A Macroeconomic Model with a Financial Sector*

Markus K. Brunnermeier and Yuliy Sannikov†

February 22, 2011

Abstract

This paper studies the full equilibrium dynamics of an economy with financial frictions. Due to highly non-linear amplification effects, the economy is prone to instability and occasionally enters volatile episodes. Risk is endogenous and asset price correlations are high in down turns. In an environment of low exogenous risk experts assume higher leverage making the system more prone to systemic volatility spikes - a volatility paradox. Securitization and derivatives contracts leads to better sharing of exogenous risk but to higher endogenous systemic risk. Financial experts may impose a negative externality on each other by not maintaining adequate capital cushion.

*We thank Nobu Kiyotaki, Hyun Shin, Thomas Philippon, Ricardo Reis, Guido Lorenzoni, Huberto Ennis, V. V. Chari, Simon Potter, Emmanuel Farhi, Monika Piazzesi, Simon Gilchrist, Ben Moll and seminar participants at Princeton, HKU Theory Conference, FESAMES 2009, Tokyo University, City University of Hong Kong, University of Toulouse, University of Maryland, UPF, UAB, CUFE, Duke, NYU 5-star Conference, Stanford, Berkeley, San Francisco Fed, USC, UCLA, MIT, University of Wisconsin, IMF, Cambridge University, Cowles Foundation, Minneapolis Fed, New York Fed, University of Chicago, the Bank of Portugal Conference, Econometric Society World Congress in Shanghai, Seoul National University, European Central Bank and UT Austin. We also thank Wei Cui, Ji Huang, Dirk Paulsen, Andrei Rachkov and Martin Schmalz for excellent research assistance.

†Brunnermeier: Department of Economics, Princeton University, markus@princeton.edu, Sannikov: Department of Economics, Princeton University, sannikov@gmail.com

1 Introduction

Many standard macroeconomic models are based on identical households that invest directly without financial intermediaries. This representative agent approach can only yield realistic macroeconomic predictions if, in reality, there are no frictions in the financial sector. Yet, following the Great Depression, economists such as Fisher (1933), Keynes (1936) and Minsky (1986) have attributed the economic downturn to the failure of financial markets. The current financial crisis has underscored once again the importance of the financial sector for the business cycles.

Central ideas to modeling financial frictions include heterogeneous agents with lending. One class of agents - let us call them experts - have superior ability or greater willingness to manage and invest in productive assets. Because experts have limited net worth, they end up borrowing from other agents who are less skilled at managing or less willing to hold productive assets.

Existing literature uncovers two important properties of business cycles, persistence and amplification. Persistence arises when a temporary adverse shock depresses the economy for a long time. The reason is that a decline in experts' net worth in a given period results in depressed economic activity, and low net worth of experts in the subsequent period. The causes of amplification are leverage and the feedback effect of prices. Through leverage, expert net worth absorbs a magnified effect of each shock, such as new information about the potential future earning power of current investments. When the shock is aggregate, affecting many experts at once, it results in decreased demand for assets and a drop in asset prices, further lowering the net worth of experts, further feeding back into prices, and so on. Thus, each shock passes through this infinite amplification loop, and asset price volatility created through this mechanism is sometimes referred to as endogenous risk. Bernanke and Gertler (1989), Bernanke, Gertler, and Gilchrist (1999) and Kiyotaki and Moore (1997) build a macro model with these effects, and study linearized system dynamics around the steady state.

We build a model to study full equilibrium dynamics, not just near the steady state. While the system is characterized by relative stability, low volatility and reasonable growth around the steady state, its behavior away from the steady state is very different and best resembles crises episodes as large losses plunge the system into a regime with high volatility. These crisis episodes are highly nonlinear, and strong amplifying adverse feedback loops during these incidents may take the system way below the stochastic steady state, resulting in significant inefficiencies, disinvestment, and slow recovery. Interestingly, the stationary distribution is double-humped shaped suggesting that (without government intervention) the dynamical system spends a significant amount of time in the crisis state once thrown there.

The reason why the amplification of shocks through prices is much milder near than below the stochastic steady state is because experts choose their capital cushions endogenously. In the normal regime, experts choose their capital ratios to be able to withstand reasonable losses. Excess profits are paid out (as bonuses, dividends, etc) and mild losses are absorbed by reduced payouts to raise capital cushions to a desired level. Thus, normally experts are fairly unconstrained and are able to absorb moderate shocks to net worth easily, without a significant effect on their demand for assets and market prices. Consequently, for small shocks amplification is limited. However, in response to more significant losses, experts choose to reduce their positions, affecting asset prices and triggering amplification loops. The stronger asset prices react to shocks to the net worth of experts, the stronger the feedback effect that causes further drops in net worth, due to depressed prices. Thus, it follows that below the steady state, when experts feel more constrained, the system becomes less stable as the volatility shoots up. Asset prices exhibit fat tails due to endogenous systemic risk rather than exogenously assumed rare events. This feature causes volatility smirk effects in option prices during the times of low volatility.

Our results imply that endogenous risk and excess volatility created through the amplification loop make asset prices significantly more correlated cross-sectionally in crises than in normal times. While cash flow shocks affect the values of individual assets held by experts, feedback effects affect the prices of all assets held by experts.[1]

We argue that it is typical for the system to enter into occasional volatile episodes away from the steady state because risk-taking is endogenous. This may seem surprising, because one may guess that log-linearization near the steady state is a valid approximation when exogenous risk parameters are small. In our model this guess would be incorrect, because experts choose their leverage endogenously in response to the riskiness of the assets they hold. Thus, assets with lower fundamental uncertainty result in greater leverage. Paradoxically, lower exogenous risk can make the systemic more susceptible to volatility spikes – a phenomenon we refer to as "volatility paradox". In sum, whatever the exogenous risk, it is normal for the system to sporadically enter volatile regimes away from the steady state. In fact, our results suggest that low exogenous risk environment is conducive to greater buildup of systemic risk.

We find that higher volatility due to endogenous risk also increases the experts' precautionary hoarding motive. That is, when changes in asset prices are driven by the constraints of market participants rather than changes in cash flow fundamentals, incentives to hold cash and wait to pick up assets at the bottom increase. In case prices fall further, the same amount of money can buy a larger quantity of assets, and at a lower price, increasing expected return. In our equilibrium this phenomenon leads to price drops in anticipation of the crisis, and higher expected return in times of increased endogenous risk. Aggregate equilibrium leverage is determined by experts' responses to everybody else's leverage – higher aggregate leverage increases endogenous risk, increases the precautionary motive and reduces individual incentives to lever up.[2]

We also find that due to endogenous risk-taking, derivatives hedging, securitization

[1] While our model does not differentiate experts by specialization (so in equilibrium experts hold fully diversified portfolios, leading to the same endogenous correlation across all assets), our results have important implications also for networks linked by similarity in asset holdings. Important models of network effects and contagion include Allen and Gale (2000) and Zawadowski (2009).

