period, where 'o' stands in for 'old'. At the end of the period, after firms have exited and all uncertainty has been resolved, mass n_t^e firms enter according to a condition that we discuss in Section (2.7) below. Therefore, at the end of period t, there are n_{t+1} firms that will produce next period.

$$n_t = n_t^o + n_t^e \tag{24}$$

Each variety i of the differentiated good is produced by one of the n_t unique firms, and each firm produces $y_t(i)$ each period with a constant-returns-to-scale Cobb-Douglass production technology that uses capital and labor. Factor markets are competitive, so that every firm is subject to the same costs for hiring the quantities of capital and labor they use each period. Also, every period there is an economy-wide shock to this technology, z_t. Therefore each firm produces: $y_t(i) = z_t \left(k_t(i)\right)^{\alpha} \left(l_t(i)\right)^{\alpha}$, where α determines the capital intensity of production. All firms hires capital ($k_t(i)$) and labor ($l_t(i)$) from the stock of available capital K_{t-1} and labor L_t, sold at the competitive price of r_t and W_t, respectively.

With flexible prices and frictionless factor markets, firms choose prices to maximize profits, subject to the demand for each variety, $g_t(i) = \left(\frac{p_t(i)}{P_t^D}\right)^{-\sigma} G_t^D$. The firm's cost minimization problem determines their marginal cost of production, $mc_t = \left(r_t\right)^{\alpha} \left(W_t\right)^{1-\alpha} \alpha^{-\alpha}(1-\alpha)^{\alpha-1}$, and the price paid to capital is determined by the first order optimality condition from the firm, to that we must have $r_t = \frac{\alpha}{1-\alpha} W_t L_t / K_{t-1}$. This pins down the price of capital, together with labor demand from the non-differentiated goods sector.

$$\pi_t(i) = \max_{p_t(i)} \left\{ \left(\frac{p_t(i)}{P_t^D}\right)^{-\sigma} \left[p_t(i) - mc_t\right] G_t^D \right\} \tag{25}$$

Solving this problem gives the standard expression for prices,

$$p_t(i) = mc_t \left(\frac{\sigma}{\sigma - 1}\right),$$

and profits are $\pi_t(i) = \left(p_t(i)\right)^{1-\sigma} \left(P_t^D\right)^{\sigma} G_t^D / \sigma$. The intra-period profit maximization problem is part of firms' longer term goal of maximizing the net assets. Similar to the bank of Section (2.3) the firm will retain earnings that they can invest in assets that earn them a return, as we discuss in the next section.

2.6 Assets for Differentiated Goods Firms

Firms that existed in the prior period, start every period with their stock of retained earnings, $N_t^F(i)$. As with the bank, their current amount available is the sum of their investments from last period, minus the financing costs paid to the bank through interest on loans and dividend payments. Firms have a tax incentive to borrow in bonds, so that for the rate R_t^L charged for bonds, firms only pay $R_t^L(1-\tau)$. In the calibration of the model, τ is essential for calibrating the spread between loans and deposits (given by $R_t^L - R_t$). The key asset for firms is their stock of physical capital, in which they invest each period, earn a return of r_t, and lose share δ et the end of each period to depreciation.

$$N_t^F(i) = K_{t-1}(i)\left[r_t + 1 - \delta\right] - \theta_{t-1}(i)\left(\pi_t(i) + p_t^\theta\right) - B_{t-1}(i)R_{t-1}^L(1-\tau) \tag{26}$$

They are confronted with making an optimal decision for next period by allocating retained earnings. Again, similar to the bank, they face the constraint on investments,

$$N_t^F(i) = K_t(i) - Q_t(i) + \theta_t(i)p_t^\theta. \tag{27}$$

They solve for the capital, equity shares and loans to maximize the lifetime value of net assets.

$$V_t^F(i) = \max_{Q_t(i),\theta_t(i)} E_t\Omega_{t+1}\left\{(1-\psi)N_{t+1}^F(i) + \psi V_{t+1}^F(i)\right\} \tag{28}$$

but their investment in capital, however, is constrained by an incentive compatibility constraint. The intuition here is similar to the banks' constraint in (15), where firm owners are able to abscond with share Γ_t^K of physical assets, and therefore the value of continuing in production must exceed the value of ceasing production and selling off the firm's physical assets.

$$V_t(i)^{firm} \geq \Gamma_t^K K_t(i) \tag{29}$$

The firm then solves (28) subject to (29), (26) and (27). We summarize the solution in Result 2, and leave the proof for Appendix A.

Result 2. Solving the firm's problem yields the value function as a sum of time varying coefficients ρ_t^K, ρ_t^n, and ρ_t^θ multiplied by state variables $K_t(i)$, $N_t^F(i)$, and $\theta_t(i)$.

$$V_t = \rho_t^K K_t(i) + \rho_t^n N_t^F(i) - \rho_t^\theta \theta_t(i) p_t^\theta \tag{30}$$

with

$$\rho_t^K = E_t\left(\breve{\Omega}_{t+1}\left[r_{t+1} + 1 - \delta - R_t^L(1-\tau)\right]\right) \tag{31}$$

$$\rho_t^n = E_t\left(\breve{\Omega}_{t+1}\right) R_t^L(1-\tau) \tag{32}$$

$$\rho_t^\theta = E_t\left(\breve{\Omega}_{t+1}\left[\frac{p_{t+1}^\theta + \pi_{t+1}}{p_t^\theta} - R_t^L(1-\tau)\right]\right) \tag{33}$$

$$\rho_t^\theta = 0 \tag{34}$$

where the stochastic discount factor for the firm in these expressions is defined as

$$\breve{\Omega}_{t+1} = \Omega_{t+1}\left(1 - \psi + \psi\frac{\Gamma_t^K \rho_{t+1}^N}{\Gamma_t^K - \rho_{t+1}^K}\right) \tag{35}$$

Also, the value for the firms' Lagrange multiplier is

$$\lambda_t^{firm} = \frac{\rho^K}{\Gamma_t^K - \rho^K} \tag{36}$$

Proof. See Appendix A.

