The Balance Sheet Channel

Ethan Cohen-Cole
University of Maryland at College Park

Enrique Martínez-García
Federal Reserve Bank of Dallas

We study the role of the balance sheet channel of monetary policy in an environment in which credit plays an important role in the funding of new capital investment. Specifically, we ask whether the transmission mechanism of monetary policy is altered in an environment in which financial intermediation with agency costs, aggregate risk on the performance of loans, and banking regulations are all features that can potentially amplify the impact of shocks over the cycle. Because monetary policy has empirically been asymmetric and marked by periods of pronounced action, our approach provides an alternative plausible mechanism that generates the necessary intuition to account for these patterns. Our model is consistent with current New Neoclassical Synthesis models in good times. In bad times (or crisis periods), when systemic losses are potentially large, the model can generate sharp changes in the external finance premium and, therefore, in the patterns of investment.

To illustrate these phenomena, we posit, from first principles, a model with financial intermediation as well as aggregate risk. We articulate a simple characterization of the link between policy and the real economy that passes through leveraged and regulated financial intermediaries to leveraged borrowers, and we then use the model to explore the role of monetary policy and banking regulation. Our model can provide an intuitive, simple, and micro-founded explanation of the financial accelerator. We also show that basic features of banking regulation like deposit reserve requirements or capital adequacy requirements can amplify the cycle by adding to the costs that entrepreneurs have to pay to borrow from the financial system. Hence, that lends some validity to the argument that banking regulation can help mitigate the effects of crises.

Financial Stability, Monetary Policy, and Central Banking, edited by Rodrigo A. Alfaro, Santiago, Chile.

The first-generation New Neoclassical Synthesis models are not well equipped to interpret the role of monetary policy under financial stress, as the recent crisis illustrated. They were based on a couple of classic imperfections, such as nominal rigidities and monopolistic competition, to allow for nontrivial relative price distortions. The goal, of course, was to illustrate how demand shifts could affect real output, and thus how monetary policy shifting nominal demand could have real effects. These models supported an extensive literature on the basic role of monetary policy, but they omitted details of market imperfections that are central to the questions we explore here. We conjecture that these omissions may be partly responsible for the fact that consensus Taylor rules cannot describe the path of monetary policy (Rudebusch, 2006).

A new round of (second-generation) New Neoclassical Synthesis models focuses on the implications of other frictions. Because of the current financial crisis, a huge number of new papers, this one included, have turned their attention to the role of financial and credit market imperfections by building on work by Bernanke, Gertler, and Gilchrist (1999) and Carlstrom and Fuerst (1997, 2001). In particular, there is renewed interest in a real economy link that passes through the banking sector. This channel is now widely believed to play an important role in the conduct of monetary policy. Our question is how and, specifically, how to model it.

A successful model should be able to accomplish a few key tasks. First, it should be able to characterize monetary policy in both normal and crisis periods. Second, it should do so without relying on ad hoc assumptions on the goals of monetary policy. Third, if indeed there is a financial channel or a banks' balance sheet channel, the model should provide an articulation of how this mechanism operates. Of course, the model should do so without sacrificing many of the gains of research to date in characterizing the paths of other aggregates or, crucially, parsimony.

How do we accomplish this? A combination of bank regulation and systemic risk allows us flexibility in a few important ways. First, the presence of systemic risk provides the framework to motivate state-contingent monetary policy that retains the structure of targeting output and inflation explicitly. Second, a fully described regulated banking sector allows us both to maintain the costly state verification (CSV) framework of Townsend (1979), Gale and Hellwig (1985), and Bernanke, Gertler, and Gilchrist (1999) and to introduce the bank lending channel. This provides an answer to criteria one and three directly. Indeed, we can motivate changes in the pro-inflation response in crisis periods without resorting to ad hoc financial stability targets.

To produce the desired parsimony, we build a variant of the model of Bernanke, Gertler, and Gilchrist (1999) that includes a regulated (but still competitive) banking sector and frictions in the secondary market for used capital. We take this generalization and identify a parsimonious characterization of the external financing premium that intuitively incorporates agency costs due to costly monitoring (costly state verification), as well as the costs of bank regulation on the balance sheet of the financial intermediaries. We show that the external finance premium (EFP) can be represented as follows:

EFP = f (Aggregate Shocks, Agency Cost Channel, Balance Sheet Channel).

We think our approach is useful for two main reasons. First, it reconciles the research agendas that look at stability targeting with those that want a pure monetary policy objective function. Second, it provides a simple and tractable mechanism to explain the financial channel that is consistent both with the banking literature that finds a link between monetary policy and the real economy and with the financial stability literature on the role of capital regulation for monetary policy.[1]

Our approach differs from existing work in a few ways. In one sense, it provides a tractable model through which regulation matters. Unlike models that generate financial channel effects through exogenous spread changes, our model gives an important role to banking intermediation precisely because of the trade-offs present in banking regulation and monetary policy and maintains the view that spreads are at least partly endogenous. Moreover, the model stands on its own because it provides a simple way to think about financial intermediation via leverage and regulatory constraints.

The remainder of the paper is structured as follows. We fully describe the foundations of our model in section 1 and present

1. The literature on this is wide ranging. Bernanke and Lown (1991) argue that the 1992 Basel I deadline contributed to the early 1990s credit crunch, while others suggest that capital regulation generates magnified business cycles. Some relevant papers include Berger and Udell (1994), Blum and Hellwig (1995), Brinkmann and Horvitz (1995), and Thakor (1996). More recent papers include Goodhart, Surinard, and Tsomocos (2006), Estrella (2004), Kashyap and Stein (2004), and Gordy and Howells (2006). See Borio and Zhu (2007) for a comprehensive literature review.

our characterization of the external finance premium in section 2. Moreover, it articulates the intuition of the model for monetary policy and banking regulation. It also discusses a couple of areas for future research, particularly with respect to our characterization of the stance of monetary policy and the banking sector. Section 3 concludes.

1. The Building Blocks of the Model

The financial system is hampered by asymmetries of information between borrowers and lenders and costly state verification, but it is also constrained by regulatory features like capital adequacy and deposit reserve requirements. The economy is populated by a continuum of households and entrepreneurs, each with unit mass. In addition, we include three types of nonfinancial firms (capital goods producers, wholesale producers, and retailers) and one type of financial institution (the banks). All firms, whether financial or nonfinancial, operate under perfect competition, except for the retailers that exploit a monopoly power in their own varieties to add a retail mark-up on their prices. Ownership of all the firms is given to the household, except for wholesale producers who are owned and operated by entrepreneurs.

The banks originate the loans and channel household savings toward the investment needs of the entrepreneurs. The central bank, in turn, has the power to set both banking regulation and monetary policy. Monetary policy is characterized by an interest rate feedback rule in the tradition of Taylor (1993). Banking regulation is summarized in a compulsory reserve requirement ratio on deposits and a capital adequacy requirement on bank capital (or bank equity). The fiscal authority plays a mostly passive role.

In the financial accelerator model of Bernanke, Gertler, and Gilchrist (1999), the relevant friction arises from asymmetric information between entrepreneurs-borrowers and banks-lenders. Monitoring costs make external financing costly for entrepreneurs, so the borrowers' balance sheet conditions play an important role over the business cycle. Otherwise, banks act as a third party inserted between the households and the entrepreneurs whose mission is to intermediate the flow of savings toward investment. In other words, the balance sheet of the lenders that originate the loans becomes passive because loan supply must be equal to the bank deposits demanded by households.

Our benchmark extends the Bernanke-Gertler-Gilchrist model to enhance the role of the banking balance sheet. In particular, we explore the role that banking regulation has on the banks' lending channel and its relevance for monetary policy We also investigate the interaction between banking regulation and monetary policy. We still fit in the tradition of Bernanke, Gertler, and Gilchrist (1999), however, since the basic structure of banking relationships, intermediation, and contract loans is taken as given, rather than arising endogenously, and since we also maintain the illusion of a perfectly competitive banking system. Our model also shares an important characteristic with the framework of Kiyotaki and Moore (1997) in that asset price movements serve to reinforce credit market imperfections.

We depart from Bernanke, Gertler, and Gilchrist (1999) in that banking regulation affects the decisions of banks and, therefore, alters the transmission mechanism in the financial accelerator model. We also depart from their model by introducing systemic (or aggregate) risk on capital income to help us analyze the interest rate spreads, the borrower-lender relationship, and the business cycle dynamics in response to rare or unusual events of large capital income losses.

1.1 Households

There is a continuum of households of unit mass. Households are infinitely lived agents with an identical utility function that is additively separable in consumption, C_t, and labor, H_t. That is,

$$\sum_{\tau=0}^{+\infty} \beta^{\tau} E_t \left[\frac{1}{1-\sigma^{-1}} (C_{t+\tau})^{1-\sigma^{-1}} - \frac{1}{1+\varphi^{-1}} (H_{t+\tau})^{1+\varphi^{-1}} \right], \tag{1}$$

where $0 < \beta < 1$ is the subjective intertemporal discount factor, $\sigma > 0$ ($\sigma \neq 1$) is the elasticity of intertemporal substitution, and $\varphi > 0$ is the Frisch elasticity of labor supply. Households' income comes from renting nonmanagerial labor to the wholesale producers at competitive nominal wages, W_t. It also comes from the ownership of retailers and capital producers, which rebate their total nominal profits (or losses) to them in every period, Π_t^r and Π_t^k, respectively. The unanticipated profits of the banking system are also fully rebated in each period, Π_t^b. Households also obtain their income from interest on their one-period nominal deposits in the banking system, D_t, and from yields on their stake on bank capital, B_{t+1}. With this disposable income, households finance their aggregate consumption, C_t, open new deposits, D_{t+1}, buy new bank shares, B_{t+1}, and pay their nominal (lump-sum) tax bill, T_t.

Accordingly, the households' sequence of budget constraints is described by

$$P_t C_t + T_t + D_{t+1} + B_{t+1} \leq W_t H_t + I_t D_t + (1 - \iota^h) R_t^b B_t + \Pi_t^r + \Pi_t^k + \Pi_t^b, \qquad (2)$$

where I_t is the nominal short-term interest rate offered to depositors, R_t^b is the yield on bank capital, and P_t is the consumer price index (CPI). The nominal tax on bank equity, ι^h, is a convenient simplification to capture the differential tax treatment of capital gains from equity holdings and deposits in many tax codes around the world. As a matter of convention, D_{t+1} and B_{t+1} denote nominal deposits and bank equity held from time t to $t+1$. Therefore, the interest rate I_{t+1} paid at $t+1$ is known and determined at time t, but the yield on bank equity R_{t+1}^b could potentially depend on the state of the world at time $t+1$. Household optimization yields the standard first-order conditions for consumption-savings and labor supply,

$$\frac{1}{I_{t+1}} = \beta E_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\sigma^{-1}} \frac{P_t}{P_{t+1}} \right], \qquad (3)$$

$$1 = \beta E_t \left[(1 - \iota^h) R_{t+1}^b \left(\frac{C_{t+1}}{C_t} \right)^{-\sigma^{-1}} \frac{P_t}{P_{t+1}} \right], \qquad (4)$$

and

$$\frac{W_t}{P_t} = (C_t)^{\sigma^{-1}} (H_t)^{\varphi^{-1}}, \qquad (5)$$

plus the appropriate no-Ponzi transversality condition. It also implies that each period, budget constraint holds with equality.

As we discuss below, the problem of the banks is such that the yield on bank capital is also known and determined at time t. By simple arbitrage between equations (3) and (4), it follows that (1−

ι^h) $R^b_{t+1} = I_{t+1}$ is necessary for an interior solution to exist (where households hold both bank deposits and bank equity).