[2] The fact that in reality risk taking by leveraged market participants is not observable to others can lead to risk management strategies that are in aggregate mutually inconsistent. Too many of them might be planning to sell their capital in case of an adverse shock, leading to larger than expected price drops. Brunnermeier, Gorton, and Krishnamurthy (2010) argue that this is one contributing factor to systemic risk.

and other forms of financial innovation may make the financial system less stable. That is, volatile excursion away from the steady state may become more frequent with the use of mechanisms that allow intermediaries to share risks more efficiently among each other. For example, securitization of home loans into mortgage-backed securities allows institutions that originate loans to unload some of the risks to other institutions. More generally, institutions can share risks through contracts like credit-default swaps, through integration of commercial banks and investment banks, and through more complex intermediation chains (e.g. see Shin (2010)). To study the effects of these risk-sharing mechanisms on equilibrium, we add idiosyncratic shocks to our model. We find that when expert can hedge idiosyncratic shocks among each other, they become less financially constrained and take on more leverage, making the system less stable. Thus, while securitization is in principle a good thing - it reduces the costs of idiosyncratic shocks and thus interest rate spreads - it ends up amplifying systemic risks in equilibrium.

Financial frictions in our model lead not only to amplification of exogenous risk through endogenous risk but also to inefficiencies. Externalities can be one source of inefficiencies as individual decision makers do not fully internalize the impact of their actions on others. Pecuniary externalities arise since individual market participants take prices as given, while as a group they affect them.

Literature review. Financial crises are common in history - having occurred at roughly 10-year intervals in Western Europe over the past four centuries, according Kindleberger (1993). Crises have become less frequent with the introduction of central banks and regulation that includes deposit insurance and capital requirements (see Allen and Gale (2007) and Cooper (2008)). Yet, the stability of the financial system has been brought into the spotlight again by the events of the current crises, see Brunnermeier (2009).

Financial frictions can limit the flow of funds among heterogeneous agents. Credit and collateral constraints limit the debt capacity of borrowers, while equity constraints bound the total amount of outside equity. Both constraints together imply the solvency constraint. That is, net worth has to be nonnegative all the time. The literature on credit constraints typically also assumes that firms cannot issue any equity. In addition, in Kiyotaki and Moore (1997) credit is limited by the expected price of the collateral in the next period. In Geanakoplos (1997, 2003) and Brunnermeier and Pedersen (2009) borrowing capacity is limited by possible adverse price movement in the next period. Hence, greater future price volatility leads to higher haircuts and margins, further tightening the liquidity constraint and limiting leverage. Garleanu and Pedersen (2010) study asset price implications for an exogenous margin process. Shleifer and Vishny (1992) argue that when physical collateral is liquidated, its price is depressed since natural buyers, who are typically in the same industry, are likely to be also constrained. Gromb and Vayanos (2002) provide welfare analysis for a setting with credit constraints. Rampini and Viswanathan (2011) show that highly productive firms go closer to their debt capacity and hence are harder hit in a downturns. In Carlstrom and Fuerst (1997) and Bernanke, Gertler, and Gilchrist (1999) entrepreneurs do not face a credit constraint but debt becomes more expensive as with higher debt level default probability increases.

In this paper experts can issue some equity but have to retain "skin in the game" and hence can only sell off a fraction of the total risk. In Shleifer and Vishny (1997) fund managers are also concerned about their equity constraint binding in the future. He and Krishnamurthy (2010b,a) also assume an equity constraint.

One major role of the financial sector is to mitigate some of the financial frictions. Like Diamond (1984) and Holmström and Tirole (1997) we assume that financial intermediaries have a special monitoring technology to overcome some of the frictions. However, the intermediaries' ability to reduce these frictions depends on their net worth. In Diamond and Dybvig (1983) and Allen and Gale (2007) financial intermediaries hold long-term assets financed by short-term liabilities and hence are subject to runs, and He and Xiong (2009) model general runs on non-financial firms. In Shleifer and Vishny (2010) banks are unstable since they operate in a market influenced by investor sentiment.

Many papers have studied the amplification of shocks through the financial sector near the steady state, using log-linearization. Besides the aforementioned papers, Christiano, Eichenbaum, and Evans (2005), Christiano, Motto, and Rostagno (2003, 2007), Curdia and Woodford (2009), Gertler and Karadi (2009) and Gertler and Kiyotaki (2011) use the same technique to study related questions, including the impact of monetary policy on financial frictions.

We argue that the financial system exhibits the types of instabilities that cannot be adequately studied by steady-state analysis, and use the recursive approach to solve for full equilibrium dynamics. Our solution builds upon recursive macroeconomics, see Stokey and Lucas (1989) and Ljungqvist and Sargent (2004). We adapt this approach to study the financial system, and enhance tractability by using continuous-time methods, see Sannikov (2008) and DeMarzo and Sannikov (2006).

A few other papers that do not log-linearize include Mendoza (2010) and He and Krishnamurthy (2010b,a). Perhaps most closely related to our model is He and Krishnamurthy (2010b). The latter studies an endowment economy to derive a two-factor asset pricing model for assets that are exclusively held by financial experts. Like in our paper, financial experts issue outside equity to households but face an equity constraint due to moral hazard problems. When experts are well capitalized, risk premia are determined by aggregate risk aversion since the outside equity constraint does not bind. However, after a severe adverse shock experts, who cannot sell risky assets to households, become constrained and risk premia rise sharply. He and Krishnamurthy (2010a) calibrate a variant of the model and show that equity injection is a superior policy compared to interest rate cuts or asset purchasing programs by the central bank.

Pecuniary externalities that arise in our setting lead to socially inefficient excessive borrowing, leverage and volatility. These externalities are studied in Bhattacharya and Gale (1987) in which externalities arise in the interbank market and in Caballero and Krishnamurthy (2004) which study externalities an international open economy framework. On a more abstract level these effects can be traced back to inefficiency results within an incomplete markets general equilibrium setting, see e.g. Stiglitz (1982) and Geanakoplos and Polemarchakis (1986). In Lorenzoni (2008) and Jeanne and Korinek (2010) funding constraints depend on prices that each individual investor takes as given. Adrian and Brunnermeier (2010) provide a systemic risk measure and argue that financial regulation should focus on these externalities.

We set up our baseline model in Section 2. In Section 3 we develop methodology to solve the model, and characterize the equilibrium that is Markov in the experts' aggregate net worth and presents a computed example. Section 4 discusses equilibrium asset allocation and leverage, endogenous and systemic risk and equilibrium dynamics in normal as well as crisis times. We also extend the model to multiple assets, and show that endogenous risk makes asset prices much more correlated in cross-section in crisis times. In Section 5 focuses on the "volatility paradox". We show that the financial system is always prone to instabilities and systemic risk due endogenous risk taking. We also argue that hedging of risks within the financial sector, while reducing inefficiencies from idiosyncratic risks, may lead to the amplification of systemic risks. Section 6 is devoted to efficiency and externalities. Section 7 microfounds experts' balance sheets in the form that we took as given in the baseline model, and extend analysis to more complex intermediation chains. Section 8 concludes.