At least one results follows from this setup: equity prices and loan spreads will move in opposite directions. From the firm problem, taking (33), and (34), we have the price of equities traded between banks and firms, $p_t^\theta = \frac{E_t\left(\frac{\lambda_{t+1}}{\lambda_t}[d_{t+1}+\rho_{t+1}]\right)}{R_t^L}$. Equity prices are decreasing in the rate charged by banks, as this lowers firm profits and directs financing away from dividend payments towards debt servicing.

2.7 Firms Entry, Exit and Aggregating

A mass of n_t^e new firms enter at the end of each period and face identical constraints to existing firms, namely the ICC (in equation 29). They enter with a fixed amount of startup capital priced in labor units, $\omega^s W_t$, that they can use as collateral for leveraging new investments. Startup capital is essential for entering firms, because without it, they could not purchase capital for next period. To see this, the ICC constraint can be re-arranged so that $K_t(i) = N_t(i)\rho_t^n/(\Gamma_t^K - \rho_t^k)$. If $N_t(i) = 0$, so must $K_t(i) = 0$. While they have startup funds, entering firms are also subject to a sunk investment of c^s that is financed out of labor, and only appears in the ICC.

The value function, $V_t^e(i)$, must then satisfy two constraints: one for investing in capital that all firms face, and one that must be satisfied by free entry. After the initial period, the firms' problem resolves to the scenario in Section (2.6).

$$\text{Free entry: } V_t^e(i) = \omega^s W_t + c^s \quad (37)$$

$$\text{Incentive Compatibility: } V_t^e(i) = \Gamma_t^K \left(K_t(i) + c^s\right) \quad (38)$$

A key feature of this setup is that new firms will devote a small *share* of their net assets to capital because they are more constrained. By using the ICC constraint for new firms, and comparing this to the constraint for old firms in (29), we see that new firms invest less by $-\frac{c^s \Gamma_t^K}{\Gamma_t^K - \rho_t^K}$. In effect, sunk costs hinder new entrants from taking on leverage. This is because a sunk cost is an investment that does not also serve as collateral. It therefore causes the ICC constraint to bind more tightly for newly entering firms.

This feature of the model, that new firms start with a lower ratio of capital to net assets, is key to generating high degrees of persistence as the economy responds to shocks. With a large jump in entry, the average ratio of $K_t(i)/N_t(i)$ falls, and can even cause the aggregate stock of K_t to fall, depending on the calibration of the model. The fall in capital per-firm mutes the full impact of the technology shock, and only after many periods, during which the capital ratio is improving, will the result of the shock start to dissipate. Accounting for the added constraints faced by entering firms allows this model to easily account for the hump-shaped response of output, consumption, and firm-entry often seen in the data following technology shocks.

With this feature comes a cost: all firms are not identical. As we show below, we need only track new firms for one period. After the initial period of entry, all firms allocate identical shares of their net assets to $\theta_t(i)$ and $Q_t(i)$. As long as the investment shares are identical, tracking the average level of firm net assets is sufficient to solve the model.

2.8 Aggregating New and Old Firms

To arrive at the total flow of firms' net assets across time, we integrate across all new and existing firms at the end of period. A key feature of the model makes this straightforward–all existing firms choose the same ratio of capital, equity shares and loans when scaled by their level of net assets. In the Appendix A, we show that all firms' net assets can be aggregated into a unified expression describing the evolution of net assets for 'old' firms. Here we use 'bars' to indicate averages, and drop subscripts for variety i. Below is the expression for 'old' firms net assets, $\bar{N}_t^o$. From this value, total net assets are easily found because $n_{t+1}^o \bar{N}_t^o = N_t$

and $n_t\psi = n^o_{t+1}$.

$$\begin{aligned}\bar{N}^o_t =& \psi \bar{N}^o_{t-1} R^L_{t-1}(1-\tau) + \frac{n^e_t}{n^o_t}\omega^s W_{t-1} R^L_{t-1}(1-\tau) \\ & - \bar{\theta}_{t-1} p^\theta_{t-1}\left[\frac{p^\theta_t + \pi_t}{p^\theta_{t-1}} - R^L_{t-1}(1-\tau)\right] \\ & + \left[\bar{K}^e_{t-1}\frac{n^e_t}{n^o_t} + \bar{K}^o_{t-1}\psi\right]\left[r_t + 1 - \delta - R^L_{t-1}(1-\tau)\right]\end{aligned} \tag{39}$$

This equations makes clear the impact of entering firms on the average level of net assets. Since $\bar{K}^e_t < \bar{K}^o_t$ from our discussion around (38), then $\bar{N}^o_{t+1}$ must fall below $\bar{N}^o_t$ when n^e_t is large. Part of this is caused by the assumption that new firms have the same max problem, except for the constraint on capital. For this reason, all firms, new and old, make the same decisions about borrowing through issuing equity shares, $\theta_t(i)$. Therefore, while capital investment is lower for new firms, borrowing by issuing shares is not lower, the net impact on net assets is negative.