1.2 Retailers

There is a continuum of retail firms of unit mass. The retail sector transforms wholesale output into differentiated goods using a linear technology. For simplicity, we assume that no capital or labor is needed in the retail sector, so the wholesale good is the only input of production. Each retail variety is then sold to households, entrepreneurs, and capital goods producers, and bundled up for either consumption or investment (only capital goods producers acquire these varieties for investment purposes). The retailers add a brand name to the wholesale good to introduce differentiation. Variety is valued by all potential consumers, so retailers gain monopolistic power to charge a retail mark-up on them.

1.2.1 Aggregation

We denote the differentiated varieties as $Y_t(z)$, where the index $z \in [0,1]$ identifies each individual retailer. Final goods used for consumption and investment, Y_t, are bundles of these differentiated varieties, $Y_t(z)$, aggregated by means of a common constant elasticity of substitution (CES) index, as follows:

$$Y_t = \left[\int_0^1 Y_t(z)^{\frac{\theta-1}{\theta}} dz \right]^{\frac{\theta}{\theta-1}}. \tag{6}$$

The elasticity of substitution across varieties is represented by $\theta > 1$. The corresponding consumption price index (CPI) is given by

$$P_t = \left[\int_0^1 P_t(z)^{1-\theta} dz \right]^{\frac{1}{1-\theta}}, \tag{7}$$

where $P_t(z)$ is the price charged by retailer z for its variety. The optimal allocation of expenditure to each variety, that is,

$$Y_t(z) = \left[\frac{P_t(z)}{P_t} \right]^{-\theta} Y_t, \tag{8}$$

implies that retailers face a downward-sloping demand function.

1.2.2 Optimal pricing

Retailers set prices to maximize profits, but their ability to reoptimize is constrained because they face nominal rigidities à la Calvo (1983). The retailer maintains its previous period price with an exogenous probability $0 < \alpha < 1$ in each period. However, with probability $(1 - \alpha)$, the retailer is allowed to optimally reset its price. Whenever reoptimization is possible, a retailer z chooses its price, $\tilde{P}_t(z)$, to maximize the expected discounted value of its net nominal profits, that is,

$$\sum_{\tau=0}^{+\infty} E_t \left\{ \alpha^{\tau} M_{t,t+\tau} \tilde{Y}_{t,t+\tau}(z) \left[\tilde{P}_t(z) - (1-\iota^r) P^w_{t+\tau} \right] \right\}, \tag{9}$$

where

$$M_{t,t+\tau} \equiv \beta^{\tau} \left(\frac{C_{t+\tau}}{C_t} \right)^{-\sigma^{-1}} \frac{P_t}{P_{t+\tau}},$$

is the household's stochastic discount factor (SDF) for τ-periods-ahead nominal payoffs, $P^w_{t+\tau}$ is the nominal price of wholesale goods, and

$$\tilde{Y}_{t,t+\tau}(z) = \left[\frac{\tilde{P}_t(z)}{P_{t+\tau}} \right]^{-\theta} Y_{t+\tau},$$

is the demand at time $t + \tau$ given that prices remain fixed at $\tilde{P}_t(z)$ (see equation 8). We also include a subsidy on inputs for retailers, ι^r, which is used by the government to eliminate the retail mark-up distortion whenever $\iota^r = 1/\theta$.

The solution to the retailer's maximization problem satisfies the following first-order condition:

$$\sum_{\tau=0}^{+\infty} E_t \left\{ (\alpha\beta)^{\tau} \left(\frac{C_{t+\tau}}{C_t} \right)^{-\sigma^{-1}} \tilde{Y}_{t,t+\tau}(z) \left[\frac{\tilde{P}_t(z)}{P_{t+\tau}} - \frac{\theta(1-\iota^r)}{\theta-1} \frac{P^w_{t+\tau}}{P_{t+\tau}} \right] \right\} = 0, \tag{10}$$

where $\theta/(\theta - 1)$ denotes the retail mark-up, and P^w_t/P_t denotes the price of wholesale output in units of consumption. The latter provides a measure for the real marginal costs before the government subsidy. The first-order condition in equation (10) is often referred to as the price-setting rule. Given that a fraction α of retailers maintains prices in period t and that all reoptimizing retailers face a symmetric problem, the aggregate CPI in equation (7) can be rewritten in the following terms:

$$P_t = \left[\alpha P_{t-1}^{1-\theta} + (1-\alpha)\tilde{P}_t\ (z)^{1-\theta}\right]^{\frac{1}{1-\theta}}, \tag{11}$$

where $\tilde{P}_t(z)$ is the (symmetric) optimal price implied by equation (10).

Technically, there is no aggregate production function for the final output, Y_t. However, there is a simple way to account for the distribution of resources. By market clearing, the sum of the individual retailers' demands of the wholesale good has to be equal to the total production of the wholesale producers, that is,

$$\int_0^1 Y_t\,(z)dz = Y_t^w. \tag{12}$$

Using the optimal allocation of expenditure in equation (8), we get

$$Y_t = \left(\frac{P_t^*}{P_t}\right)^{\theta} Y_t^w, \tag{13}$$

where

$$P_t^* \equiv \left[\int_0^1 P_t(z)^{-\theta} dz\right]^{-\frac{1}{\theta}} = \left[\alpha(P_{t-1}^*)^{-\theta} + (1-\alpha)\tilde{P}_t(z)^{-\theta}\right]^{-\frac{1}{\theta}}. \tag{14}$$

The term $p_t^* \equiv (P_t^* / P_t)^{\theta} \leq 1$ characterizes the magnitude of the efficiency distortion due to sticky prices.

Since households own the retailers, we assume that all profits (or losses) from the retail activity are rebated to the households as a lump sum in every period. After a bit of algebra, the aggregate nominal profits received by the households can be computed as

$$\begin{aligned} \Pi_t^r &= \int_0^1 \left\{Y_t(z)\left[P_t(z) - (1-\iota^r)P_t^w\right]dz\right\} \\ &= P_t\left(\frac{P_t^*}{P_t}\right)^{\theta} Y_t^w - (1-\iota^r)P_t^w Y_t^w, \end{aligned} \tag{15}$$

where the second equality follows from the optimal allocation of expenditure in each variety described in equation (8), the aggregation formulas in equations (6) and (7), and the relationship between final output and wholesale output implied by equation (12).

1.3 Capital Goods Producers

There is a continuum of capital goods producers of unit mass. At time t, these producers combine aggregate investment goods, X_t, and depreciated capital, $(1-\delta)K_t$, to manufacture new capital goods, K_{t+1}. The production of new capital is limited by technological constraints. We assume that the aggregate stock of new capital evolves according to the following law of motion:

$$K_{t+1} \leq (1-\delta)K_t + \Phi\,(X_t, X_{t-1}, K_t)\,X_t, \tag{16}$$

where X_t is real aggregate investment, K_t stands for real aggregate capital, and $0 < \delta < 1$ is the depreciation rate. The function $\Phi(X_t, X_{t-1}, K_t)$ implicitly characterizes the technology available to the capital goods producers to transform investment goods into new capital.

We explore three different specifications of the technological constraint. The neoclassical adjustment case (NAC) assumes that the transformation of investment goods into new capital can be attained at a one-to-one rate:

$$\Phi\,(X_t, X_{t-1}, K_t) = 1. \tag{17}$$

The specification for the so-called capital adjustment case (CAC), favored by Bernanke, Gertler, and Gilchrist (1999) and others, takes the following form:

$$\Phi\left(\frac{X_t}{K_t}\right) = 1 - \frac{1}{2}\chi\frac{\left[(X_t/K_t)-\delta\right]^2}{(X_t/K_t)}, \tag{18}$$

where X_t/K_t denotes the investment-to-capital ratio. Finally, the investment adjustment case (IAC), preferred by Christiano, Eichenbaum, and Evans (2005) takes the following form:

$$\Phi\left(\frac{X_t}{X_{t-1}}\right) = 1 - \frac{1}{2}\kappa\frac{\left[(X_t/X_{t-1})-1\right]^2}{(X_t/X_{t-1})}, \tag{19}$$

where X_t/X_{t-1} denotes the gross investment growth rate. The parameters $\chi > 0$ and $\kappa > 0$ regulate the degree of concavity of the technological constraint and, therefore, the sensitivity of investment in new capital. In steady state, the CAC function satisfies that $\Phi(\delta) = 1$, $\Phi'(\delta) = 0$, and $\Phi''(\delta) = -(\chi/\delta) < 0$. Similarly, the IAC function satisfies that $\Phi(1) = 1$, $\Phi'(1) = 0$, and $\Phi''(1) = -\kappa < 0$.

Capital goods producers choose their investment demand, X_t, and their output of new capital, K_{t+1}, to maximize the expected discounted value of their net profits:

$$\sum_{\tau=0}^{+\infty} E_t \left\{ M_{t,t+\tau} P_{t+\tau} \left[Q_{t+\tau} K_{t+\tau+1} - (1-\delta) \bar{Q}_{t+\tau} K_{t+\tau} - X_{t+\tau} \right] \right\}, \tag{20}$$

subject to the law of motion for capital described in equation (16). Here,

$$M_{t,t+\tau} \equiv \beta^{\tau} \left(\frac{C_{t+\tau}}{C_t} \right)^{-\sigma^{-1}} \frac{P_t}{P_{t+\tau}},$$

is the household's stochastic discount factor for τ-periods-ahead nominal payoffs, since households own the capital goods producers. As a matter of convention, K_{t+1} denotes the real stock of capital built (and determined) at time t for use at time $t + 1$.

The investment good is bundled in the same fashion as the consumption good and is bought at the same price, P_t. The depreciated capital is bought at a resale price, $\bar{Q}_t$, in units of the consumption good. However, the new capital is sold to the entrepreneurs at a price Q_t, which determines the relative cost of investment in unit of consumption and is often referred to as Tobin's Q. We assume that frictions in the secondary market for used capital prevent arbitrage between the resale value of old capital and the sale value of new capital, that is, $\bar{Q}_t = o_t Q_t$ where $o_t \neq 1$. Those frictions are left unmodeled, although we also assume that the parties involved in the secondary market (namely, entrepreneurs and capital goods producers) view them as entirely out of their control. Hence, they treat the wedge, o_t, as an exogenous and random shock.

Moreover, there is no centralized market that ensures a uniform pricing for used capital, so each individual entrepreneur and capital producer pair matched in the secondary market gets a different draw of this random wedge. In other words, o_t is modeled not as an aggregate shock, but as an idiosyncratic one. Nonetheless, we map this resale shock into the Bernanke-Gertler-Gilchrist framework as closely as possible. That keeps our departure from their model to a minimum, but requires us to note that the wedge, o_t , has a component that depends on other endogenous variables that have an influence on the capital returns that the entrepreneurs can generate.

The optimization of the capital goods producers yields a standard first-order condition that determines the linkage between Tobin's Q, Q_t, and investment, X_t, that is,

$$Q_t\left[\Phi(X_t, X_{t+1}, K_t) + \frac{\partial\Phi(X_t, X_{t-1}, K_t)}{\partial X_t} X_t\right] + \beta\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma^{-1}} Q_{t+1}\frac{\partial\Phi(X_{t+1}, X_t, K_{t+1})}{\partial X_t} X_{t+1}\right] = 1, \tag{21}$$

which does not depend on the wedge, o_t. The law of motion for capital is binding in each period. Given our alternative specifications of the technological constraint, we could rewrite the first-order condition in equation (21) more compactly as follows:

$$\begin{cases} Q_t = 1, \quad \text{if NAC;} \\ Q_t\left[\Phi\left(\frac{X_t}{K_t}\right) + \Phi'\left(\frac{X_t}{K_t}\right)\frac{X_t}{K_t}\right] = 1, \quad \text{if CAC;} \\ Q_t\left[\Phi\left(\frac{X_t}{X_{t-1}}\right) + \Phi'\left(\frac{X_t}{X_{t-1}}\right)\frac{X_t}{X_{t-1}}\right] = 1 + \beta\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\sigma^{-1}} Q_{t+1}\Phi'\left(\frac{X_{t+1}}{X_t}\right)\left(\frac{X_{t+1}}{X_t}\right)^2\right], \text{ if IAC.} \end{cases} \tag{22}$$

The neoclassical adjustment case is of particular interest because without the asset price fluctuations captured by Tobin's Q, the Bernanke-Gertler-Gilchrist framework loses the characteristic that asset price movements serve to reinforce credit market imperfections. For more details on the derivations of the Tobin's Q equations, see Martínez-García and Sondergaard (2008).