2 The Baseline Model

In an economy without financial frictions and complete markets, the distribution of net worth does not matter as the flow of funds to the most productive agents is unconstrained. In our model financial frictions limit the flow of funds from less productive households to more productive entrepreneurs. Hence, higher net worth in the hands of the entrepreneurs leads to higher overall productivity. In addition, financial intermediaries can mitigate financial frictions and improve the flow of funds. However, they need to have sufficient net worth on their own. In short, the two key variables in our economy are entrepreneurs' net worth and financial intermediaries' net worth. When the net worth's of intermediaries and entrepreneurs become depressed, the allocation of resources (such as capital) in the economy becomes less efficient and asset prices become depressed.

In our baseline model we study equilibrium in a simpler system governed by a single state variable, "expert" net worth. We interpret it as an aggregate of intermediary and entrepreneur net worth's. In Section 7 we partially characterize equilibrium in a more general setting and provide conditions under which the more general model of intermediation reduces to our baseline setting.

Technology. We consider an economy populated by experts and less productive households. Both types of agents can own capital, but experts are able to manage it more productively. The experts' ability to hold capital and equilibrium asset prices will depend on the experts' net worths in our model.

We denote the aggregate account of efficiency units of capital in the economy by K_t, where $t \in [0, \infty)$ is time, and capital held by an individual agent by k_t. Physical capital k_t held by experts produces output at rate

$$y_t = ak_t,$$

per unit of time, where a is a parameter. The price of output is set equal to one and serves as numeraire. Experts can create new capital through internal investment. When held by an expert, capital evolves according to

$$dk_t = (\Phi(\iota_t) - \delta)k_t\, dt + \sigma k_t\, dZ_t$$

where $\iota_t k_t$ is the investment rate (i.e. ι_t is the investment rate per unit of capital), the function $\Phi(\iota_t)$ reflects (dis)investment costs and dZ_t are exogenous Brownian aggregate shocks. We assume that that $\Phi(0) = 0$, so in the absence of new investment capital depreciates at rate δ when managed by experts, and that the function $\Phi(\cdot)$ is increasing and concave. That is, the marginal impact of internal investment on capital is decreasing when it is positive, and there is "technological illiquidity," i.e. large-scale disinvestments are less effective, when it is negative.

Households are less productive and do not have an internal investment technology. The capital that is managed by households produces only output of

$$\underline{y}_t = \underline{a}\, \underline{k}_t$$

with $\underline{a} \leq a$. In addition, capital held in households' hands depreciates at a faster rate $\underline{\delta} \geq \delta$. The law of motion of k_t when managed by households is

$$d\underline{k}_t = -\underline{\delta}\, \underline{k}_t\, dt + \sigma \underline{k}_t\, dZ_t.$$

The Brownian shocks dZ_t reflect the fact that one learns over time how "effective" the capital stock is.[3] That is, the shocks dZ_t captures changes in expectations about the future productivity of capital, and k_t reflects the "efficiency units" of capital, measured in expected future output rather than in simple units of physical capital (number of machines). For example, when a company reports current earnings it not only reveals information about current but also future expected cashs flow. In this sense our model is also linked to the literature on connects news to business cycles, see e.g. Jaimovich and Rebelo (2009).

Preferences. Experts and less productive households are risk neutral. Households discount future consumption at rate r, and they may consume both positive and negative amounts. This assumption ensures that households provide fully elastic lending at the risk-free rate of r. Denote by $\underline{c}_t$ the cumulative consumption of an individual household until time t, so that $d\underline{c}_t$ is consumption at time t. Then the utility of a household is given by[4]

$$E\left[\int_0^\infty e^{-rt}\, d\underline{c}_t\right].$$

[3]Alternatively, one can also assume that the economy experiences aggregate TFP shocks a_t with $da_t = a_t \sigma dZ_t$. Output would be $y_t = a_t \kappa_t$, where capital κ is now measured in physical (instead of efficiency) units and evolves according to $d\kappa_t = (\Phi(\iota_t/a_t) - \delta)\kappa_t dt$. To preserve the tractable scale invariance property one has to modify the adjustment cost function to $\Phi(\iota_t/a_t)$. The fact that adjustment costs are higher for high a_t can be justified by the fact that high TFP economies are more specialized.

[4]Note that we do not denote by $c(t)$ the flow of consumption and write $E\left[\int_0^\infty e^{-\rho t} c(t)\, dt\right]$, because consumption can be lumpy and singular and hence $c(t)$ may be not well defined.

In contrast, experts discount future consumption at rate $\rho > r$, and they cannot have negative consumption. That is, cumulative consumption of an individual expert c_t must be a *nondecreasing* process, i.e. $dc_t \geq 0$. Expert utility is

$$E\left[\int_0^\infty e^{-\rho t}\, dc_t\right].$$

Market for Capital. There is a fully liquid market for physical capital, in which experts can trade capital among each other or with households. Denote the market price of capital (per efficiency unit) in terms of output by q_t and its law of motion by[5]

$$dq_t = \mu_t^q q_t \;\, dt + \sigma_t^q q_t \, dZ_t$$

In equilibrium q_t is determined *endogenously* through supply and demand relationships. Moreover, $q_t > \underline{q} \equiv \underline{a}/(r + \underline{\delta})$, since even if households had to hold the capital forever, the Gordon growth formula tells us that they would be willing to pay $\underline{q}$.

When an expert buys and holds k_t units of capital at price q_t, by Ito's lemma the value of this capital evolves according to[6]

$$d(k_t q_t) = (\Phi(\iota_t) - \delta + \mu_t^q + \sigma\sigma_t^q)(k_t q_t)\, dt + (\sigma + \sigma_t^q)(k_t q_t)\, dZ_t. \tag{1}$$

Note that the total risk of holding this position in capital consists of *fundamental* risk due to news about the future productivity of capital $\sigma\, dZ_t$, and *endogenous* risk due to the allocation of capital between experts and less productive households, $\sigma_t^q\, dZ_t$. Capital also generates output net of investment of $(a - \iota_t)k_t$, so the total return from one unit of wealth invested in capital is

$$\underbrace{\left(\frac{a - \iota_t}{q_t} + \Phi(\iota_t) - \delta + \mu_t^q + \sigma\sigma_t^q\right)}_{\equiv E_t[r_t^k]} dt + (\sigma + \sigma_t^q)\, dZ_t.$$

We denote the experts' expected return on capital by $E_t[r_t^k]$.

Experts' problem. The evolution of expert's net worth n_t depends on how much debt and equity he issues. Less productive households provide fully elastic debt funding at a discount rate $r < \rho$ to any expert with positive net worth, as long as he can guarantee to repay the loan with probability one.[7]

[5]Note that q_t follows a diffusion process because all new information in our economy is generated by the Brownian motion Z_t.

[6]The version of Ito's lemma we use is the product rule $d(X_t Y_t) = Y_t dX_t + X_t dY_t + \sigma_x \sigma_y\, dt$. Note that unlike in standard portfolio theory, k_t is not a finite variation process and has volatility σk_t, hence the term $\sigma\sigma_t^q(k_t q_t)$.

[7]In the short run, an individual expert can hold an arbitrarily large amount of capital by borrowing through risk-free debt because prices change continuously in our model, and individual experts are small and have no price impact.