3 Market Clearing

In this section we close the economy by discussing the market clearing conditions. First, investment is defined as the difference between the new stock of capital and the depreciated old stock of capital. Also, total capital bought by new and old firms must sum to the total capital stock.

$$\begin{aligned} I_t &= K_t - K_{t-1}(1-\delta) \\ K_t &= n^o_{t+1}\bar{K}^o_t + n^E_{t+1}\bar{K}^e_t \end{aligned}$$

Total final goods output is used in consumption, investment, adjustment costs for entering firms, and adjustment costs for consumer's deposits (bonds).

$$G_t = C_t + I_t + AC_t$$

Total borrowing among firms must equal total lending from banks

$$Q_t = n^o_{t+1}\bar{Q}^o_t + n^e_{t+1}\bar{Q}^e_t$$

and the total shares sold by firms, along with the total dividends issued by firms, must equal total shares bought by banks and the total dividends received by consumers.

$$\theta_t = n_{t+1}\bar{\theta}_t$$

With constant returns to scale, total output in the differentiated goods sector is $Y_t = z_t \left(K_{t-1}\right)^{\alpha} \left(L_t^D\right)^{1-\alpha}$, and the total output must be $y_t(i)n_t = Y_t$. Finally, labor supply must equal labor demand, so that

$$L_t = L_t^D + L_t^N + n_t^e \omega^s$$

where L_t^D is the sum of labor used by all firms, $n_t l_t(i)$.

4 Shocks

As discussed, there are three shocks to the model. The shock to productivity, z_t hits production in the differentiated goods sector. Firms in this sector also suffer shocks γ_t that decrease Γ_t^F in their ICC constraint, effectively reducing their 'trustworthy-ness'. Banks suffer a similar shock to their ICC constraint as they are hit by κ_t. The process for each of theses shocks is,

$$log(z_t) = \rho_z log(z_{t-1}) + \epsilon_{z,t} \tag{40}$$

$$log(\gamma_t) = \rho_\eta log(\gamma_{t-1}) + \epsilon_{\gamma,t} \tag{41}$$

$$log(\kappa_t) = \rho_\kappa log(\kappa_{t-1}) + \epsilon_{\kappa,t} \tag{42}$$

The shocks γ_t and κ_t only hit the economy through the incentive compatibility constraints. Specifically, an increase in either of these shocks makes it easier for firm and bank owners to abscond with a assets.

$$\Gamma_t^Q = \Gamma^Q/\gamma_t$$

$$\Gamma_t^\theta = \Gamma^\theta/\gamma_t$$

$$\Gamma_t^K = \Gamma^Q/\kappa_t$$

The persistence, variance and correlation of shocks are estimated through a standard Bayesian Toolkit using U.S. data on real GDP, investment and total labor supply. We explain the principle results in the section below, and document the data used in the Appendix B at the end of this paper.

5 Estimation and Results

In the calibration of the model, we choose standard values for the parameters commonly found in the RBC literature. The coefficient of relative risk aversion (γ) is set to 2, the elasticity of labor supply is set to 1/1.9, and the discount factor β is set to 1/1.011 to fit the 90 day annualized interest rate on Treasuries. The elasticity of substitution in the differentiated goods sector is $\sigma = 6$ to give a 20% markup, and ϕ (the elasticity of substitution between homogeneous and differentiated goods) is also 6. The tax benefit of borrowing with loans versus equity issuance follows (Jermann & Quadrini, 2012) so that $\tau = 0.35(1 - 1/R^L)$ where R^L is the steady state return on bank loans. The rate of depreciation for capital is $\delta = 0.025$

Parameters ω^s, c^s, Γ_t^Q and Γ_t^θ are chosen to match mean values for (1) the interest rate spread on loans (2) the average bank-leverage ratio of 7.6 in the U.S. (3) the average capital to net asset ratio of U.S. manufacturing firms of 1.45 in the U.S. economy. In particular the steady state loan-deposit interest rate spread in the model is $R^L = R\frac{\Gamma_t^Q - \Gamma_t^\theta}{(1-\tau)\Gamma_t^Q - \Gamma_t^\theta}$. Given that $R_t^L/R_t - 1$ in U.S. data is approximately 0.021% quarterly, and with a given τ, this implies a relationship between Γ_t^θ and Γ_t^Q such that $\Gamma_t^\theta = (1 - \tau - \tau\frac{R}{R^L - R})\Gamma_t^Q$. The exit rates for firms and banks, respectively ψ and ξ are set to match 6% and 2% annual exit rates for firms and banks.

Finally, we estimate using the standard Bayesian Toolkit in Dynare the values governing the stochastic process for technology shocks and financial shocks to banks and firms. Estimated parameters include the standard errors, persistence and correlation of shocks. Table 3 documents the assumptions about priors, and plots in Figure (4) show the results for the estimated shocks.

The key results of the estimated model are presented in the variance decomposition in Table (2) below. The table shows that the driving force of volatility across most variables are shocks to the firm incentive constraint. Perhaps this is not so surprising, since firms own productive capital that is essential for output. To test the robustness of this results, it would be straightforward to test the sensitivity of this decomposition to different shares of capital in production, or by including a consumer durable good or housing sector over which banks have ownership.