Profits (or losses) may arise since X_{t-1} and K_t are predetermined at time t and cannot be adjusted freely. The aggregate profits at each point in time for the capital goods producers, that is,

$$\begin{aligned} \prod\nolimits_t^k &\equiv P_t Q_t K_{t+1} - (1-\delta) P_t \left[\int_0^1 \overline{Q}_t \mu_t^o(o_t)\, do_t\right] K_t - P_t X_t \\ &= P_t Q_t \Phi(X_t, X_{t-1}, K_t) X_t - \left[\int_0^1 o_t \mu_t^o(o_t)\, do_t - 1\right](1-\delta) P_t Q_t K_t - P_t X_t, \end{aligned} \tag{23}$$

must be added to the households' budget constraint (since households are their only shareholders). Here, $\mu_t^o(o_t)$ denotes the mass of capital goods producers receiving a given realization of the idiosyncratic shock, o_t.

1.4 Wholesale Producers

There is a continuum of mass one of wholesale producers. Wholesale producers combine the nonmanagerial labor provided by the households with the managerial labor supplied and the capital rented from the entrepreneurs to produce wholesale goods according to the following Cobb-Douglas technology:

$$Y_t^w \leq e^{a_t}(K_t)^{1-\psi-\varsigma}(H_t)^{\psi}(K_t^e)^{\varsigma}, \tag{24}$$

where Y_t^w is the output of wholesale goods, K_t is the aggregate capital rented, and H_t and H_t^e are the demands for nonmanagerial and managerial labor, respectively.

With a constant returns-to-scale technology, the nonmanagerial and managerial labor shares in the production function are determined by the coefficients $0 < \psi < 1$ and $0 < \varsigma < 1$. In keeping with Bernanke, Gertler, and Gilchrist (1999), the managerial share is often assumed to be very small, that is, ς would be close to zero. The productivity shock, a_t, allows a first-order autoregressive, or AR(1), process of the following form:

$$a_t = \zeta_a a_{t-1} + \varepsilon_t^a, \tag{25}$$

where ε_t^a is a zero mean, uncorrelated, and normally distributed innovation. The parameter $-1 < \zeta_a < 1$ determines the persistence of the productivity shock, and $\sigma_a^2 > 0$ the volatility of its innovation.

Wholesale producers maximize their static profit:

$$\prod\nolimits_t^w \equiv P_t^w Y_t^w - R_t^w K_t - W_t H_t - W_t^e H_t^e, \tag{26}$$

subject to the technological constraint implied by equation (24). Wholesale producers rent labor from households and entrepreneurs at competitive nominal wages W_t and W_t^e, respectively, and they compensate the entrepreneurs with a nominal return per unit of capital rented, R_t^w. The optimization of the wholesale producers results in the following well-known rules to compensate the factors of production:

$$R_t^w = (1-\psi-\zeta)\frac{P_t^w Y_t^w}{K_t}; \tag{27}$$

$$W_t = \psi\frac{P_t^w Y_t^w}{H_t}; \tag{28}$$

$$W_t^e = \zeta\frac{P_t^w Y_t^w}{H_t^e}. \tag{29}$$

The optimization of the wholesale producer can be summarized in these first-order conditions plus the technological constraint in equation (24) holding with equality. Wholesale producers make zero profits in every period (that is, $\Pi_t^w = 0$), so the entrepreneurs who own them do not receive any dividends. All the income entrepreneurs extract comes from their supply of two key inputs in the production function—managerial labor and, especially, capital. Wholesale producers rent the capital they use from the entrepreneurs and return the depreciated capital after production has taken place.

As we discuss shortly, uncertainty about the resale value of depreciated capital is the underlying risk that distorts the relationship between borrowers (the entrepreneurs) and lenders (the banks). In fact, asymmetries of information on this type of risk and costly state verification lead to a distorted allocation of households' savings toward the productive capital investments operated by the entrepreneurs.

1.5 Entrepreneurs

There is a continuum of entrepreneurs of unit mass. Entrepreneurs are infinitely lived agents with identical preferences that are linear in consumption, C_t^e:

$$\sum_{\tau=0}^{\infty} (\beta\eta)^{\tau} E_t \left(C_{t+\tau}^e \right), \tag{30}$$

where $0 < \beta\eta < 1$ is the subjective intertemporal discount factor. Entrepreneurs inelastically supply one unit of managerial labor:

$$H_t^e = 1, \quad \forall t. \tag{31}$$

The entrepreneurs' utility function also differs from that of the households because they are risk neutral (linear utility), and they discount utility at a higher rate (that is, $0 < \eta < 1$). The relative impatience is intended to ensure that entrepreneurs never save enough resources to overcome their financing constraints. The assumption of risk neutrality implies that entrepreneurs care only about expected returns, which considerably simplifies the financial contract (see the appendix).

At the end of period t, the entrepreneur receives a competitive nominal wage, W_t^e, and earns income from the capital rented at the beginning of the period for the production of wholesale goods, $R_t^w K_t$, as well as from the resale value on the depreciated capital bought by the capital goods producers, $(1-\delta)P_t\bar{Q}_t K_t$.[2] After repaying their outstanding

2. Distortions in the secondary market create a random wedge between the acquisition cost of new capital and the resale value of old capital in each period.

loans to the banking system, L_t, entrepreneurs can appropriate a fraction of the aggregate capital income, that is, a share of $R_t^w K_t + (1-\delta)P_t\bar{Q}_t K_t$. The entrepreneurs own the wholesale producers, but these firms generate zero profits after paying for the factors of production and, therefore, produce no dividends for the entrepreneurs.

Using the resources coming from managerial wages and capital rental rates, the entrepreneurs must buy the new capital, K_{t+1}, and decide how much to consume, C_t^e. New capital is needed for the production of wholesale goods at time $t + 1$. Net of consumption, the entrepreneurs set aside a portion of their income in the form of entrepreneurial net worth, N_{t+1}. Entrepreneurial net worth is, in effect, a form of savings for the entrepreneur that can be applied partly to acquire new capital. The entrepreneurs use these savings, N_{t+1}, as well as external loans from the banking system, L_{t+1}, to fund the acquisition of the entire stock of new capital, $P_t Q_t K_{t+1}$:

$$P_t Q_t K_{t+1} = N_{t+1} + L_{t+1}. \tag{32}$$

Equation (32) also implies that new capital is the only asset in which entrepreneurs can invest their savings. As in Bernanke, Gertler, and Gilchrist (1999), we rule out a more complex portfolio setting for entrepreneurs.

1.5.1 Idiosyncratic and anticipated systemic risk

We define the returns on capital relative to its acquisition cost whenever the resale value of capital and the cost of new capital are equalized as

$$R_t^e \equiv \frac{R_t^w K_t + (1-\delta)P_t Q_t K_t}{P_{t-1}Q_{t-1}K_t}.$$

For an individual entrepreneur, we define the returns on the capital that was acquired at time $t - 1$, $\omega_t R_t^e$, as the total income generated by a unit of capital at time t after accounting for the effects of the distortion in the secondary market:[3]

3. To be more precise, we define the rate of return on capital, R_t^e, as the rate that would prevail if the secondary market for used or depreciated capital led to arbitrage between the resale value of capital and the cost of acquiring new capital, that is, $\bar{Q}_t = Q_t$. The returns on capital are realized under distortions in the secondary market, so the actual rate of return on capital is $\omega_t R_t^e$, as defined in equation (33). For convenience, we implicitly capture the randomness of the wedge in the resale value, o_t, by positing that ω_t is a purely exogenous random variable.

$$\omega_t R_t^e \equiv \frac{R_t^w K_t + (1-\delta) P_t \bar{Q}_t K_t}{P_{t-1} Q_{t-1} K_t} = \left[\frac{(R_t^w / P_t) + (1-\delta) o_t Q_t}{Q_{t-1}} \right] \frac{P_t}{P_{t-1}}, \tag{33}$$

where the rental rate on capital, R_t^w, is defined in equation (27). Returns on capital are subject to idiosyncratic shocks, ω_t, which reflect the impact of the random resale distortion:

$$o_t \equiv O\left(\omega_t, \frac{R_t^w}{P_t Q_t} \right).$$

The function that links the wedge on the secondary market, o_t, to the idiosyncratic shock, ω_t, can be expressed as

$$\begin{aligned} o_t \equiv O\left(\omega_t, \frac{R_t^w}{P_t Q_t} \right) &= \frac{\omega_t R_t^e P_{t-1} Q_{t-1} K_t - R_t^w K_t}{(1-\delta) P_t Q_t K_t} \\ &= \omega_t + (\omega_t - 1) \frac{R_t^w}{(1-\delta) P_t Q_t}, \end{aligned} \tag{34}$$

where the second equality follows from the definition of R_t^e.

We interpret the shock $\omega_{t+1} \in (0, +\infty)$ as a reduced-form representation of the exogenous losses on the resale value of the depreciated capital due to frictions in the secondary market. Those frictions, which are left unmodeled, imply a wedge between the resale value of capital and the acquisition cost of new capital (of Tobin's Q) within the period. We denote $\phi(\omega_{t+1} | s_{t+1})$ the density and $\Phi(\omega_{t+1} | s_{t+1})$ the cumulative distribution of ω_{t+1} conditional on a given realization of the aggregate shock s_{t+1}.

We assume that the expected capital return of each entrepreneur is a function of the aggregate shock s_{t+1} (for example, Faia and Monacelli, 2007). The aggregate shock s_{t+1} captures our notion of systemic risk on the resale value of depreciated capital, which has the effect of shifting the mean of the distribution of the risky capital returns. The systemic risk shock, s_t, follows an AR(1) process of the following form:

$$s_t = \rho_s s_{t-1} + \varepsilon_t^s, \tag{35}$$

where ε_t^s is a zero mean, uncorrelated, and normally distributed innovation. The parameter $-1 < \rho_s < 1$ determines the persistence of the systemic shock, and $\sigma_s^2 > 0$ the volatility of its innovation.

We assume that the realization of the time $t+1$ shock is publicly observed at time t. Therefore, these systemic shocks are interpreted as anticipated (rather than unanticipated) losses.