For an expert who only finances his capital holding of $q_t k_t$ through debt, without issuing any equity, the net worth evolves according to

$$dn_t = rn_t\,dt + (k_t q_t)[(E_t[r_t^k] - r)\,dt + (\sigma + \sigma_t^q)\,dZ_t] - dc_t. \tag{2}$$

In this equation, the exposure to capital k_t may change over time due to trading, but trades themselves do not affect expert net worth because we assume that individual experts are small and have no price impact. The terms in the square brackets reflect the excess return from holding one unit of capital.

Experts can in addition issue some (outside) equity. Equity financing leads to a modified equation for the law of motion of expert net worth. We assume that the amount of equity that experts can issue is limited. Specifically, they are required to hold at least a fraction of $\tilde{\varphi}$ of *total* risk of the capital they hold, and they are able to invest in capital only when their net worth is *positive*. That is, experts are bound by an *equity* constraint and a *solvency* constraint. In Section 7 we microfound these financing constraints using an agency model, and explain its relation to contracting and observability and also fully model the intermediary sector that monitors and lends to more productive households.

When experts holds a fraction $\varphi_t \geq \tilde{\varphi}$ of capital risk and unload the rest to less productive households through equity issuance, the law of motion of expert net worth (2) has to be modified to

$$dn_t = rn_t\,dt + (k_t q_t)[(E_t[r_t^k] - r)\,dt + \varphi_t(\sigma + \sigma_t^q)\,dZ_t] - dc_t. \tag{3}$$

Equation (3) takes into account that, since less productive households are risk-neutral, they require only an expected return of r on their equity investment. Figure 1 illustrates the balance sheet of an individual expert at a fixed moment of time t.[8]

[8]Equation (3) captures the essence about the evolution of experts' balance sheets. To fully characterize the full mechanics note first that equity is divided into *inside equity* with value n_t, which is held by the expert and *outside equity*, with value $(1-\varphi_t)n_t/\varphi_t$, held by less productive households. At any moment of time t, an expert holds capital with value $k_t q_t$ financed by equity n_t/φ_t and debt $k_t q_t - n_t/\varphi_t$. The equity stake of less productive households changes according to

$$r(1-\varphi_t)/\varphi_t n_t\,dt + (1-\varphi_t)(k_t q_t)(\sigma + \sigma_t^q)\,dZ_t - (1-\varphi_t)/\varphi_t dc_t,$$

where $(1-\varphi_t)/\varphi_t dc_t$ is the share of dividend payouts that goes to outside equity holders.

Since the expected return on capital held by experts is higher than the risk-free rate, inside equity earns a higher return than outside equity. This difference can be implemented through a fee paid by outside equity holders to the expert for managing assets. From equation (3), the earnings of inside equity in excess of the rate of return r are

$$(k_t q_t)(E_t[r_t^k] - r).$$

Thus, to keep the ratio of outside equity to inside equity at $(1-\varphi_t)/\varphi_t$, the expert has to raise outside equity at rate

$$(1-\varphi_t)/\varphi_t(k_t q_t)(E_t[r_t^k] - r).$$

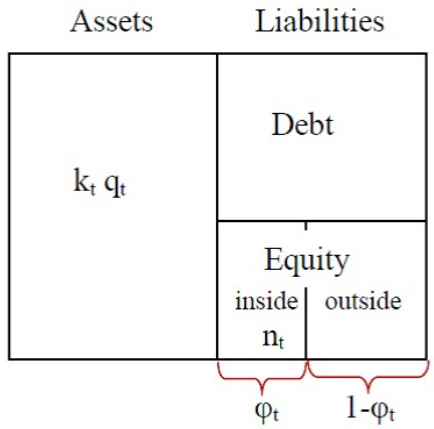

Figure 1: Expert balance sheet with inside and outside equity

Formally, each expert solves

$$\max_{dc_t \geq 0, \iota_t, k_t \geq 0, \varphi_t \geq \tilde{\varphi}} E\left[\int_0^\infty e^{-\rho t} dc_t\right],$$

subject to the solvency constraint $n_t \geq 0, \forall t$ and the dynamic budget constraint (3).

Households' problem. Each household may lend to experts at the risk-free rate r, buy experts' outside equity, or buy physical capital from experts. Let $\underline{\xi}_t$ denote the amount of risk that the household is exposed to through its holdings of outside equity of experts and $d\underline{c}_t$ is the consumption of an individual household. When a household with net worth $\underline{n}_t$ buys capital $\underline{k}_t$ and invests the remaining net worth, $\underline{n}_t - \underline{k}_t q_t$ at the risk-free rate and in experts' outside equity, then

$$d\underline{n}_t = r\underline{n}_t\, dt + \underline{\xi}_t(\sigma + \sigma_t^q)\, dZ_t + (\underline{k}_t q_t)[(E_t[\underline{r}_t^k] - r)\, dt + (\sigma + \sigma_t^q)\, dZ_t] - d\underline{c}_t. \quad (4)$$

Analogous to experts, we denote households' expected return of capital by

$$E_t[\underline{r}_t^k] \equiv \frac{\underline{a}}{q_t} - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q.$$

Formally, each household solves

$$\max_{d\underline{c}_t, \underline{k}_t \geq 0, \underline{\xi}_t \geq 0} E\left[\int_0^\infty e^{-rt} d\underline{c}_t\right],$$

subject to $\underline{n}_t \geq 0$ and the evolution of $\underline{n}_t$ given by (4). Note that unlike that of experts, household consumption $d\underline{c}_t$ can be both positive and negative.

In sum, experts and households differ in three ways: First, experts are more productive since $a \geq \underline{a}$ and/or $\delta < \underline{\delta}$. Second, experts are less patient than households, i.e. $\rho > r$. Third, experts' consumption has to be positive while we allow for negative households consumption to ensure that the risk free rate is always r.[9]

[9]Negative consumption could be interpreted as the disutility from an additional labor input to produce extra output.

Equilibrium. Informally, an equilibrium is characterized by market prices of capital $\{q_t\}$, investment and consumption choices of agents such that, given prices, agents maximize their expected utilities and markets clear. To define an equilibrium formally, we denote the set of experts to be the interval $\mathbb{I} = [0, 1]$, and index individual experts by $i \in \mathbb{I}$, and similarly denote the set of less productive households by $\mathbb{J} = (1, 2]$ with index j.