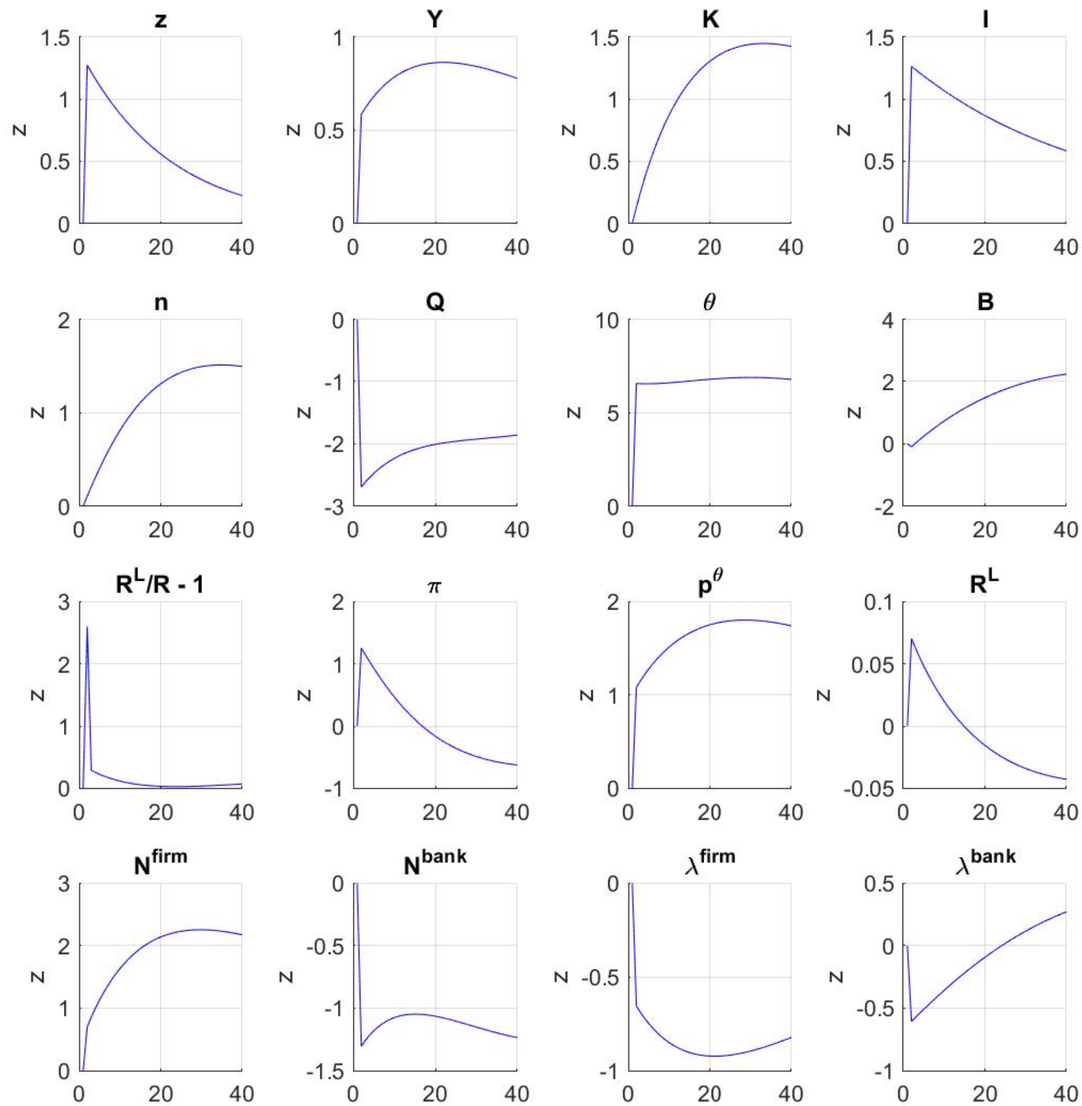
Besides labor, the variable that is most effected by the bank shocks is the choice of capital for existing firms, K_t^o. An apparent puzzle, however, is that the capital stock K_t is driven by firm shocks. The puzzle is easy to resolve by concluding that that much of the variance in the capital stock is coming from new firms, rather than old firms. New firms are especially constrained because they need to finance sunk costs, and are therefore especially sensitive to exogenous changes in the firm constraint.

The impulse responses from the second order approximation of the model are shown in Figures (2), (3) and (1). The shock that is being innovated is placed in the upper left corner of each figure. Among the things to note from these plots is, first, that firm entry is extremely persistent, hump shaped, and accords with intuition. In particular, firms enter after a financial shocks as profits increase and firms enter to satisfy the free entry condition. Firms exit after financial shocks again because the free entry condition–but for at least two distinct reasons. Under bank shocks, the cost of borrowing shoots up, decreasing returns to net assets. Under firm financial shocks, the ICC constraint binds too tightly, keeping firm investment, and therefore profits low, so that the earnings incentive is not there to motivate entry.