The expected idiosyncratic shock on capital income, ω_{t+1}, conditional on the realization of the aggregate shock, s_{t+1}, is given by

$$E_t(\omega_{t+1}|s_{t+1}) = 1 - J(s_{t+1}), \tag{36}$$

where $0 \leq \lambda \equiv J(0) < 1$ determines the level of the expected losses in steady state, and $-\infty < \xi \equiv J'(0) < +\infty$ characterizes the sensitivity of the expected losses. This specification is flexible enough to allow for catastrophic losses due to a sizeable systemic risk shock s_{t+1}. By choosing λ sufficiently close to zero, we ensure that the expected idiosyncratic shock remains relatively close to one most of the time, that is, $E_t(\omega_{t+1}|s_{t+1}) \cong 1$. That means that entrepreneurs get, on average, a capital return that is approximately equal to R_t^e, which is what is expected whenever the acquisition cost and the resale value of capital are equalized within each period.[4]

1.5.2 The loan contract

At time t, the entrepreneurs-borrowers and the banks-lenders must agree on a contract that facilitates the acquisition of new capital,

4. Given the characterization of the idiosyncratic shock ω_t in equation (33) and the definition of the capital return under equalization between the resale value of capital and the acquisition cost, R_t^e, we can argue that the expected or average value of depreciated capital is equal to

$$\begin{aligned}(1-\delta)P_tQ_tK_t\left[\int_0^1 o_t\mu_t^o(o_t)do_t\right] &= \left[\int_0^{+\infty}\omega_t\phi(\omega_t \mid s_t)d\omega_t - 1\right]R_t^wK_t \\ &+ \left[\int_0^{+\infty}\omega_t\phi(\omega_t \mid s_t)d\omega_t\right](1-\delta)P_tQ_tK_t \\ &= (1-\delta)P_tQ_tK_t - J(s_t)\left[R_t^wK_t + (1-\delta)P_tQ_tK_t\right] \\ &= (1-\delta)P_tQ_tK_t - P_{t-1}Q_{t-1}K_tR_t^eJ(s_t),\end{aligned}$$

where we use the fact that $1 - J(s_t)$ is the expectation of ω_t. Given this, we can rewrite the aggregate profits for the capital goods producers in equation (23) as

$$\begin{aligned}\Pi_t^k &= P_tQ_tK_{t+1} - (1-\delta)P_tQ_tK_t - P_tX_t + (1-\delta)P_tQ_tK_t\left[1 - \int_0^1 o_t\mu_t^o(o_t)do_t\right] \\ &= P_tQ_tK_{t+1} - (1-\delta)P_tQ_tK_t - P_tX_t + P_{t-1}Q_{t-1}K_tR_t^eJ(s_t).\end{aligned}$$

We can see from this aggregate profit function that what we call systemic losses for the entrepreneur are additional profits for the capital goods producers.

K_{t+1}, and that has to be repaid at time $t + 1$. The entrepreneurs operate in a legal environment that ensures them limited liability. Hence, in case of default at time $t + 1$, the banks can only appropriate the total capital return of the entrepreneur at time $t + 1$, that is, $\omega_{t+1} R^e_{t+1} P_t Q_t K_{t+1}$. The loan is restricted to take the standard form of a one-period risky debt contract as in Townsend (1979), Gale and Hellwig (1985), and Bernanke, Gertler, and Gilchrist (1999).[5]

We assume that the idiosyncratic shock ω_{t+1} is not known at time t when the loan contract is signed, and that the realization of the idiosyncratic shock can only be observed privately by the entrepreneurs himself at time $t + 1$. Banks, however, observe the systemic shock s_{t+1} at time t and have access to a costly monitoring technology that permits them to uncover the true realization of the idiosyncratic shock ω_{t+1} at a cost, that is, at a cost of $\mu \omega_{t+1} R^e_{t+1} P_t Q_t K_{t+1}$, where $0 < \mu < 1$.

Default on a loan signed at time t occurs whenever the capital returns obtained by the entrepreneur at time $t + 1$ after the realization of the idiosyncratic shock ω_{t+1}, that is $\omega_{t+1} R^e_{t+1} P_t Q_t K_{t+1}$, fall short of the amount that needs to be repaid. Hence the default space is implicitly characterized by

$$\omega_{t+1} R^e_{t+1} P_t Q_t K_{t+1} \leq I^l_{t+1} L_{t+1}, \tag{37}$$

where I^l_{t+1} is short-hand notation for the repayment amount agreed at time t per unit of loan, and L_{t+1} represents the loan size. A risky one-period loan contract at time t can be defined in terms of a threshold on the idiosyncratic shock, $\bar{\omega}_t$, and a measure of capital returns, $R^e_{t+1} P_t Q_t K_{t+1}$, such that the repayment is equal to

$$I^l_{t+1} L_{t+1} = \bar{\omega}_{t+1} R^e_{t+1} P_t Q_t K_{t+1}. \tag{38}$$

Given the terms of the loan contract, the lenders will commit to supply as much external funding as the entrepreneurs choose to demand under those conditions. Another way to interpret the implication of equations (37) and (38) is that making a loan to the entrepreneurs entitles the lenders to share in their capital returns.

When default occurs, that is, when $\omega_t < \bar{\omega}_t$, the entrepreneur

5. For a discussion of optimal contracts in a dynamic costly state verification framework, see Monnet and Quintin (2005).

cannot repay the amount it owed based on the capital returns derived from investment. To avoid misreporting on the part of defaulting entrepreneurs, the lender must verify the individual entrepreneur's income statement. That requires the lender to expend resources by an amount of $\mu\omega_{t+1}R^e_{t+1}P_tQ_tK_{t+1}$ in monitoring costs. In case of default, the lender always chooses to monitor and the entrepreneur gets nothing, while the bank appropriates $(1-\mu)\omega_{t+1}R^e_{t+1}P_tQ_tK_{t+1}$ for itself. If the entrepreneur does not default, that is, if $\omega_t > \bar{\omega}_t$, then the entrepreneur pays $\bar{\omega}_{t+1}R^e_{t+1}P_tQ_tK_{t+1}$ back to the lender and keeps the rest for himself. In other words, the entrepreneur gets to keep $(\omega_{t+1} - \bar{\omega}_{t+1})R^e_{t+1}P_tQ_tK_{t+1}$.

We take this defaulting rule and the implied sharing agreement of capital returns between the entrepreneur-borrower and the bank-lender as given. At time $t + 1$, the capital returns net of borrowing costs expected by the entrepreneurs after observing all aggregate shocks, but before the realization of its own idiosyncratic shock ω_{t+1}, can be computed as follows:[6]

$$\begin{aligned}
&\int_{\bar{\omega}_{t+1}}^{+\infty}\left[\omega_{t+1}R^e_{t+1}P_tQ_tK_{t+1} - I^l_{t+1}L_{t+1}\right]\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1} \\
&\quad = R^e_{t+1}P_tQ_tK_{t+1}\left[\int_{\bar{\omega}_{t+1}}^{+\infty}(\omega_{t+1} - \bar{\omega}_{t+1})\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1}\right] \qquad (39)\\
&\quad = R^e_{t+1}P_tQ_tK_{t+1}f(\bar{\omega}_{t+1}, s_{t+1}),
\end{aligned}$$

where

$$f(\bar{\omega}_{t+1}, s_{t+1}) \equiv \int_{\bar{\omega}_{t+1}}^{+\infty}\omega_{t+1}\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1} - \bar{\omega}_{t+1}\left[1 - \Phi(\bar{\omega}_{t+1} \mid s_{t+1})\right]. \qquad (40)$$

By the law of large numbers, equation (40) can be interpreted also as the fraction of the expected capital return obtained by the average entrepreneur. In a similar fashion, the capital returns net of monitoring costs expected by the lenders after observing all aggregate shocks at time $t + 1$ would be equal to

$$\begin{aligned}
&(1-\mu)\int_0^{\bar{\omega}_{t+1}}\left[\omega_{t+1}R^e_tP_tQ_tK_{t+1}\right]\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1} + \int_{\bar{\omega}_{t+1}}^{+\infty}\left[I^l_{t+1}L_{t+1}\right]\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1} \\
&\quad = R^e_{t+1}P_tQ_tK_{t+1}\left[(1-\mu)\int_0^{\bar{\omega}_{t+1}}\omega_{t+1}\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1} + \bar{\omega}_{t+1}\int_{\bar{\omega}_{t+1}}^{+\infty}\phi(\omega_{t+1} \mid s_{t+1})d\omega_{t+1}\right] \qquad (41)\\
&\quad = R^e_{t+1}P_tQ_tK_{t+1}g(\omega_{t+1}, s_{t+1}),
\end{aligned}$$

6. Here, aggregate shocks includes the productivity shocks, a_{t+1}, the monetary shock, m_{t+1}, and the systemic risk shocks, s_{t+1}.

where

$$g(\overline{\omega}_{t+1}, s_{t+1}) \equiv (1-\mu)\int_0^{\overline{\omega}_{t+1}} \omega_{t+1}\phi(\omega_{t+1}|s_{t+1})d\omega_{t+1} + \overline{\omega}_{t+1}\left[1-\Phi(\overline{\omega}_{t+1}, s_{t+1})\right]. \quad (42)$$

By the law of large numbers, equation (42) can be interpreted as the fraction of the expected capital returns that accrues to the average lender.

As explained in the appendix, the formal contracting problem reduces to choosing the quantity of physical capital, K_{t+1}, and the threshold, ω_{t+1}, that maximize the entrepreneurs' expected nominal return on capital net of the loan costs (see equation 39):

$$P_t Q_t K_{t+1} E_t(R^e_{t+1})\left[1 - J(s_{t+1}) - \Gamma(\overline{\omega}_{t+1}, s_{t+1})\right], \quad (43)$$

subject to the participation constraint for lenders (see equation 41), that is,

$$\begin{aligned} &P_t Q_t K_{t+1} E_t(R^e_{t+1})\Gamma(\overline{\omega}_{t+1}, s_{t+1}) - \mu G(\overline{\omega}_{t+1}, s_{t+1}) \geq I^b_{t+1} L_{t+1} \\ &= I^b_{t+1}(P_t Q_t K_{t+1} - N_{t+1}). \end{aligned} \quad (44)$$

We write the share of capital returns going to the entrepreneurs as

$$f(\overline{\omega}_{t+1}, s_{t+1}) = 1 - J(s_{t+1}) - \Gamma(\overline{\omega}_{t+1}, s_{t+1}) \quad (45)$$

and the share going to the lender as

$$g(\overline{\omega}_{t+1}, s_{t+1}) = \Gamma(\overline{\omega}_{t+1}, s_{t+1}) - \mu\ \mathrm{G}(\overline{\omega}_{t+1}, s_{t+1}). \quad (46)$$

For more details on the characterization of the functions $\Gamma(\overline{\omega}_{t+1}, s_{t+1})$ and $G(\overline{\omega}_{t+1}, s_{t+1})$, see the appendix.

Solving this optimization problem results in two additional equilibrium conditions. On the one hand, the participation constraint for the lenders becomes

$$\frac{P_t Q_t K_{t+1}}{N_{t+1}} = \frac{1}{1 - \left\{\frac{\Psi(\overline{\omega}_{t+1}, s_{t+1}) + \Gamma(\overline{\omega}_{t+1}, s_{t+1}) - \left[1 - J(s_{t+1})\right]}{\Psi(\overline{\omega}_{t+1}, s_{t+1})}\right\}}, \quad (47)$$

which implies that the threshold $\bar{\omega}_{t+1}$ can be viewed as a function of variables that are either known or observed at time t, that is,

$$\bar{\omega}_{t+1} \equiv \bar{\omega}\left(\frac{P_t Q_t K_{t+1}}{N_{t+1}}, s_{t+1}\right).$$

The expression $\Psi(\bar{\omega}_{t+1}, s_{t+1})$ is defined in the appendix as

$$\begin{aligned}\Psi(\bar{\omega}_{t+1}, s_{t+1}) \equiv\, & 1 - J(s_{t+1}) - \Gamma(\bar{\omega}_{t+1}, s_{t+1}) \\ & + \lambda(\bar{\omega}_{t+1}, s_{t+1})\left[\Gamma(\bar{\omega}_{t+1}, s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1})\right],\end{aligned} \tag{48}$$

where $\lambda(\bar{\omega}_{t+1}, s_{t+1})$ is the Lagrange multiplier on the lenders' participation constraint in equation (44) (and represents the shadow cost of enticing the participation of the lenders). The threshold depends not only on the anticipated systemic risk shock, s_{t+1}, but also on the asset-to-net-worth ratio of the entrepreneur-borrower, $P_t Q_t K_{t+1}/N_{t+1}$. Given the relationship in equation (32), the asset-to-net-worth ratio can be related to the leverage borrower as

$$\frac{P_t Q_t K_{t+1}}{N_{t+1}} = 1 + \frac{L_{t+1}}{N_{t+1}}, \tag{49}$$

where L_{t+1}/N_{t+1} is a conventional measure of the debt-to-net-worth ratio of the entrepreneur. Moreover, it can be argued that a formulation for the external financing premium arises in the following terms:

$$E_t(R^e_{t+1}) = s\left(\frac{P_t Q_t K_{t+1}}{N_{t+1}}, s_{t+1}\right) I^b_{t+1}. \tag{50}$$

This characterization of the external financing premium expands the Bernanke-Gertler-Gilchrist framework by adding the explicit possibility that the spread itself be affected by the impact of an anticipated aggregate shock, s_{t+1}. However, we preserve the key feature of the financial accelerator model, which is the linkage between the spread on capital returns and the leverage of the entrepreneurs-borrowers. Moreover, the costly state verification theory implies that external funding (loans) is more expensive than internal funding (the entrepreneurs' savings).