Definition 1 *For any initial endowments of capital $\{k_0^i, \underline{k}_0^j; i \in \mathbb{I},\ j \in \mathbb{J}\}$ such that*

$$\int_{\mathbb{I}} k_0^i \, di + \int_{\mathbb{J}} \underline{k}_0^j \, dj = K_0,$$

an equilibrium is described by a group of stochastic processes on the filtered probability space defined by the Brownian motion $\{Z_t, t \geq 0\}$: the price process of capital $\{q_t\}$, net worths $\{n_t^i \geq 0\}$, capital holdings $\{k_t^i \geq 0\}$, investment decisions $\{\iota_t^i \in \mathbb{R}\}$, fractions of equity retained $\{\varphi_t^i \geq \tilde{\varphi}\}$ and consumption choices $\{dc_t^i \geq 0\}$ of individual experts $i \in \mathbb{I}$, and net worths $\{\underline{n}_t^j\}$, capital holdings $\{\underline{k}_t^j\}$, investments in outside equity $\{\underline{\xi}_t^j\}$ and consumption choices $\{d\underline{c}_t^j\}$ of each less productive household $j \in \mathbb{J}$; such that

(i) initial net worths satisfy $n_0^i = q_0 k_0^i$ and $\underline{n}_0^j = q_0 \underline{k}_0^j$, for $i \in \mathbb{I}$ and $j \in \mathbb{J}$,

(ii) each expert $i \in \mathbb{I}$ solve his problem given prices

(iii) each household $j \in \mathbb{J}$ solve his problem given prices

(iv) markets for consumption goods,[10] equity, and capital clear

$$\int_{\mathbb{I}} (dc_t^i) \, di + \int_{\mathbb{J}} (d\underline{c}_t^j) \, dj = \left(\int_{\mathbb{I}} (a - \iota_t^i) k_t^i \, di + \int_{\mathbb{J}} \underline{a} \ \underline{k}_t^j \, dj \right) dt,$$

$$\int_{\mathbb{I}} (1 - \varphi_t^i) k_t^i \, di = \int_{\mathbb{J}} \underline{\xi}_t^j \, dj, \qquad \text{and} \qquad \int_{\mathbb{I}} k_t^i \, di + \int_{\mathbb{J}} \underline{k}_t^j \, dj = K_t,$$

where $\quad dK_t = \left(\int_{\mathbb{I}} k_t^i (\Phi(\iota_t) - \delta) \, di - \int_{\mathbb{J}} \underline{\delta} \ \underline{k}_t^j \, dj \right) dt + \sigma K_t \, dZ_t.$

Note that if three of the markets clear, then the remaining market for risk-free lending and borrowing at rate r automatically clears by Walras' Law.

3 Solving for the Equilibrium

To solve for the equilibrium, we first derive conditions for households' and experts' optimal capital holding given prices q_t, and use them together with the market-clearing conditions to solve for prices, and investment and consumption choices simultaneously. We proceed in two steps. First, we derive equilibrium conditions that the stochastic

[10] In equilibrium while aggregate consumption is continuous with respect to time, the experts' and households' consumption is not. However, their singular parts cancel out in the aggregate.

equations for the price of capital and the marginal value of net worth have to satisfy in general. Second, we show that the dynamics of our basic setup can be described by a single state variable and derive the system of equations to solve for the price of capital and the marginal value of net worth as functions of this state variable.

Intuitively, we expect the equilibrium prices to fall after negative macro shocks, because those shocks lead to expert losses and make them more constrained. At some point, prices may drop so far that less productive households may find it profitable to buy capital from experts. Less productive households are speculative as they hope to make capital gains. In this sense they are liquidity providers as they pick up some of the functions of the traditional financial sector in times of crises.[11]

Households' optimization problem is straightforward as they are not financially constrained. In equilibrium they must earn a return of r, their discount rate, on investments in the risk-free assets and expert's equity. Their expected return on physical capital cannot exceed r in equilibrium, since otherwise they would demand an infinite amount of capital. Formally, denote the fraction of physical capital held by households by

$$1 - \psi_t = \frac{1}{K_t} \int_{\mathbb{J}} \underline{k}_t^j \, dj.$$

Households expected return has to be exactly r when $1 - \psi_t > 0$, and not greater than r when $1 - \psi_t = 0$. This leads to the equilibrium condition

$$\underbrace{\frac{a}{q_t} - \underline{\delta} + \mu_t^q + \sigma \sigma_t^q}_{E_t[\underline{r}_t^k]} \le r, \text{ with equality if } 1 - \psi_t > 0. \tag{H}$$

Experts' optimization problems are significantly more complex because experts are financially constrained and the problem that they face is *dynamic*. That is, their decisions on how much to lever up depend not only on the current price levels and their production technologies, but also on the whole future law of motion of prices. They face the following trade-off: greater leverage leads to both higher profit and greater risk. Even though experts are risk-neutral with respect to consumption streams in our model, our analysis shows that they exhibit risk-averse behavior (in aggregate) because their investment opportunities are time-varying. Taking on greater risk leads experts to suffer greater losses exactly in the events when they value funds the most - after negative shocks when prices become depressed and profitable opportunities arise.

Before discussing dynamic optimality of experts' strategies, note that one choice that experts make, internal investment ι_t, is static. Optimal investment maximizes

$$k_t q_t \Phi(\iota_t) - k_t \iota_t.$$

[11]Investors like Warren Buffet have helped institutions like Goldman Sachs and Wells Fargo with capital infusions. More generally, governments through backstop facilities have played a huge role in providing capital to financial institutions in various ways and induced large shifts in asset holdings (see He, Khang, and Krishnamurthy (2010)). Our model does not capture the important role the government played in providing various lending facilities during the great recession.

The first-order condition is $q_t \Phi'(\iota_t) = 1$ (marginal Tobin's q) which implies that the optimal level of investment and the resulting growth rate of capital are functions of the price q_t, i.e.

$$\iota_t = \iota(q_t) \text{ and } \Phi(\iota_t) - \delta = g(q_t).$$

From now on, we assume that experts are optimizing with respect to internal investment, and take $E_t[r_t^k]$ to incorporate the optimal choice of ι_t.

Unlike internal investment, expert choices with respect to the trading of capital k_t, consumption dc_t and the fraction of risk $\varphi_t \geq \varphi$ they hold are fully dynamic.[12] To solve the experts' dynamic optimization problems, we define the experts' value functions and write their Bellman equations. The value function of an expert summarizes how his continuation payoff depends on his wealth and market conditions. The following lemma highlights an important property of the expert value functions: they are proportionate to their wealth, because of the assumption that experts are atomistic and act competitively. That is, expert A whose wealth differs from that of expert B by a factor of ς can get the payoff of expert B times ς by scaling the strategy of expert B proportionately. We denote the proportionality coefficient that summarizes how market conditions affect the experts' expected payoff per dollar of net worth by the process θ_t. The process θ_t is determined endogenously in equilibrium.

Lemma 1 *There exists a process θ_t such that the value function of any expert with net worth n_t is of the form $\theta_t n_t$.*

Lemma 2 characterizes expert optimization problem via the Bellman equation.

Lemma 2 *Let $\{q_t, t \geq 0\}$ be a price process for which the experts' value functions are finite.*[13] *Then the following two statements are equivalent*

(i) the process $\{\theta_t, t \geq 0\}$ represents the marginal value of net worth and $\{k_t, dc_t, \varphi_t, \iota_t; t \geq 0\}$ is an optimal strategy

(ii) the Bellman equation

$$\rho\theta_t n_t \, dt = \max_{\substack{k_t \geq 0,\, dc_t \geq 0,\, \varphi_t \geq \tilde{\varphi} \\ \text{s.t. (3) holds}}} dc_t + E[d(\theta_t n_t)], \tag{5}$$

together with transversality condition that $E[e^{-\rho t}\theta_t n_t] \to 0$ as $t \to \infty$ hold.