Table 1: **Baseline Model Parameters**

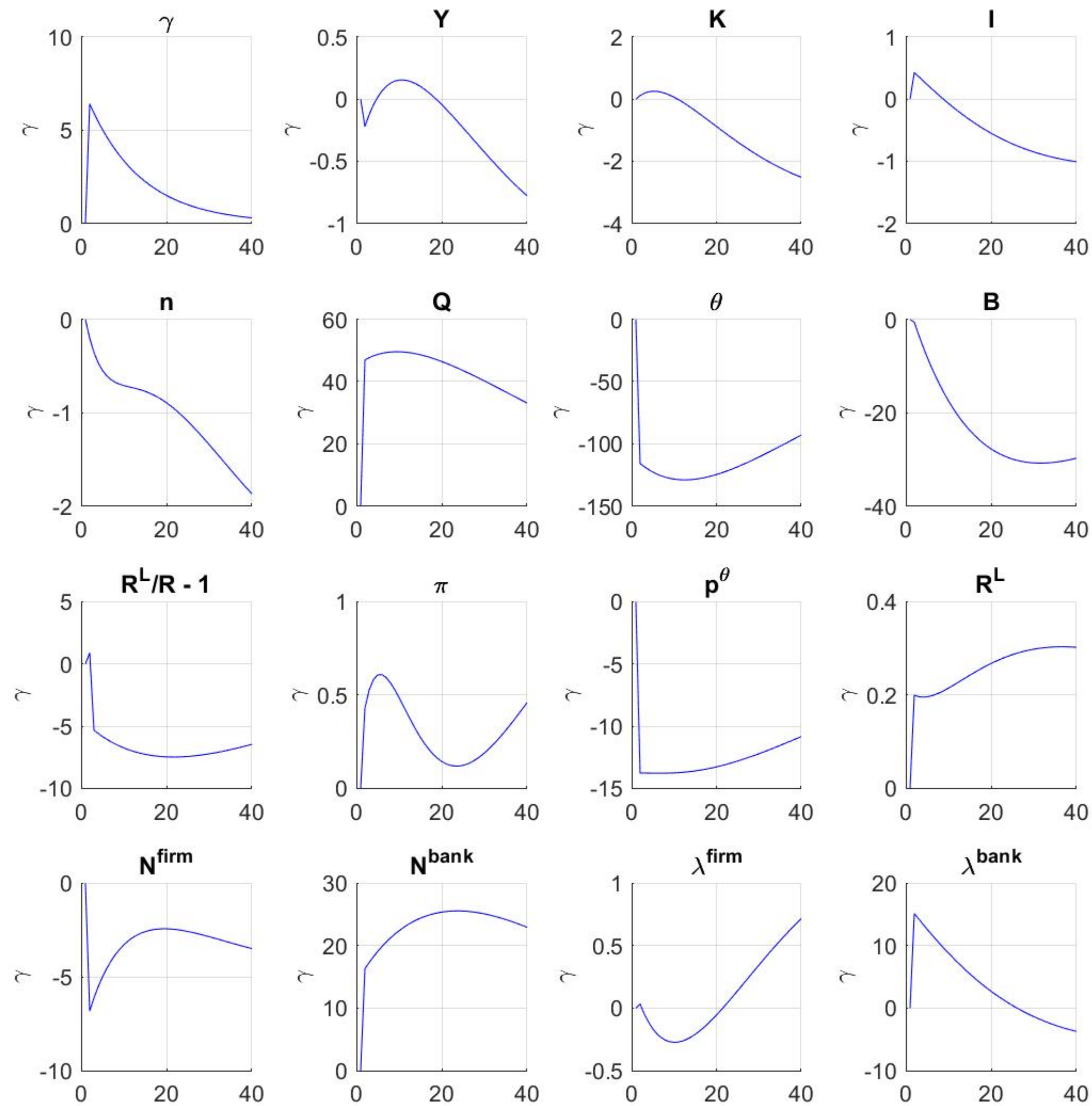
Parameter	**Value**	**Description**
ψ^L	0.5	Frische elasticity of labor supply
β	0.991	Consumer discount factor
γ	2	Coefficient of relative risk aversion
ϕ	6	Elasticity of substitution, final goods
α	0.4	Capital share in production
ϕ	0.3	Preference for non-differentiated goods
σ	6	Elasticity of substitution, differentiated goods
ξ	.981	Persistence rate for banks
ψ	0.4	Persistence rate for firms
Γ^Q	0	Constraint parameter for loans
Γ^θ	0	Constraint parameter for equities
Γ^K	0	Constraint parameter for capital
τ	0.12	Corresponding to $R^L(1-\tau) = 1 + r^L(1-0.35)$
ω^s	0.012	New firm startup capital ($\star$)
c^s	0.025	Sunk entry costs

Figure 1: Technology Shock (z)



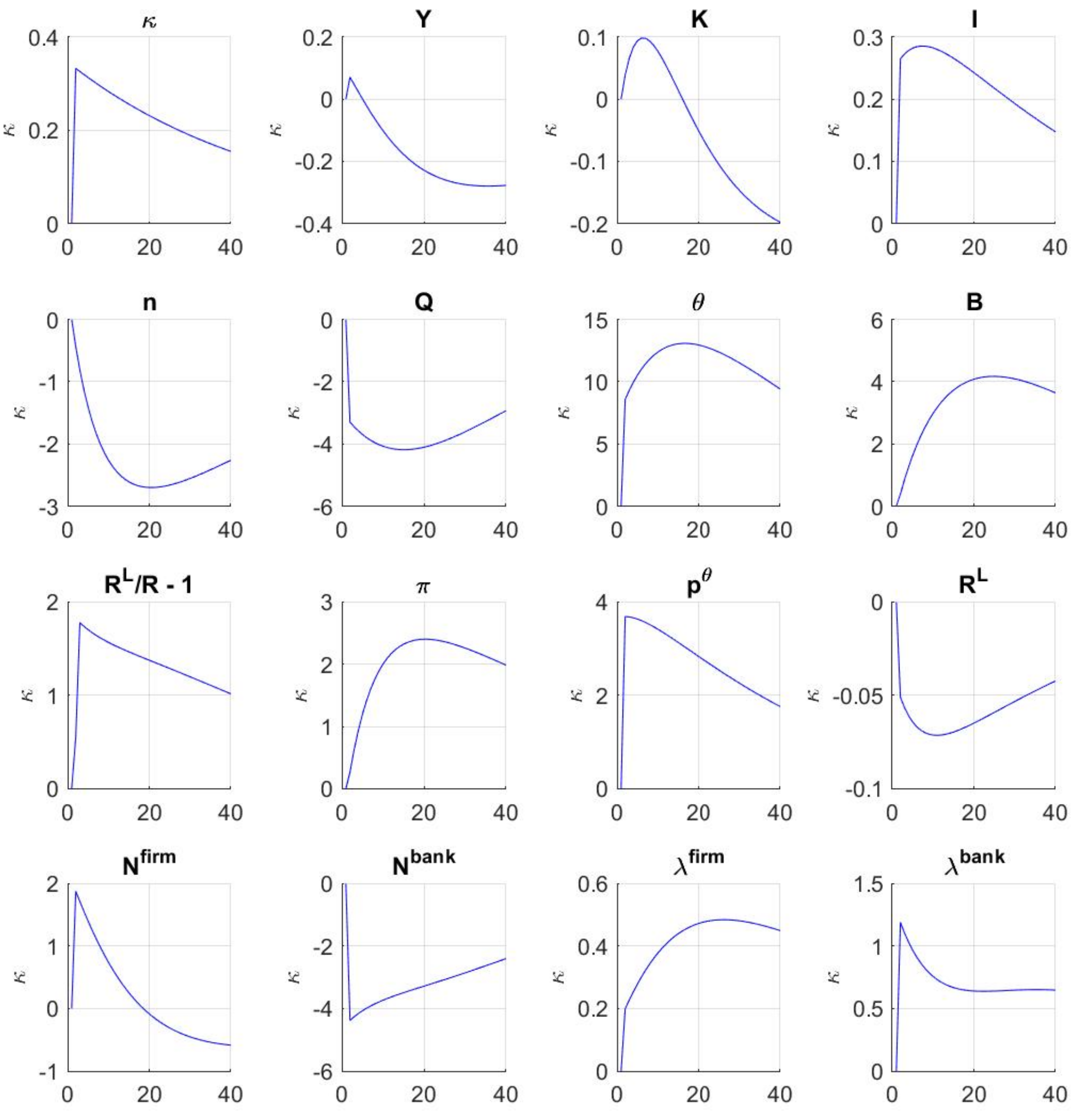
(a) Responses to a 1-standard deviation persistent shock to technology, z, in percent deviations from steady state.