1.5.3 The optimal capital investment for the entrepreneurs

As noted earlier, the entrepreneurs obtain income from managerial labor at a competitive nominal wage, W^e_t, and from renting capital to wholesale firms and reselling the depreciated capital to the capital goods producers, $\omega_{t+1} R^e_{t+1} P_t Q_t K_{t+1}$. With these resources at hand, each entrepreneur must repay the previous period loans at the agreed rate (that is, they must repay $I^l_t L_t = \bar{\omega}_{t+1} R^e_{t+1} P_t Q_t K_{t+1}$) or choose to default. The entrepreneur must also finance his own consumption, C^e_t, acquire new capital from the capital producers, $P_t Q_t K_{t+1}$, and repay L_{t+1}. In this environment, the budget constraint of a representative entrepreneur can be described in the following terms:

$$P_t C^e_t + P_t Q_t K_{t+1} \leq W^e_t H^e_t + \left[1 - J(s_t) - \Gamma(\bar{\omega}_t, s_t)\right] R^e_t P_{t-1} Q_{t-1} K_t - P_t C^e_t. \quad (51)$$

Using the equilibrium participation constraint as expressed in equation (47) to replace $P_{t-1} Q_{t-1} K_t$, it immediately follows that

$$N_{t+1} \leq W^e_t H^e_t + \Psi(\bar{\omega}_t, s_t) R^e_t N_t - P_t C^e_t. \quad (52)$$

Based on this characterization of the budget constraint of the representative entrepreneur, we can infer that an interior solution of his optimization problem in which equation (52) holds with equality can be obtained as the solution to an equivalent maximization problem, according to which the entrepreneur chooses his real net worth, N_{t+1}/P_t, to maximize

$$\sum_{\tau=0}^{+\infty} (\beta\eta)^{\tau} E_t \left[\frac{W^e_{t+\tau}}{P_{t+\tau}} + \Psi(\bar{\omega}_{t+\tau}, s_{t+\tau}) R^e_{t+\tau} \frac{P_{t+\tau-1}}{P_{t+\tau}} \frac{N_{t+\tau}}{P_{t+\tau-1}} - \frac{N_{t+\tau+1}}{P_{t+\tau}} \right], \quad (53)$$

where we implicitly use the fact that managerial labor is inelastically supplied and normalized to one (as pointed out in equation 31).

This intertemporal optimization must satisfy the following Euler equation:

$$1 = \beta\eta E_t \left[\Psi(\bar{\omega}_{t+1}, s_{t+1}) R^e_{t+1} \frac{P_t}{P_{t+1}} \right], \quad (54)$$

which determines the consumption-savings margin for the representative entrepreneur. The left-hand side of equation (54) is the marginal utility of entrepreneurs' consumption. The right-hand side is the expected discounted real rate of return of acquiring a unit of capital after taking into account the costs associated with the need for external funding. The latter term has two components. The first term, $\Psi(\bar{\omega}_{t+1}, s_{t+1})$, captures the effect of default on external borrowing costs and also accounts for the role of anticipated systemic losses. The second component, $R^e_{t+1}(P_t / P_{t+1})$, is the real rate of return on capital whenever the resale value of depreciated capital and the acquisition cost of new capital are equalized.

1.6 Banks

There is a continuum of banks of unit mass. All banks are systemic and perfectly competitive, so they take all prices as given. The bank offers the households two types of assets for investment purposes: one that we call bank equity and another that we call one-period deposits. Deposits offer a nominal risk-free rate, while equity is rewarded with a riskless return in every period that induces households-shareholders to hold bank capital as well. All households who own bank equity must be indifferent between investing in equity or simply making a deposit.

For convenience, we define the safe return promised to the equity holders in terms of a yield, R^e_{t+1}, over the value of the banks equity, B_{t+1}. Household deposits are perfectly insured and pay a risk-free rate, I_{t+1}. Banks use all the resources they attract (deposits and bank capital) to offer one-period loans to the entrepreneurs with the conditions described above. At the end of each loan contract, all unanticipated profits accrued by the bank are rebated (lump-sum) to the households independently of their portfolio allocation between the bank's liabilities (deposits) and equity.

At the end of period t, the balance sheet of the banking system can be summarized as follows:

$$L_{t+1} + \varpi D_{t+1} = B_{t+1} + D_{t+1}, \tag{55}$$

where the right-hand side describes the liabilities (that is, the deposits) taken at time t, D_{t+1}, and the equity offered at the same time, B_{t+1}. The left-hand side shows the assets, $L_{t+1} + \varpi D_{t+1}$. Among the assets, we count the reserves on deposits maintained at the central bank (that is, ϖD_{t+1}, where $0 \leq \varpi \leq 1$ represents the compulsory reserve requirement on nominal deposits set by the regulator) and the loans offered at time t, L_{t+1}. As a matter of convention, D_{t+1} denotes nominal deposits and L_{t+1} nominal loans held from time t to $t+1$. Similarly, B_{t+1} is the bank capital outstanding between time t and time $t+1$.

We can rewrite the balance sheet more conveniently as

$$L_{t+1} = \left(\frac{1-\varpi}{1-\upsilon_{t+1}} \right) D_{t+1}, \tag{56}$$

where we define the leverage ratio on bank capital as $\upsilon_{t+1} \equiv (B_{t+1}/L_{t+1})$. In other words, the rate at which deposits are transformed into loans is affected by the compulsory reserve requirement and by the bank's capital leverage policy. In Bernanke, Gertler, and Gilchrist (1999), with $\varpi = 0$ and no bank equity (that is, $\upsilon_{t+1} = 0$), the transformation rate is one to one. It thus holds that $L_{t+1} = D_{t+1}$. Although the model preserves the basic underlying structure of the bank's balance sheet in Bernanke, Gertler, and Gilchrist (1999), equation (56) indicates that the regulatory features should play a significant role on the cost structure of loan supply.

The banks' profits on a given one-period loan contract are realized at time $t+1$. We can express the profits of the banking system as

$$\begin{aligned} \Pi^b_{t+1} \equiv\ & R^e_{t+1} P_t Q_t K_{t+1} \left[\Gamma(\bar{\omega}_{t+1}, s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1}) \right] \\ & + \varpi \bar{I}_{t+1} D_{t+1} - R^b_{t+1} B_{t+1} - I_{t+1} D_{t+1}, \end{aligned} \tag{57}$$

while the expected profits at the time the loan is contracted should be

$$E_t \left(\Pi^b_{t+1} \right) \equiv I^b_{t+1} L_{t+1} + \varpi \bar{I}_{t+1} D_{t+1} - R^b_{t+1} B_{t+1} - I_{t+1} D_{t+1}. \tag{58}$$

The required nominal participation returns on loans, I^b_{t+1}, are determined at time t when the loans are signed between the banks-lenders and the entrepreneurs-borrowers (see the participation constraint in equation 44). Deposits held at the central bank in the form of reserves are also returned to the banks. We assume that they earn an interest on reserves, $\bar{I}_{t+1}$, which is known at time t and designed as a two-part rate:

$$\bar{I}_{t+1} \equiv (1-c) + \zeta(I_{t+1} - 1), \tag{59}$$

whereby banks pay a fixed fee as a management cost per unit of reserve held at the central bank, and $0 < \zeta < 1$ denotes the discount rate relative to the monetary net short-term rate at which reserves are compensated. Although in most instances the practice is to set this rate of return to zero (that is, $c = \zeta = 0$), there are precedents for paying interest on reserves.[7] We also make the simplifying assumption that there is full deposit insurance, so that deposits are riskless and the gross interest rate paid on deposits is equal to the risk-free nominal rate, I_{t+1}, which is known at time t.

Bank capital shareholders (that is, the households) have to be compensated with a certain nominal yield, R^b_{t+1}, determined at time t. Since at time t expected profits depend exclusively on variables that are chosen and known at that time by the banks and the households, competitive banks must end up offering a yield to the shareholders that is also known at time t. By arbitrage implied in equations (3) and (4), it must therefore be the case that

$$(1-\iota^h)R^b_{t+1} = I_{t+1}, \tag{60}$$

which ensures that households remain indifferent between holding bank capital or deposits. For a competitive banking sector, the expected profit function in equation (58) must satisfy a zero-expected profit condition (that is, $\Pi^b_{t+1} = 0$) in the following terms:

$$E_t\left(\Pi^b_{t+1}\right) \equiv \left[I^b_{t+1} - \upsilon_{t+1}R^b_{t+1} - (1-\upsilon_{t+1})\left(\frac{I_{t+1} - \varpi\bar{I}_{t+1}}{1-\varpi}\right)\right]L_{t+1} = 0. \tag{61}$$

After using the balance sheet equation in equation (56). The banks' problem is to optimize their capital structure to reflect the trade-off between bank equity and deposits, subject to the constraint that banks must offer a yield on bank capital that makes households indifferent given the existing option of a risk-free rate on deposits as given by equation (60). Of course, this problem is also subject to the feature of the central bank's policy of paying reserves as given

7. Until very recently, reserve requirements held at the Federal Reserve did not earn interest. The Federal Reserve announced changes to reserve management after winning the power to pay interest on required and excess reserves on 3 October 2008. The Federal Reserve has argued that paying interest would deter banks from lending out excess reserves and as such would make it easier for the Fed to attain its target rate. We do not attempt to model this feature explicitly.

by equation (61) and subject to a regulatory constraint on capital adequacy that implies banks must satisfy

$$1 \geq \upsilon_{t+1} \equiv \frac{B_{t+1}}{L_{t+1}} \geq \upsilon, \tag{62}$$

where $0 \leq \upsilon < 1$ is equal to the minimum mandatory capital adequacy requirement set by the regulator.[8] The lower bound, υ, may also reflect a buffer above the minimum regulatory requirement implied by the statutory requirements of the banks themselves, and it could even be time-varying over the cycle.

We make two key parametric assumptions to simplify the problem of the banks, and we leave the exploration of more complex banking cost structures for future research. Our goal at this stage is to make only the smallest possible departure from the original Bernanke-Gertler-Gilchrist framework. We assume that $\zeta = 1 - c$ and that taxes on bank equity are bounded by $0 < 1 - \iota^h < (1 - \varpi) / (1 - \zeta\varpi)$. Whenever $\xi = 0$, this bound implies that $\iota^h > \varpi$; whenever $\xi = 1$, it merely requires that $\iota^h > 0$. Given the fact that tax rates are quite often much higher than the minimum reserve ratios, these bounds are not excessively restrictive.

The two assumptions together imply that

$$R^b_{t+1} > \left(\frac{I_{t+1} - \varpi \overline{I}_{t+1}}{1 - \varpi} \right). \tag{63}$$

In other words, it is costlier for banks to finance themselves with bank equity than with deposits. Therefore, the lower bound on the leverage ratio must be binding at all times.