From the Bellman equation, we can derive more specific conditions that stochastic laws of motion of q_t and θ_t, together with the experts' optimal strategies, have to satisfy. We conjecture that in equilibrium $\sigma_t^q \geq 0$, $\sigma_t^\theta \leq 0$ and $\psi_t > 0$, i.e. capital prices rise after positive macro shocks (which make experts less constrained) and drop after negative shocks, the marginal value of expert net worth rises when prices fall,

[12] Of these choices, the fraction of risk that the experts retain is straightforward, $\varphi_t = \tilde{\varphi}$, as we verify later. That is, experts wish to minimize their exposure to aggregate risk.

[13] In our setting, because experts are risk-neutral, their value functions under many price processes can be easily infinite (although, of course, in equilibrium they are finite).

and experts always hold positive amounts of capital. Under these assumptions we derive necessary and sufficient conditions for the optimality of expert' strategies in the following proposition.

Proposition 1 *Consider a pair of processes*

$$\frac{dq_t}{q_t} = \mu_t^q\, dt + \sigma_t^q\, dZ_t \quad \text{and} \quad \frac{d\theta_t}{\theta_t} = \mu_t^\theta\, dt + \sigma_t^\theta\, dZ_t$$

such that $\sigma_t^q \geq 0$ *and* $\sigma_t^\theta \leq 0$. *Then* $\theta_t < \infty$ *represents the expert's marginal value of net worth and* $\{k_t \geq 0, dc_t > 0, \varphi_t \geq \tilde{\varphi}\}$ *is an optimal strategy if and only if*

(i) $\theta_t \geq 1$ *at all times, and* $dc_t > 0$ *only when* $\theta_t = 1$,

(ii)
$$\mu_t^\theta = \rho - r \tag{E}$$

(iii) either
$$\underbrace{\frac{a-\iota}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q - r}_{\text{expected excess return on capital, } E[r_t^k]-r} = \underbrace{-\tilde{\varphi}\sigma_t^\theta(\sigma+\sigma_t^q)}_{\text{risk premium}} \tag{EK}$$

$k_t > 0$ *and* $\varphi_t = \tilde{\varphi}$, *or* $E[r_t^k] - r \leq -\tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q)$ *and* $k_t = 0$,[14]

(iv) and the transversality condition holds.

Our definition of an equilibrium requires three conditions: household and expert optimization and market clearing. Household problem is characterized by condition (H), that of experts, by conditions (E) and (EK) of Proposition 1. According to Proposition 1, as long as (EK) holds, any nonnegative amount of capital in experts' portfolio is consistent with experts' utility maximization, so markets for capital clear automatically. Markets for consumption clear because the risk-free rate is r and households' consumption may be positive or negative, and markets for expert's outside equity clear because it generates an expected return of r.

Proof. Consider a process θ_t that satisfies the Bellman equation, and let us justify (i) through (iii). For (i), θ_t can never be less than 1 because an expert can guarantee a payoff of n_t by consuming his entire net worth immediately. When $\theta_t > 1$, then the maximization problem inside the Bellman equation requires that $dc_t = 0$. Intuitively, when the marginal value of an extra dollar is worth more on the expert's balance sheet, it is not optimal to consume. Therefore, (i) holds.

Using the laws of motion of θ_t and n_t as well as Ito's lemma, we transform the Bellman equation to

$$\begin{aligned} \rho\theta_t n_t \quad = & \quad \max_{k_t \geq 0, \varphi_t \geq \tilde{\varphi}} \theta_t \left\{ r n_t + (k_t q_t) \left(\frac{a - \iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q - r \right) \right\} \\ & + \quad \theta_t \mu_t^\theta n_t + \sigma_t^\theta \theta_t\, (k_t q_t) \varphi_t (\sigma + \sigma_t^q) + \underbrace{\max_{dc_t \geq 0} \left(dc_t - \theta_t dc_t \right)}_{0}. \end{aligned}$$

[14]Without the assumptions that $\sigma_t^q \geq 0$ and $\sigma_t^\theta \leq 0$, condition (iii) has to be replaced with $\max E[r_t^k] - r + \varphi_t \sigma_t^\theta(\sigma + \sigma_t^q) \leq 0$, with strict inequality only if $k_t = 0$.

When some value $k_t > 0$ solves the maximization problem above, then (EK) must hold as the first-order condition with respect to k_t but with φ_t instead of $\tilde{\varphi}$. Moreover, because $\sigma_t^\theta(\sigma + \sigma_t^q) \le 0$, it follows that $\varphi_t = \tilde{\varphi}$ maximizes the right hand side. When (EK) holds then any value of k_t maximizes the right hand side, and we obtain

$$\rho\theta_t n_t = \theta_t r n_t + \theta_t \mu_t^\theta n_t \Rightarrow \rho - r = \mu_t^\theta.$$

When only $k_t = 0$ solves the maximization problem in the Bellman equation, then $E[r_t^k] - r < -\tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q)$, because otherwise it would be possible to set $\varphi_t = \tilde{\varphi}$ and increase k_t above 0 without hurting the right hand side of the Bellman equation. With $k_t = 0$, the Bellman equation also implies $\rho - r = \mu_t^\theta$.

Conversely, it is easy to show that if (i) through (iii) hold then the Bellman equation also holds. ■

Equation (EK) is instructive. Experts earn profit by levering up to buy capital, but at the same time taking risk. The risk is that they lose $\tilde{\varphi}(\sigma+\sigma_t^q)\,dZ_t$ per dollar invested in capital exactly in the event that better investment opportunities arise as θ_t goes up by $\sigma_t^\theta\theta_t\,dZ_t$. Thus, while the left hand side of (EK) reflects the experts' incentives to hold more capital, the expression $\tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q)$ on the right hand side reflects the experts' *precautionary* motive. If endogenous risk ever made the right hand side of (EK) greater than the left hand side, experts would choose to hold cash in volatile times waiting to pick up assets at low prices at the bottom ("flight to quality"). The subsequent analysis shows how this trade-off leads to an equilibrium choice of leverage, because individual experts' incentives to take risk are decreasing in the risks taken by other experts in the aggregate.

While not directly relevant to our derivation of the equilibrium, it is interesting to note that θ_t can be related to the stochastic discount factor (SDF) that experts use to price assets. Note that experts are willing to pay price

$$\theta_t x_t = E_t[e^{-\rho s}\theta_{t+s}x_{t+s}]$$

for an asset that pays x_{t+s} at time $t+s$, since their marginal value of a dollar of net worth at time t is θ_t and at time $t+s$, θ_{t+s}. Thus, $e^{-\rho s}\theta_{t+s}/\theta_t$ is the experts' stochastic discount factor (SDF) at time t, which prices all assets that experts invest in (i.e. capital minus the outside equity and the risk-free asset).[15]

Scale Invariance. Define the aggregate net worth of experts in our model by

$$N_t \equiv \int_{\mathbb{I}} n_t^i di,$$

and the level of expert net-worth per unit of aggregate capital by

$$\eta_t \equiv \frac{N_t}{K_t}.$$

[15] Note that returns are linear in portfolio weights in our basic model. With decreasing returns the SDF $e^{-\rho s}\theta_{t+s}/\theta_t$ prices only the experts' optimal portfolios under optimal leverage.