Figure 2: Firm Financial Shock (γ)



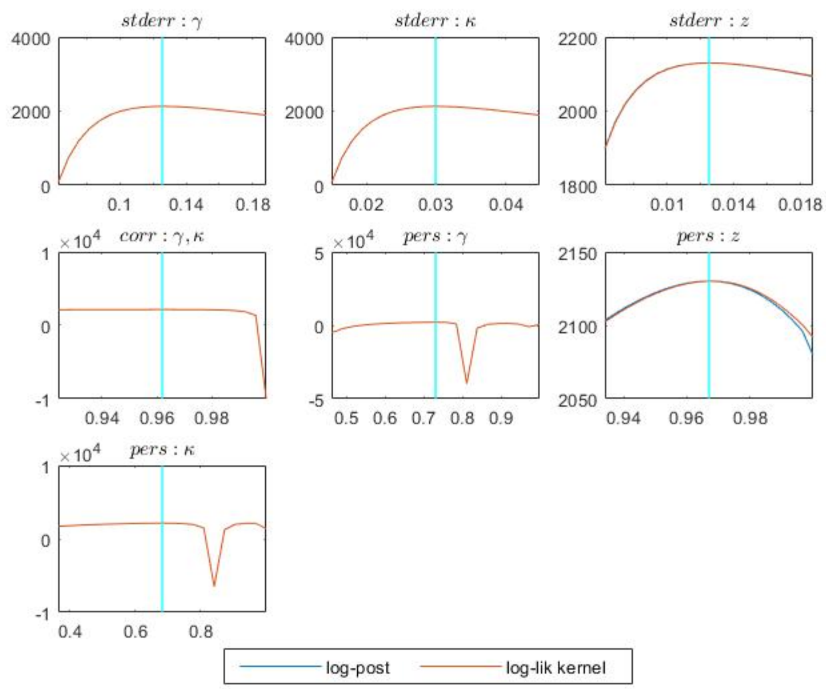
(a) Responses to a 1-standard deviation 1-period shock to bank net assets, γ, in percent deviations from steady state.

Figure 3: Bank Financial Shock (κ)



(a) Responses to a 1-standard deviation 1-period shock to bank net assets, κ, in percent deviations from steady state.

Figure 4: Posterior Results from Estimation



(a) Results from the log-likelihood estimation, where *pers* is the persistence parameter, *corr* is the correlation, and *stderr* is the standard error.

Table 2: **Variance Decomposition of Estimated Model**

Variable	Bank (κ)	Technology(z)	Firm (γ)
GDP	2.7	26.6	70.6
I	1.0	20.6	78.3
C	3.4	27.2	69.3
K	0.7	15.9	83.2
L	60.0	1.4	38.4
N^{firms}	2.2	15.8	81.9
N^{banks}	1.4	0.4	98.0
Q	0.7	0.3	98.9
R	2.5	2.5	94.9
R^L	2.5	2.5	94.9
n	33.6	13.5	52.8
λ^{banks}	1.0	0.5	98.3
λ^{firms}	6.4	2.3	91.2
$B^e(i)$	0.4	0.5	98.9
$B^o(i)$	0.1	0.9	98.9
$K^o(i)$	95.2	0.3	4.4
$K^e(i)$	11.5	10.5	77.8
$N^o(i)$	42.6	4.9	52.4
$R^L/R-1$	1.8	0.2	98.0
$p(i)$	34.3	13.3	52.3
π	63.8	8.1	28.1
p^θ	3.5	2.2	94.2
r	0.9	13.6	85.3
$\theta(i)$	0.5	0.7	98.7
W	6.2	25.3	68.3
$y(i)$	57.5	8.7	33.7

Table 3: Prior information (Parameters)