8. The current regulatory regime was shaped primarily by the 1988 international Basel Accord and the 1991 Federal Deposit Insurance Corporation Improvement Act (FDICIA). The Basel Accord established minimum capital requirements as ratios of two aggregates of accounting capital to risk-weighted assets (and certain off-balance-sheet activities). The risk weights are supposed to reflect credit risk. For example, commercial and industrial loans have a weight of one, while U.S. government bonds have zero weight and consequently do not require any regulatory capital. Primary or tier 1 (core) capital (equal to the book value of the bank's stock plus retained earnings) is required to exceed 4 percent of risk weighted assets, while total (tier 1 plus tier 2) capital must be at least 8 percent. In calculating the risk-weighted capital asset ratio, all loans are assumed to be in the highest risk category in the sense of the Basel Accord, with a risk weight of 100 percent. This category includes all claims to the nonbank private sector, except for mortgages on residential property, which receive a risk weight of 50 percent. The riskless securities are in the lowest risk category, with a weight of zero. Typical examples are Treasury bills and short loans to other depository institutions.

These assumptions imply that the participation rate of return required by the banks to fund the entrepreneurs is fully determined by the cost structure of the banks themselves, as follows:

$$\begin{aligned} I_{t+1}^{b} &= \upsilon R_{t+1}^{b} + (1-\upsilon)\left(\frac{I_{t+1} - \varpi \overline{I}_{t+1}}{1-\varpi}\right) \\ &= \left[\upsilon\left(\frac{1}{1-\iota^{h}}\right) + (1-\upsilon)\left(\frac{1-\varpi\zeta}{1-\varpi}\right)\right] I_{t+1}. \end{aligned} \tag{64}$$

This is what we call the balance sheet channel of banking regulation. Without capital adequacy requirements (that is, $\upsilon = 0$) and without reserve requirements (that is, $\varpi = 0$), we would be back to the world of Bernanke, Gertler, and Gilchrist (1999), where $I_{t+1}^{b} = I_{t+1}$. Our equation (64) is a heavily parametrized version of the following expression for returns on the portfolio of loans under constant returns to scale:

$$\frac{I_{t+1}^{b}}{I_{t+1}} \equiv \upsilon_{t+1} \times \frac{\text{cost}(\textit{bank equity}_{t+1})}{I_{t+1}} + (1-\upsilon_{t+1}) \times \frac{\text{cost}(\textit{deposits}_{t+1})}{I_{t+1}}, \tag{65}$$

where υ_{t+1} represents the leverage ratio as before. The realized profits at the time the loan contract expires in equations (57) can, alternatively, be represented as

$$\Pi_{t+1}^{b} \equiv \left[R_{t+1}^{b} - E_t(R_{t+1}^{b})\right] P_t Q_t K_{t+1} \left[\Gamma(\overline{\omega}_{t+1}, s_{t+1}) - \mu G(\overline{\omega}_{t+1}, s_{t+1})\right], \tag{66}$$

where we have used the participation constraint in equation (44) appropriately. Hence, the realized profits reflect the intertemporal aggregate risks associated with the portfolio of loans supplied to the entrepreneurs (which is captured by the margin $R_{t+1}^{e} - E_t(R_{t+1}^{e})$) on the asset side of the banks' balance sheet. The assumption that all realized profits are rebated to the households (no profits are retained by the banks) transfers the consequences of the aggregate risks to the households, who cannot avoid them by adjusting their portfolio between bank equity and bank deposits. We leave for future research the exploration of a more complex environment in which banks' dividends are related to equity holdings and, more interestingly, in which retained profits can affect the evolution of bank equity and expose bank capital to aggregate risks.

1.7 Government

We close our description of the model with the specification of a consolidated (and balanced) budget constraint and an interest rate rule for monetary policy. We assume that government expenditures and the subsidy on inputs for retailers are financed through lump-sum taxes on households, taxes on bank equity, and seigniorage, that is,

$$\begin{aligned} P_t G_t + T_t + \iota^h R_t^b B_t + M_{t+1} &= \iota^r P_t^w \left[\int_0^1 Y_t(z) dz \right] + \bar{I}_t M_t \\ &= \iota^r P_t^w Y_t^w + \bar{I}_t M_t, \end{aligned} \tag{67}$$

where G_t denotes real government expenditure. We assume for simplicity that government consumption is equal to zero in every period, that is, $G_t = 0$. The characteristics and bounds on the tax subsidy for retailers, ι^r, and the tax rate on dividends, ι^h, as well as the nature of the nondistortionary (lump-sum) tax or transfer to the households, T_t, have already been discussed elsewhere. The government also funds its operations by issuing high-powered money (the monetary base), M_{t+1}, at time t.

For the purpose of defining the monetary base, money consists only of the total reserves of the banking sector on their accounts at the central bank. Therefore, given the compulsory requirement on reserves, the equilibrium in the money market requires that

$$M_{t+1} = \varpi D_{t+1}. \tag{68}$$

As noted before, reserves deposited at time t accrue a rate of return $\bar{I}_t$, which is characterized by the formula in equation (61). For simplicity, money plays exclusively the role of a unit of account and acts as the counterpart for deposit reserves on the balance sheet of the central bank.

The central bank policy is modeled by means of an interest rate reaction function. In the spirit of Taylor (1993), the policy rule targets the short-term nominal interest rate, I_{t+1}, and is linear in the logs of the relevant arguments:

$$i_{t+1} = \rho_i i_t + (1 - \rho_i) \left[\psi_\pi \ln \left(\frac{P_t}{P_{t+1}} \right) + \psi_q \ln(Q_t) + \psi_y \ln(Y_t) \right] + m_t, \tag{69}$$

where $i_t \equiv \ln(I_t)$ is the logarithm of the risk-free rate. In line with most of the literature, we assume that the monetary authority is willing to smooth changes in the actual short-term nominal interest rate, that is, $0 \leq \rho_i \leq 1$, where ρ_i is the smoothing parameter. The other parameters in the reaction function satisfy $\psi_\pi \geq 1$, $-\infty < \psi_1 < +\infty$, and $\psi_y \geq 0$. The monetary shock in logs, m_t, follows an AR(1) process of the following form:

$$m_t = \rho_m m_{t-1} + \varepsilon_t^m, \tag{70}$$

where ε_t^m is a zero mean, uncorrelated, and normally distributed innovation. The parameter $-1 < \rho_m < 1$ determines the persistence of the monetary shock and $\sigma_m^2 > 0$ the volatility of its innovation.

A few observations on the specification of equation (69) are in order. First, we model monetary policy in terms of an implementable rule, whereby the central bank sets the short-term nominal interest rate in response to observable variables only. Second, this general specification allows the monetary policy instrument to react to deviations of the relative price of capital goods, Q_t, from its long-run value of one. This is the channel through which we allow asset price fluctuations to feed into the setting of monetary policy.

Third, equation (71) can always be rewritten in terms of a pure trade-off between inflation and output, as follows:

$$\begin{aligned} i_{t+1} = & \left\{ \rho_i + \psi_q (1-\rho_i) \left[\frac{\ln(Q_t)}{i_t - \ln\left(\frac{P_t}{P_{t-1}}\right)} \right] \right\} i_t \\ & + (1-\rho_i) \left\{ \left[\psi_\pi - \psi_q \frac{\ln(Q_t)}{i_t - \ln\left(\frac{P_t}{P_{t-1}}\right)} \right] \ln\left(\frac{P_t}{P_{t-1}}\right) + \psi_y \ln(Y_t) \right\} + m_t, \end{aligned} \tag{71}$$

where the coefficient on inflation and the inertia parameter vary depending on whether Tobin's Q is growing faster than the ex post real interest rate. This is obviously one of many observationally equivalent rules that we could write that are consistent with the structure of equation (69). In more general terms, it would be rather appealing to fix monetary policy in terms of a well-known trade-off between inflation and output, but at the same time allow flexibility for the rule to respond differently to systemic risk, which is a critical source of uncertainty in our framework.

The specification of the Taylor rule that we have in mind would take the following form:

$$
\begin{aligned}
i_{t+1} &= \rho_i (s_t - \bar{s}) i_t \\
&\quad + \left[1 - \rho_i (s_t - \bar{s})\right] \left[\psi_\pi (s_t - \bar{s}) \ln\left(\frac{P_t}{P_{t-1}} \right) + \psi_y (s_t - \bar{s}) \ln(Y_t) \right] + m_t,
\end{aligned}
\tag{72}
$$

where the inertia and the weights on inflation and output are a function of the perceived riskiness of the current environment as determined by the distance of the actual systemic risk shock realization, s_t, relative to the breaking point after which losses in the secondary market for used capital become catastrophic.

1.7.1 Resource constraint

Equilibrium in the final goods market requires that the production of the final good be allocated to total private consumption by households and entrepreneurs (and possibly the government), to investment by capital goods producers, and to covering the costs that originate from the monitoring technology required to enforce the loan contract described earlier (and in the appendix). That is,

$$
Y_t = C_t + C^e + G_t + X_t + \underbrace{\mu G(\bar{\omega}, s_t) R_t^e \frac{P_{t-1}}{P_t} Q_{t-1} K_t}_{\text{Loss from monitoring costs}},
\tag{73}
$$

where final output and wholesale output are related as $Y_t = (P_t^* / P_t)^\theta Y_t^w$. In the above equation, the impact of government consumption is trivial since we have assumed for simplicity that G_t=0. In the Bernanke-Gertler-Gilchrist model, government consumption evolves exogenously and is assumed to be financed through lump-sum taxes. A similar extension can be implemented in our setting.

2. Discussion and Interpretation

The relationship in equation (64) clearly ties the participation return, I^b_{t+1}, to the risk-free rate, I_{t+1}, which happens to also be the relevant instrument for monetary policy. The regulatory restriction on capital adequacy in equation (62) does not prevent bad outcomes from happening. Instead, the purpose of this regulatory constraint is to effectively give the monetary authority a way to regulate the supply of loans without having to manipulate the interest rate directly. In that sense, we can visualize the banks' balance sheet channel in this framework by combining equations (50) and (64) as follows:

$$E_t(R^e_{t+1}) = \underbrace{s\left(\frac{P_t Q_t K_{t+1}}{N_{t+1}}, s_{t+1}\right)}_{\text{Agency costs channel, as in BGG (1999)}} \underbrace{\left[\upsilon\left(\frac{1}{1-\iota^h}\right) + (1-\upsilon)\left(\frac{1-\varpi\zeta}{1-\varpi}\right)\right]}_{\text{Balance sheet channel} \geq 1} I_{t+1}. \quad (74)$$

This equation shows that the balance sheet channel has the potential to amplify the external financing premium spread. However, because this channel is regulated by the central bank, the monetary authority can potentially manipulate the requirements to reduce the amplification effect when the agency cost component is rising.

We have a fairly standard setting that quite closely follows the derivation of the equilibrium conditions in Bernanke, Gertler, and Gilchrist (1999), so our linearization shows obvious similarities with theirs. The main differences arise because we have introduced frictions in the secondary market for used capital that have the potential to alter the conditions under which borrowers and lenders operate in this economy, and because we have expanded the balance sheet of the banks-lenders to give banking regulation a role in loan pricing decisions.

Entrepreneurs cannot borrow at the riskless rate as revealed in equation (74). The cost of external financing differs from the risk-free rate because the idiosyncratic component to their returns on capital is unobservable from the banks' point of view. To infer the realized return of the entrepreneur, the bank has to pay a monitoring cost. The banks monitor the entrepreneurs that default, pay the verification cost, and seize the remaining capital income. In equilibrium, entrepreneurs borrow up to the point at which the expected return on capital equals the cost of external financing:

$$E_t(\hat{r}^e_{t+1}) \approx \hat{i}_{t+1} + \vartheta\left(\hat{p}_t + \hat{q}_t + \hat{k}_{t+1} - \hat{n}_{t+1}\right) + \Lambda\hat{\upsilon}_{t+1} + \Theta\hat{s}_{t+1}, \quad (75)$$

where $\hat{k}_{t+1}$ denotes capital, $\hat{n}_{t+1}$ is the entrepreneur's net worth, $\hat{q}_t$ is Tobin's Q, $\hat{p}_t$ is the CPI, $\hat{i}_{t+1}$ is the risk-free rate, $\hat{\upsilon}_{t+1}$ determines changes in banking regulation (capital adequacy) or the bank's leverage policy, and $\hat{s}_{t+1}$ stands for the systemic risk shock that capture the distortions in the secondary market for used capital. The composite parameters ϑ, Λ, and Θ can be expressed as a function of the structural parameters of the model, and all variables in lowercase letters with an over hat represent log deviations from the steady state.