Our model has scale-invariance properties, which imply that inefficiencies with respect to investment and capital allocation as well as that the level of prices depend on η_t. That is, under our assumptions an economy with aggregate expert net worth ςN_t and aggregate capital ςK_t has the same properties as an economy with aggregate expert net worth N_t and capital K_t, scaled by a factor of ς. More specifically, if (q_t, θ_t) is an equilibrium price-value function pair in an economy with aggregate expert net worth N_t and capital K_t, then it can be an equilibrium pair also in an economy with aggregate expert net worth ςN_t and aggregate capital ςK_t.

We will characterize an equilibrium that is Markov in the state variable η_t. Before we proceed, Lemma 3 derives the equilibrium law of motion of $\eta_t = N_t/K_t$ from the equations for dN_t and dK_t. In Lemma 3, we do not assume that the equilibrium is Markov.[16]

Lemma 3 *The equilibrium law of motion of η_t is*

$$d\eta_t = \mu_t^\eta \eta_t \, dt + \sigma_t^\eta \eta_t \, dZ_t - d\zeta_t, \tag{6}$$

where

$$\mu_t^\eta = r - \psi_t g(q_t) + (1-\psi_t)\underline{\delta} + \psi_t \frac{q_t}{\eta_t}(E_t[r_t^k] - r) - \sigma \sigma_t^\eta,$$

$$\sigma_t^\eta = \frac{\psi_t \varphi_t q_t}{\eta_t}(\sigma + \sigma_t^q) - \sigma, \qquad d\zeta_t = \frac{dC_t}{K_t},$$

$dC_t = \int_{\mathbb{I}} (dc_t^i)\, di$ are aggregate payouts to experts and $\varphi_t = \frac{1}{\psi_t K_t} \int_{\mathbb{I}} (\varphi_t^i k_t^i)\, di$. Moreover, if $\sigma_t^q \geq 0, \sigma_t^\theta \leq 0$ and $\psi_t > 0$, then expert optimization implies that $\varphi_t = \tilde{\varphi}$ and

$$\mu_t^\eta = r - \psi_t g(q_t) + (1-\psi_t)\underline{\delta} - \sigma_t^\eta(\sigma + \sigma_t^\theta) - \sigma\sigma^\theta.$$

Markov Equilibrium. Because of scale invariance, it is natural to look for an equilibrium that is Markov in the state variable η_t. In a Markov equilibrium, q_t, θ_t and ψ_t are functions of η_t, so

$$q_t = q(\eta_t), \; \theta_t = \theta(\eta_t) \text{ and } \psi_t = \psi(\eta_t).$$

Equation (5), the law of motion of η_t, expresses how the state variable η_t is determined by the path of aggregate shocks $\{Z_s, s \leq t\}$, and q_t, θ_t and ψ_t are determined by η_t. In the following proposition, we characterize a Markov equilibrium via a system of differential equations. We conjecture that $\sigma_t^q \geq 0$, $\sigma_t^\theta \leq 0$ and $\psi_t > 0$ and use conditions (E), (EK) and (H) together with Ito's lemma to mechanically express μ_t^q, μ_t^θ, σ_t^q, and σ_t^θ through the derivatives of $q(\eta)$ and $\theta(\eta)$.

[16] We conjecture that the Markov equilibrium we derive in this paper is unique, i.e. there are no other equilibria in the model (Markov or non-Markov). While the proof of uniqueness is beyond the scope of the paper, a result like Lemma 3 should be helpful for the proof of uniqueness.

Proposition 2 *The equilibrium domain of functions $q(\eta)$ and $\theta(\eta)$ is an interval $[0, \eta^*]$. For $\eta \in [0, \eta^*]$, these functions can be computed from the differential equations*

$$q''(\eta) = \frac{2(\mu_t^q q_t - q'(\eta)\mu_t^\eta \eta)}{(\sigma_t^\eta)^2 \eta^2} \qquad \text{and} \qquad \theta''(\eta) = \frac{2\left[(\rho - r)\theta_t - \theta'(\eta)\mu_t^\eta \eta\right]}{(\sigma_t^\eta)^2 \eta^2},$$

where $q_t = q(\eta_t)$, $\theta_t = \theta(\eta_t)$, $\psi_t = \psi(\eta_t)$, $\mu_t^\eta = r - \psi_t g(q_t) + (1-\psi_t)\underline{\delta} - \sigma_t^\eta(\sigma + \sigma_t^\theta) - \sigma\sigma^\theta$,

$$\mu_t^q = -\left(\frac{a - \iota_t}{q_t} + g(q_t) + \sigma\sigma_t^q - r + \tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q)\right),$$

and σ_t^η, σ_t^q and σ_t^θ are determined as follows

$$\sigma_t^\eta = \frac{\frac{\psi_t \tilde{\varphi} q_t}{\eta} - 1}{1 - \psi_t \tilde{\varphi} q'(\eta_t)}\sigma, \quad \sigma_t^q = \frac{q'(\eta_t)}{q_t}\sigma_t^\eta \eta_t, \quad \text{and} \quad \sigma_t^\theta = \frac{\theta'(\eta_t)}{\theta_t}\sigma_t^\eta \eta_t.$$

Also, $\psi_t = 1$ if

$$g(q_t) + \underline{\delta} - \frac{\iota(q_t)}{q_t} + \tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q) < 0,$$

and, otherwise, ψ_t is determined by the equation

$$g(q_t) + \underline{\delta} - \frac{\iota(q_t)}{q_t} + \tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q) = 0.$$

Function $q(\eta)$ is increasing, $\theta(\eta)$ is decreasing, and the boundary conditions are

$$q(0) = \underline{q}, \ \theta(\eta^*) = 1, \ q'(\eta^*) = 0, \ \theta'(\eta^*) = 0 \ \text{and} \ \lim_{\eta \to 0} \theta(\eta) = \infty.$$

Proof. First, we derive expressions for the volatilities of η_t, q_t and θ_t. Using the law of motion of η_t from Lemma 3 and Ito's lemma, the volatility of q_t is given by

$$\sigma_t^q q_t = q'(\eta)(\psi_t \tilde{\varphi}(\sigma + \sigma_t^q)q_t - \sigma\eta_t) \Rightarrow \sigma_t^q q_t = \frac{q'(\eta_t)(\psi_t \tilde{\varphi} q_t - \eta_t)}{1 - \psi_t \tilde{\varphi} q'(\eta_t)}\sigma$$

The expressions for σ_t^η and σ_t^θ follow immediately from Ito's lemma.

Second, note that from (EK) and (H), it follows that

$$g(q_t) + \underline{\delta} - \frac{\iota(q_t)}{q_t} + \tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q) \le 0$$

with equality if $\psi_t < 1$, which justifies our procedure for determining ψ_t.