					Bounds*		90% HPDI	
	Distribution	Mean	Mode	Std.dev.	Lower	Upper	Lower	Upper
$stderr:\gamma$	Inv. Gamma	0.0010	0.0005	Inf	0.0001	56.4190	0.0003	0.0025
$stderr:\kappa$	Inv. Gamma	0.0010	0.0005	Inf	0.0001	56.4190	0.0003	0.0025
$stderr:z$	Inv. Gamma	0.0010	0.0005	Inf	0.0001	56.4190	0.0003	0.0025
$corr:\gamma,\kappa$	Beta	0.3000	0.3844	0.3000	-0.9774	0.9996	-0.2316	0.7530

(Continued on next page)

Table 3: (continued)

	Distribution	Mean	Mode	Std.dev.	Bounds* Lower	Bounds* Upper	90% HPDI Lower	90% HPDI Upper
$pers : \gamma$	Beta	0.8000	0.8462	0.1000	0.1025	0.9999	0.6146	0.9389
$pers : z$	Beta	0.8000	0.8462	0.1000	0.1025	0.9999	0.6146	0.9389
$pers : \kappa$	Beta	0.8000	0.8462	0.1000	0.1025	0.9999	0.6146	0.9389

6 Appendix

6.1 A: Proofs

(1) Proof of Result 1

The bank will solve their problem in (14) subject to their definition for net assets in (**??**), (**??**) and the constraint (15). Starting with a conjecture that the value function is linear, we set up the Lagrangian function,

$$\max_{Q_t,\theta_t} \mathcal{L} = v_t^Q Q_t + v_t^n N_t^B + v_t^\theta \theta_t p_t^\theta + \lambda_t^{bank} \left[v_t^Q Q_t + v_t^n N_t^B + v_t^\theta \theta_t p_t^\theta - \Gamma_t^Q Q_t - \Gamma_t^\theta \theta_t p_t^\theta \right] \tag{43}$$

From the first order conditions we have the value for the multiplier.

$$\lambda_t^{bank} = \frac{v_t^Q}{\Gamma_t^Q - v_t^Q} = \frac{v_t^\theta}{\Gamma_t^\theta - v_t^\theta} \tag{44}$$

From the constraint and the definition of net assets, we can arrive at

$$\frac{N_t^B \Gamma^Q v_t^n}{\Gamma^Q - v_t^Q} = V_t^{bank} \tag{45}$$

First we substitute this for the value function in the banks end-of-period optimization problem. Then using (12) and (13), we substitute N_{t+1}^B into the banks end-of-period optimization problem.

$$
\begin{aligned}
V_t^{bank} &= \max_{Q_t,\theta_t} E_t N_{t+1}^B \Omega_{t+1} \left\{ (1-\xi) + \xi \frac{\Gamma^Q v_{t+1}^n}{\Gamma^Q - v_{t+1}^Q} \right\} && (46) \\
&= \max_{Q_t,\theta_t} E_t N_{t+1}^B \hat{\Omega}_{t+1} && (47) \\
&= E_t \left(\hat{\Omega}_{t+1} \left[Q_t \left[R_t^L - R_t \right] + N_t^B R_t + \theta_t p_t^\theta \left[\frac{\pi_{t+1} + p_{t+1}^\theta}{p_t^\theta} - R_t \right] \right] \right) && (48) \\
&= Q_t E_t \left(\hat{\Omega}_{t+1} \right) \left[R_t^L - R_t \right] + N_t^B E_t \left(\hat{\Omega}_{t+1} \right) R_t + \theta_t p_t^\theta E_t \left(\hat{\Omega}_{t+1} \left[\frac{\pi_{t+1} + p_{t+1}^\theta}{p_t^\theta} - R_t \right] \right) && (49) \\
&= v_t^Q Q_t + v_t^n N_t^B + v_t^\theta p_t^\theta \theta_t && (50)
\end{aligned}
$$

Moving from (46) to (47) defines $\hat{\Omega}_{t+1}$, and moving from (48) to (49) defines v_t^Q, v_t^n and v_t^θ. From (44), we arrive at the expression that pins down p_t^θ in (20). Then by dividing the expression for v_t^Q into v_t^n and using (45), it is straightforward to arrive at (22). ■

(2) Proof of Result 2

The firm's solution follows the same procedure as the banks' solution, so this proof will resemble Result 1 above. Here we solve (28) subject to definition for net assets in (26), (27) and the constraint (29). Starting with a conjecture that the value function is linear, we set up the Lagrangian function,

$$
\max_{K_t(i),\theta_t(i)} \mathcal{L} = \rho_t^K K_t(i) + \rho_t^n N_t^F(i) + \rho_t^\theta \theta_t(i) p_t^\theta + \lambda_t^{firm} \left[\rho_t^K K_t(i) + \rho_t^n N_t^F(i) + \rho_t^\theta \theta_t(i) p_t^\theta - \Gamma^K K_t(i) \right] \quad (51)
$$

From the first order conditions we have the value for the multiplier.

$$
\lambda_t^{firm} = \frac{\rho_t^K}{\Gamma^K - \rho_t^K} \quad (52)
$$

From the constraint and the definition of net assets, we can arrive at

$$
\frac{N_t^F(i) \Gamma^K \rho_t^n}{\Gamma^K - \rho_t^K} = V_t^{firm}(i) \quad (53)
$$

First we substitute this for the value function in the banks end-of-period optimization problem. Then using (26) and (27), we substitute $N^F_{t+1}(i)$ into the banks end-of-period optimization problem.