The right-hand side of the external financing premium equation in equation (75) can be decomposed into two terms: the nominal risk-free rate and the external financing premium.[9] The parameter ϑ measures the elasticity of the external financing premium to variations in leverage of the entrepreneurs, measured by their capital expenditures relative to net worth. The larger the share of the capital purchase financed with the entrepreneurs' net worth, the closer the spread is to zero and the lower the associated moral hazard. If entrepreneurs have sufficient savings to finance the entire capital stock, then agency problems vanish, and the risk-free rate and the expected return to capital income must coincide unless either the banks' leverage, $\hat{\upsilon}_{t+1}$, or systemic risk $\hat{s}_{t+1}$, vary. So far, this is the same result found in Bernanke, Gertler, and Gilchrist (1999). Our model, however, illustrates that changes in banking regulation on capital adequacy and systemic risk add a new dimension to the external financing premium that cannot be discounted.

Two points warrant further discussion here. First, our specification of a Taylor rule in equation (72) depends on exogenous shocks that are potentially unobservable to policymakers. Second, our characterization of banks, while more complete than Bernanke, Gertler, and Gilchrist (1999), is nonetheless simple. The remainder of this section addresses these issues.

2.1 Taylor Rules

A potential disadvantage of our specification of the Taylor rule in equation (72), namely,

9. The key mechanism involves the link between the external financing premium (that is, the difference between the cost of funds raised externally and the opportunity cost of internal funds) and the net worth of the entrepreneurs-borrowers.

$$i_{t+1} = \rho_i(s_t - \bar{s})i_t + \left[1 - \rho_i(s_t - \bar{s})\right]\left[\psi_\pi(s_t - \bar{s})\ln\left(\frac{P_t}{P_{t-1}}\right) + \psi_y(s_t - \bar{s})\ln(Y_t)\right] + m_t, \tag{76}$$

is that monetary policy depends on an exogenous shock that is not necessarily observable to the policymaker, the systemic shock s_t. An alternative is to explore a policy rule reflecting the assumption that monetary authorities readjust the weights on inflation and output in response to the other observable variables every period, reacting to asset prices, Q_t, as in out conjecture in equation (69). That is,

$$i_{t+1} = \rho_i(s_t - \bar{s})i_t + (1 - \rho_i)\left[\psi_\pi \ln\left(\frac{P_t}{P_{t-1}}\right) + \psi_q \ln(Q_t) + \psi_y \ln(Y_t)\right] + m_t. \tag{77}$$

We could even explore alternative rules in which the central bank's response depends on the size of the spreads between the risk-free rate and the implied returns on capital, along the lines of Curdia and Woodford (2008). A potential specification that fits our environment is

$$i_{t+1} = \rho_i(s_t - \bar{s})i_t + (1 - \rho_i)\left[\psi_\pi \ln\left(\frac{P_t}{P_{t-1}}\right) + \psi_a \ln\left(\frac{P_t Q_t K_{t+1}}{N_{t+1}}\right) + \psi_y \ln(Y_t)\right] + m_t. \tag{78}$$

This rule targets the leverage ratio of the borrowers, since theory tells us that this is the unobservable component of the external financial premium in equation (74).

As noted above, specification (77) is comparable to the Taylor rule presented in equation (76), and they produce similar results when implemented as long as Tobin's Q is a sufficient statistic for the unobservable systemic shock. The same can be said of the specification in equation (78). Whether such Taylor rules are optimal relative to a rule with constant coefficients will likely depend on whether the observable variables (Tobin's Q or the spreads) are good proxies for signaling trouble in the secondary market for used capital. Monetary policy is likely to improve its performance if it can react to strong signals, but it probably will not do better than under an old-fashioned Taylor rule with constant coefficients if the signal is weak or gives the wrong message depending on the nature of the shock that hits the economy.

The transmission mechanism that affects the dynamics of the economy over the business cycle is also quite important here. Monetary policy has no direct effect on the systemic shock in equation (75), since this shock is assumed to be exogenous. However, the central bank can either alter the bank regulatory requirements, $\hat{\upsilon}_{t+1}$, or the short-term interest rate, $\hat{i}_{t+1}$, to offset fluctuations of the spread that tend to increase the volatility of the cost of external borrowing for the entrepreneurs and potentially lead to periods of excessive investment or underinvestment. Monetary policy, whether implemented conventionally through interest rate movements or by changes in banking regulation, would nonetheless have an indirect effect on the equilibrium spreads, which can limit the effectiveness of those actions.

2.2 Banking Sector

Arguably, our model remains a very naïve characterization of the behavior of banks. We are far from having an integrated model of the business cycle in which banks operate in multiple periods, with a portfolio of loans of different maturities, and simultaneously confront friction in their lending operations and nontrivial distortions in the way they raise capital or attract depositors. However, this characterization of the economy emphasizes the regulatory power to alter the operational costs of the banking system. Even in this simplified framework, it immediately transpires that the regulator is able to alter the terms of the banks' operating costs. The regulator thus has at hand a tool to either amplify or reduce the loan supply without directly changing the short-term interest rate. This framework offers a way to explore how the model responds to monetary policy and regulatory features.

We have already noted that regulatory features can be modified with the intention of offsetting fluctuations in the spread faced by borrowers on external funding. In principle, given the fact that reserve requirements and capital adequacy requirements are not excessively punitive in most developed countries, one might expect that changes in banking regulation would have small effects on the cost structure of banks and, therefore, would have less of an impact on the cost of borrowing for entrepreneurs. However, in the extreme case in which $\hat{\upsilon}_{t+1} = -(\Theta / \Lambda)\hat{s}_{t+1}$, it might be possible to entirely eliminate the effect of systemic risk on shocks without altering the interest rate. It might therefore be possible to limit the impact of the systemic risk shock on the economy without having to alter the entrepreneurs' incentives to invest and the households' incentives to save.

While the potential for banking regulation to play a countercyclical role is present in the model, and noted in our comments, it is not easy to obtain a clear signal of the risks confronted. In most instances, the systemic risks, $\hat{s}_{t+1}$, are simply not observable, and relying on observables to define the cyclical patterns of banking regulation is as difficult as it was for setting the interest rate rule. In practice, however, the banking leverage ratios tend to be procyclical and contribute to amplifying the cycle, so the policy debate is more oriented toward policies that would reduce those tendencies than turning banking regulation into a cyclical counterbalance.

3. Concluding Remarks

Our paper has set forth a model of the economy that generalizes the Bernanke-Gertler-Gilchrist model to include a compact characterization of both the financial accelerator and the role of the financial sector in propagating monetary policy to the real economy. We have identified the output costs of systemic risk and the agency costs of costly state verification, as well as their role in determining the external finance premium. Equation (74) neatly summarizes this relationship and makes clear how the financial sector can amplify the cycle as discussed in Bernanke, Gertler, and Gilchrist (1999). This characterization provides a parsimonious explanation that can be compared with existing research on the interaction between monetary policy and bank regulation. This result arises as part and parcel of a model designed to explain the transmission and amplification of monetary action.

A model that includes this type of lending channel can go some length toward explaining the monetary policy asymmetries that Taylor rules have been unable to account for in the last few years. Moreover, since our model is built around the existence of a regulatory capital constraint, it provides the basis for discussing the implication of joint determination of monetary policy and regulation. Indeed, the presence of differences in monetary policy discussed in this model implies a strong incentive for the joint monetary/regulatory authority to ensure that financial institutions remain above the capital constraint. In times of falling asset values, banks will approach or fall below capital requirements, rendering monetary policy ineffective at stimulating lending. At this point, the monetary/regulatory authority has a stronger incentive to lower capital requirements in order to facilitate monetary intervention. If falling asset values were due to a realization of inaccurate risk measurements, reduced capital levels may simply encourage reckless lending.

With this framework in place, there are potentially more open questions ahead (and, unfortunately, beyond the scope of this paper). For example, while the model appears to do a reasonably good job of describing the stylized patterns of the U.S. monetary authority during the recent crisis (at least in suggesting that the reduction of interest rates are a plausible policy response to systemic shocks and bank lending constraints), it is nonetheless potentially rejected by the European case. The European Central Bank held interest rates constant until late 2008. Though there are many possible reasons for this, we speculate that it emerges, in part, from differences in mandate. The Federal Reserve has responsibility for both monetary policy and bank regulation. This produces well-known conflicts between the goals of monetary policy and bank regulation. It also produces an incentive to keep banks above regulatory thresholds through the use of monetary policy (see Cecchetti and Li, 2008, on neutralization of the capital constraint).

Appendix
The Loan Contract

With regard to the aggregate sharing of capital income, we define the following two variables for simplicity of notation,

$$\Gamma(\bar{\omega}_{t+1}, s_{t+1}) \equiv \int_0^{\bar{\omega}_{t+1}} \omega_{t+1} \phi(\omega_{t+1} | s_{t+1}) d\omega_{t+1} + \bar{\omega}_{t+1} \left[1 - \phi(\bar{\omega}_{t+1}, s_{t+1})\right]; \tag{A1}$$

$$\mu G(\bar{\omega}_{t+1}, s_{t+1}) \equiv \mu \int_0^{\bar{\omega}_{t+1}} \omega_{t+1} \phi(\omega_{t+1} | s_{t+1}) d\omega_{t+1}. \tag{A2}$$

Then, we can rewrite the share of capital returns going to the lenders in equation (42) more compactly as

$$g(\bar{\omega}_{t+1}, s_{t+1}) = \Gamma(\bar{\omega}_{t+1}, s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1}). \tag{A3}$$

Given the definition of the capital returns share going to entrepreneurs in equation (40), it also follows that

$$\begin{aligned} f(\bar{\omega}_{t+1}, s_{t+1}) &= \int_0^{+\infty} \omega_{t+1} \phi(\omega_{t+1}, s_{t+1}) d\omega_{t+1} - \Gamma(\bar{\omega}_{t+1}, s_{t+1}) \\ &= 1 - J(s_{t+1}) - \Gamma(\bar{\omega}_{t+1}, s_{t+1}), \end{aligned} \tag{A4}$$

where the second equality follows from our characterization of the expectation of the idiosyncratic shock in equation (36). Based on these definitions, we can infer that the capital income sharing rule resulting from this financial contract satisfies

$$f(\bar{\omega}_{t+1}, s_{t+1}) + g(\bar{\omega}_{t+1}, s_{t+1}) = 1 - J(s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1}), \tag{A5}$$

where $J(s_{t+1}) \equiv 1 - E_t(\omega_{t+1} \mid s_{t+1})$ accounts for the expected systemic losses on the resale value of capital and $\mu G(\bar{\omega}_{t+1}, s_{t+1})$ characterizes the conventional monitoring costs and probability of default associated with the costly-state verification framework.

The functions $f(\bar{\omega}_t, s_t)$ and $g(\bar{\omega}_t, s_t)$ represent the sharing rule between entrepreneurs-borrowers and banks-lenders on the capital returns required by the entrepreneur's partial use of one-period external loans to fund its risky capital investment. Both of them depend on the realization of the systemic risk shock, s_{t+1}. However, as can be inferred from equation (A5), they do not add up to one.

A fraction of the capital income, $J(s_{t+1})$, is transferred to the capital goods producers as a result of inefficiencies in the secondary market for used capital, while another fraction, $\mu G(\bar{\omega}_{t+1}, s_{t+1})$, is lost due to the burden of monitoring. Only monitoring costs result in a direct loss of capital income that detracts resources, as shown in the resource constraint in equation (75), but the fact that resources are siphoned out of the hands of borrowers and lenders due to market imperfections somewhere else still has the potential to substantially distort the incentives of both parties involved in the loan contract and, therefore, to affect the funding of investment in new capital.

The optimization problem

We conjecture that the threshold, $\bar{\omega}_{t+1}$, would be defined as a function of the systemic risk shock, s_{t+1}, and the assets-to-net-worth ratio at time *t*, $P_t\, Q_t\, K_{t+1}/N_{t+1}$. Given our conventions, both are either observed or determined by all parties at time *t*. Therefore, equation (39) implies that with the information available at time *t*, entrepreneurs expect capital returns equal to

$$P_t Q_t K_{t+1} E_t(R^e_{t+1}) f(\bar{\omega}_{t+1}, s_{t+1}). \tag{A6}$$

Similarly, equation (41) implies that with the information available at time *t*, lenders expect income equal to

$$P_t Q_t K_{t+1} E_t(R^e_{t+1}) g(\bar{\omega}_{t+1}, s_{t+1}). \tag{A7}$$

The formal contracting problem reduces to choosing the quantity of physical capital, K_{t+1}, and the threshold, $\bar{\omega}_{t+1}$, that maximize the entrepreneurs' expected nominal return on capital net of loan costs (see equations A6 and A4). That is,

$$P_t Q_t K_{t+1} E_t(R^e_{t+1}) \left[1 - J(s_{t+1}) - \Gamma(\bar{\omega}_{t+1}, s_{t+1})\right], \tag{A8}$$

subject to the participation constraint for the lenders (see equations A7 and A3). That is

$$\begin{aligned} &P_t Q_t K_{t+1} E_t(R^e_{t+1}) \left[\Gamma(\bar{\omega}_{t+1}, s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1})\right] \geq I^b_{t+1} L_{t+1} \\ &= I^b_{t+1}(P_t Q_t K_{t+1} - N_{t+1}), \end{aligned} \tag{A9}$$

where the equality on the right-hand side follows from equation (32). It is implicitly agreed that if lenders participate in this contract, they always supply enough loans, L_{t+1}, as long as a noncontingent participation rate, I^b_{t+1}, is guaranteed to them in expectation. In other words, we do not explicitly consider the possibility of credit rationing, while we view the (risk-neutral) banks as bearing part of the aggregate risk. All banks share equally in the aggregate size of the loan.

The first-order condition with respect to $\bar{\omega}_{t+1}$ defines the function $\lambda_{t+1} \equiv \lambda(\bar{\omega}_{t+1}, s_{t+1})$ in the following terms:

$$\Gamma_1(\bar{\omega}_{t+1}, s_{t+1}) - \lambda(\bar{\omega}_{t+1}, s_{t+1})\left[\Gamma_1(\bar{\omega}_{t+1}, s_{t+1}) - \mu G_1(\bar{\omega}_{t+1}, s_{t+1})\right] = 0, \tag{A10}$$

where λ_{t+1} is the Lagrange multiplier on the lenders' participation constraint. By virtue of this optimality condition, we say that the shadow cost of enticing the participation of the lenders in this contract is given by

$$\lambda(\bar{\omega}_{t+1}, s_{t+1}) = \frac{\Gamma_1(\bar{\omega}_{t+1}, s_{t+1})}{\Gamma_1(\bar{\omega}_{t+1}, s_{t+1}) - \mu G_1(\bar{\omega}_{t+1}, s_{t+1})}, \tag{A11}$$

which, in turn, implies that the participation constraint must be binding since the multiplier is nonzero. The binding participation constraint can be rewritten as

$$\frac{P_t Q_t K_{t+1}}{N_{t+1}} E_t\left(\frac{R^e_{t+1}}{I^b_{t+1}}\right)\left[\Gamma(\bar{\omega}_{t+1}, s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1})\right] = \left(\frac{P_t Q_t K_{t+1}}{N_{t+1}} - 1\right), \tag{A12}$$

or, more compactly,

$$\frac{P_t Q_t K_{t+1}}{N_{t+1}} = \frac{1}{1 - E_t\left(\frac{R^e_{t+1}}{I^b_{t+1}}\right)\left[\frac{\Psi(\bar{\omega}_{t+1}, s_{t+1}) + J(s_{t+1}) + \Gamma(\bar{\omega}_{t+1}, s_{t+1}) - 1}{\lambda(\bar{\omega}_{t+1}, s_{t+1})}\right]}, \tag{A13}$$

where we define $\Psi(\bar{\omega}_{t+1}, s_{t+1})$ as,

$$\begin{aligned}\Psi(\bar{\omega}_{t+1}, s_{t+1}) \equiv{} & 1 - J(s_{t+1}) - \Gamma(\bar{\omega}_{t+1}, s_{t+1}) \\ & + \lambda(\bar{\omega}_{t+1}, s_{t+1})\left[\Gamma(\bar{\omega}_{t+1}, s_{t+1}) - \mu G(\bar{\omega}_{t+1}, s_{t+1})\right].\end{aligned} \tag{A14}$$

The optimization also requires the following first-order condition with respect to capital, K_{t+1}, to hold:

$$E_t\left(\frac{R^e_{t+1}}{I^b_{t+1}}\right)\Psi(\bar{\omega}_{t+1},s_{t+1})-\lambda(\bar{\omega}_{t+1},s_{t+1})=0, \tag{A15}$$

where we implicitly use the conjecture that $\bar{\omega}_{t+1}$ is conditioned on variables known at time t. Simply rearranging gives us the following expression:

$$E_t\left(\frac{R^e_{t+1}}{I^b_{t+1}}\right)=\frac{\lambda(\bar{\omega}_{t+1},s_{t+1})}{\Psi(\bar{\omega}_{t+1},s_{t+1})}, \tag{A16}$$

which determines the excess returns per unit of capital above the participation returns on bank loans that would be required to make the financial contract worthwhile to both entrepreneurs-borrowers and banks-lenders.

If we combine equations (A16) and (A13), then it immediately follows that

$$\begin{aligned}\frac{P_tQ_tK_{t+1}}{N_{t+1}}&=\frac{1}{1-\frac{\lambda(\bar{\omega}_{t+1},s_{t+1})}{\Psi(\bar{\omega}_{t+1},s_{t+1})}\left[\frac{\Psi(\bar{\omega}_{t+1},s_{t+1})+J(s_{t+1})+\Gamma(\bar{\omega}_{t+1},s_{t+1})-1}{\lambda(\bar{\omega}_{t+1},s_{t+1})}\right]}\\&=\frac{1}{1-\left\{\frac{\Psi(\bar{\omega}_{t+1},s_{t+1})+\Gamma(\bar{\omega}_{t+1},s_{t+1})-\left[1-J(s_{t+1})\right]}{\Psi(\bar{\omega}_{t+1},s_{t+1})}\right\}},\end{aligned} \tag{A17}$$

which validates our conjecture on the threshold, implying that

$$\bar{\omega}_{t+1}\equiv\bar{\omega}\left(\frac{P_tQ_tK_{t+1}}{N_{t+1}},s_{t+1}\right).$$

Given the relationships in equations (A16) and (A17), a formulation for the external financing premium arises in the following terms:

$$E_t\left(\frac{R^e_{t+1}}{I^b_{t+1}}\right)=s\left(\frac{P_tQ_tK_{t+1}}{N_{t+1}},s_{t+1}\right). \tag{A18}$$

This characterization of the external financing premium expands the Bernanke-Gertler-Gilchrist framework by adding the explicit possibility that the spread itself be affected by the impact of an anticipated aggregate shock, s_{t+1}. The participation return on loans is set at the time the contract is signed, so I^b_{t+1} is known at time t and can be taken out of the expectation. That is,

$$E_t(R^e_{t+1}) = s\left(\frac{P_t Q_t K_{t+1}}{N_{t+1}}, s_{t+1}\right) I^b_{t+1}. \tag{A19}$$

This relationship is the key feature of the financial accelerator model.

References

Berger, A.N. and G. Udell. 1994. "Did Risk-Based Capital Allocate Bank Credit and Cause a 'Credit Crunch' in the United States?" *Journal of Money, Credit, and Banking* 26(3): 585–682.

Bernanke, B.S., M. Gertler, and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics,* vol. 1, edited by J.B. Taylor and M. Woodford. Amsterdam: Elsevier Science.

Bernanke, B.S. and C.S. Lown. 1991. "The Credit Crunch." *Brooking Papers on Economic Activity* 1991(2): 205–48.

Blum, J. and M. Hellwig. 1995. "The Macroeconomic Implications of Capital Adequacy Requirements for Banks." *European Economic Review* 39(3–4): 739–49.

Brinkmann, E.J. and P.M. Horvitz. 1995. "Risk-Based Capital Standards and the Credit Crunch." *Journal of Money, Credit, and Banking* 27(3): 848–64.

Borio, C. and H. Zhu. 2007. "Capital Regulation, Risk-Taking, and Monetary Policy." Paper prepared for the conference on The Implications of Changes in Banking and Financing on the Monetary Policy Transmission. European Central Bank, Frankfurt, 29–30 November.

Calvo, G.A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12(3): 383–98.

Carlstrom, C.T. and T.S. Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review* 87(5): 893–910.

————. 2001. "Monetary Shocks, Agency Costs, and Business Cycles." *Carnegie-Rochester Conference Series on Public Policy* 54(1): 1–27.

Cecchetti, S.G. and L. Li. 2008. "Do Capital Adequacy Requirements Matter for Monetary Policy?" *Economic Inquiry* 46(4): 643–59.

Christiano, L.J., M. Eichenbaum, and C.L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113(1): 1–45.

Curdia, V. and M. Woodford. 2008. "Credit Frictions and Optimal Monetary Policy." Discussion paper 0809-02. Columbia University.

Estrella, A. 2004. "The Cyclical Behavior of Optimal Bank Capital." *Journal of Banking and Finance* 28(6): 1496–98.

Faia, E. and T. Monacelli. 2007. "Optimal Interest Rate Rules, Asset Prices, and Credit Frictions." *Journal of Economic Dynamics* 31(10): 3228–54.

Gale, D. and M. Hellwig. 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem." *Review of Economic Studies* 52(3): 647–63.

Goodhart, C.A.E., P. Surinard, and D.P. Tsomocos. 2006. "A Model to Analyse Financial Fragility." *Economic Theory* 27(1): 107–42.

Gordy, M.B. and B. Howells. 2006. "Procyclicality in Basel II: Can We Treat the Disease without Killing the Patient?" *Journal of Financial Intermediation* 15(3): 396–418.

Kashyap, A. and J.C. Stein. 2004. "Cyclical Implications of the Basel II Capital Standards." *Economic Perspectives* (first quarter): 18–31. Federal Reserve Bank of Chicago.

Kiyotaki, N. and J. Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105(2): 211–48.

Martínez-García, E. and J. Sondergaard. 2008. "The Dynamic Behaviour of the Real Exchange Rate in Sticky Price Models: Does Investment Matter?" Working paper 17. Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute.

Monnet, C. and E. Quintin. 2005. "Optimal Contracts in a Dynamic Costly State Verification Model." *Economic Theory* 26(4): 867–85.

Rudebusch, G.D. 2006. "Monetary Policy Inertia: Fact or Fiction?" *International Journal of Central Banking* 2(4): 85–135.

Taylor, J.B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series* 39(1): 195–214.

Thakor, A.V. 1996. "Capital Requirements, Monetary Policy, and Aggregate Bank Lending: Theory and Empirical Evidence." *Journal of Finance* 51(1): 279–324.

Townsend, R.M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21(2): 265–93.