$$\begin{aligned}
V_t^{firm}(i) &= \max_{K_t,\theta_t(i)} E_t N^F_{t+1}(i)\Omega_{t+1}\left\{(1-\psi)+\psi\frac{\Gamma^K\rho^n_{t+1}}{\Gamma^K-\rho^K_{t+1}}\right\} && (54)\\
&= \max_{K_t,\theta_t(i)} E_t N^F_{t+1}(i)\breve{\Omega}_{t+1} && (55)\\
&= E_t\big(\breve{\Omega}_{t+1}\big[K_t(i)\left[r_{t+1}+1-\delta-R^L_t(1-\tau)\right]+N^F_t(i)R^L_t(1-\tau)- \\
&\qquad \theta_t(i)p^\theta_t\left[\frac{\pi_{t+1}+p^\theta_{t+1}}{p^\theta_t}-R^L_t(1-\tau)\right]\big]\big) && (56)\\
&= K_t(i)E_t\left(\breve{\Omega}_{t+1}\left[r_{t+1}+1-\delta-R^L_t(1-\tau)\right]\right)+N^F_t(i)E_t\left(\breve{\Omega}_{t+1}\right)R^L_t(1-\tau)- \\
&\qquad \theta_t(i)p^\theta_t E_t\left(\breve{\Omega}_{t+1}\left[\frac{\pi_{t+1}(i)+p^\theta_{t+1}}{p^\theta_t}-R^L_t(1-\tau)\right]\right) && (57)\\
&= \rho^K_t K_t(i)+\rho^n_t N^F_t(i)-\rho^\theta_t p^\theta_t\theta_t(i) && (58)
\end{aligned}$$

Moving from (54) to (55) defines $\breve{\Omega}_{t+1}$, and moving from (56) to (57) defines ρ^K_t, ρ^n_t and ρ^θ_t. Additionally, the first order condition with respect to $\theta_t(i)$ yields that $\rho^\theta_t = 0$, which pins down the quantity $\theta_t(i)$. ■

(3) Aggregation of Net Assets

Given first that

$$N^o_t(i) = K^o_{t-1}(i)\left(r_t+1-\delta\right)+\theta_{t-1}(i)\left(\frac{\pi_t+p^\theta_t}{p^\theta_t}-R^L_t(1-\tau)\right)+N^o_{t-1}(i)R^L_{t-1}(1-\tau)$$

and correspondingly for entering firms,

$$N^e_t(i) = K^e_{t-1}(i)\left(r_t+1-\delta\right)+\theta_{t-1}(i)\left(\frac{\pi_t+p^\theta_t}{p^\theta_t}-R^L_t(1-\tau)\right)+\omega_{t-1}R^L_{t-1}(1-\tau)$$

then using these expressions for the definition of total net assets,

$$N_t = \psi\left(\bar{N}^o_t n^o_t+\bar{N}^e_t n^e_t\right)$$

together with the definition for total net assets, $N_t = n^o_{t+1}\bar{N}^o_t$, and $n_t\psi = n^o_{t+1}$ gives the result.

6.2 B: Data

(1) Data

The three series used in estimation procedure for total output, investment and labor supply are provided by the U.S. Bureau of Economic Analysis. Data is logged and linearly de-trended to be stationary and represent percent deviations. The data is as follows:

- GDP is Real Gross Domestic Product in Billions of Chained 2009 Dollars on a quarterly basis from 1947-01-01 to 2018-01-01.
- Investment data is from the series on Billions of Dollars of Quarterly Gross Private Domestic Investment from 1947-01-01 to 2018-01-01.
- Labor supply is an index on a Quarterly basis, from the series: Non-farm Business Sector: Real Output Per Hour of All Persons. Again this data is from 1947-01-01 to 2018-01-01.

References

Alessandria, G., Pratap, S., & Yue, V. Z. (2013). Export dynamics in large devaluations.

Bergin, P. R., Feng, L., & Lin, C.-Y. (2017). Firm entry and financial shocks. *The Economic Journal*, *128*(609), 510–540.

Bilbiie, F. O., Ghironi, F., & Melitz, M. J. (2012). Endogenous entry, product variety, and business cycles. *Journal of Political Economy*, *120*(2), 304–345.

Bilbiie, F. O., Ghironi, F., Melitz, M. J., Midrigan, V., & Rotemberg, J. J. (2007). Monetary policy and business cycles with endogenous entry and product variety [with comments and discussion]. *NBER Macroeconomics Annual*, *22*, 299–379.

Bond, E., Tybout, J., & Utar, H. (2008). Credit rationing, macro volatility, and industrial evolution in developing countries. *International Economic Review, forthcoming*.

Bricongne, J.-C., Fontagné, L., Gaulier, G., Taglioni, D., & Vicard, V. (2012). Firms and the global crisis: French exports in the turmoil. *Journal of international Economics*, *87*(1), 134–146.

Gertler, M., & Karadi, P. (2011). A model of unconventional monetary policy. *Journal of monetary Economics*, *58*(1), 17–34.

Gertler, M., & Kiyotaki, N. (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of monetary economics* (Vol. 3, pp. 547–599). Elsevier.

Hölzl, W. (2005). Tangible and intangible sunk costs and the entry and exit of firms in a small open economy: the case of austria. *Applied Economics*, *37*(21), 2429–2443.

Jermann, U., & Quadrini, V. (2012). Macroeconomic effects of financial shocks. *American Economic Review*, *102*(1), 238–71.

Lewis, V. (2009). Business cycle evidence on firm entry. *Macroeconomic Dynamics*, *13*(5), 605–624.

Manova, K. (2012). Credit constraints, heterogeneous firms, and international trade. *Review of Economic Studies*, *80*(2), 711–744.

Manova, K., Wei, S.-J., & Zhang, Z. (2015). Firm exports and multinational activity under credit constraints. *Review of Economics and Statistics*, *97*(3), 574–588.

Uusküla, L. (2016). Monetary transmission mechanism with firm turnover. *Journal of Macroeconomics*, *50*, 1–18.