The expression for μ_t^q follows directly from (EK). The differential equation for $q''(\eta)$ follows from the law of motion of η_t and Ito's lemma: the drift of q_t is given by

$$\mu_t^q q_t = q'(\eta_t)\mu_t^\eta \eta_t + \frac{1}{2}(\sigma_t^\eta)^2 \eta_t^2 q''(\eta_t).$$

Similarly, $\mu_t^\theta = \rho - r$ and Ito's lemma imply that

$$\theta'(\eta_t)\mu_t^\eta \eta_t + \frac{1}{2}(\sigma_t^\eta)^2 \eta_t^2 \theta''(\eta) = (\rho - r)\theta(\eta_t).$$

Finally, let us justify the five boundary conditions. First, because in the event that η_t drops to 0 experts are pushed to the solvency constraint and must liquidate any capital holdings to households, we have $q(0) = \underline{q}$. In this case, households have to hold capital until it is fully depreciated and hence their willingness to pay is simply $\underline{q} = \underline{a}/(r + \underline{\delta})$. Second, because η^* is defined as the point where experts consume, expert optimization implies that $\theta(\eta^*) = 1$ (see Proposition 1). Third and fourth, $q'(\eta^*) = 0$ and $\theta'(\eta^*) = 0$ are the standard boundary conditions at a reflecting boundary. If one of these conditions were violated, e.g. if $q'(\eta^*) < 0$, then any expert holding capital when $\eta_t = \eta^*$ would suffer losses at an infinite expected rate.[17] Likewise, if $\theta'(\eta^*) < 0$, then the drift of $\theta(\eta_t)$ would be infinite at the moment when $\eta_t = \eta^*$, contradicting Proposition 1. Fifth, if η_t ever reaches 0, it becomes absorbed there. If any expert had an infinitesimal amount of capital at that point, he would face a permanent price of capital of $\underline{q}$. At this price, he is able to generate the return on capital of

$$\frac{a - \iota(\underline{q})}{\underline{q}} + g(\underline{q}) > r$$

without leverage, and arbitrarily high return with leverage. In particular, with high enough leverage this expert can generate a return that exceeds his rate of time preference ρ, and since he is risk-neutral, he can attain infinite utility. It follows that $\theta(0) = \infty$.

Note that we have five boundary conditions required to solve a system of two second-order ordinary differential equations with an unknown boundary η^*. ■

Numerical Example. Proposition 2 allows us to compute equilibria numerically, and to derive analytical results about equilibrium behavior and asset prices. To compute the example in Figure 2, we took parameter values $r = 5\%$, $\rho = 6\%$, $\underline{\delta} = 5\%$, $a = \underline{a} = 1$, $\sigma = 0.35$, $\tilde{\varphi} = 1$, and assumed that the production sets of experts are degenerate, so $g(q) = 4\%$ (so that $\delta = -4\%$) and $\iota(q) = 0$ for all q. Under these assumptions, capital, when permanently managed by less productive households, has an NPV of $\underline{q} = 10$.

As η_t increases, capital becomes more expensive (i.e. $q(\eta_t)$ goes up), and $\theta(\eta_t)$, experts' marginal value per dollar of net worth, declines. Denote by η^ψ the point that divides the state space of $[0, \eta^*]$ into the region where less productive households hold some capital directly, and the region where all capital is held by experts. In other words, when $\eta_t < \eta^\psi$, capital is so cheap that less productive households find it profitable to start speculating for capital gains, i.e. $\psi_t < 1$. Experts hold all capital in the economy when $\eta_t \in [\eta^\psi, \eta^*]$.

[17] To see intuition behind this result, if $\eta_t = \eta^*$ then $\eta_{t+\epsilon}$ is approximately distributed as $\eta^* - \bar{\omega}$, where $\bar{\omega}$ is the absolute value of a normal random variable with mean 0 and variance $(\sigma_t^\eta)^2 \epsilon$ As a result, $\eta_{t+\epsilon} \sim \eta^* - \sigma_t^\eta \sqrt{\epsilon}$, so $q(\eta^*) - q'(\eta^*)\sigma_t^\eta \sqrt{\epsilon}$. Thus, the loss per unit of time ϵ is $q'(\eta^*)\sigma_t^\eta \sqrt{\epsilon}$, and the average rate of loss is $q'(\eta^*)\sigma_t^\eta / \sqrt{\epsilon} \to \infty$ as $\epsilon \to 0$.

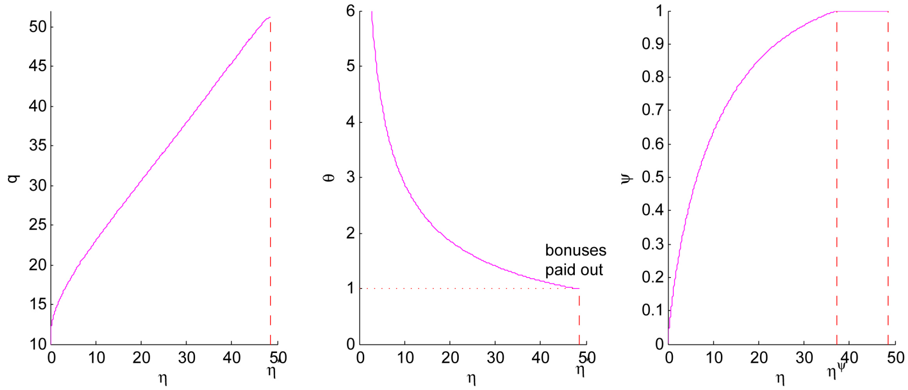

Figure 2: The price of capital, the marginal component of experts' value function and the fraction of capital managed by experts, as functions of η

In equilibrium, the state variable η_t, which determines the price of capital, fluctuates due to aggregate shocks dZ_t that affect the value of capital held by experts. To get a better sense of equilibrium dynamics, Figure 3 shows the *drift* and *volatility* of η_t for our computed example. The drift of η_t is positive on the entire interval $[0, \eta^*)$, because experts refrain from consumption and get an expected return of at least r. The magnitude of the drift is determined by the amount of capital they hold, i.e. ψ_t, and the expected return they get from investing in capital (which is related to whether capital is cheap or expensive). In expectation, η_t gravitates towards η^*, where it hits a reflecting boundary as experts consume excess net worth.

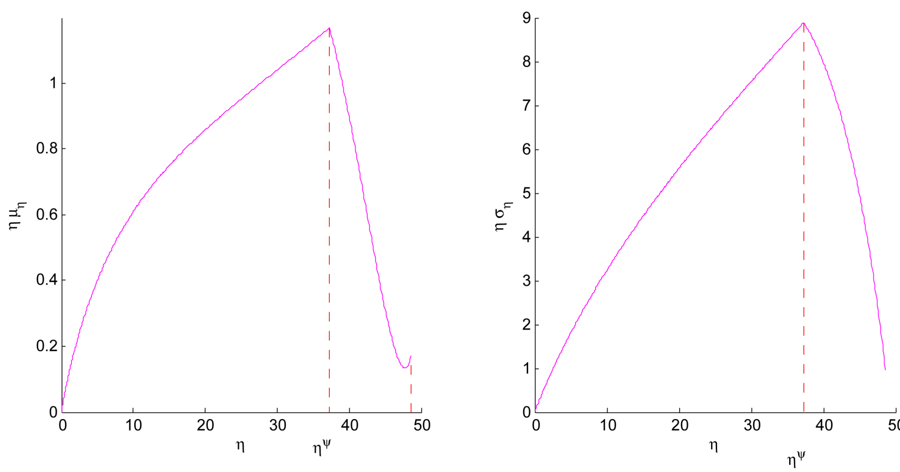

Figure 3: The drift $\eta\mu_\eta$ and volatility $\eta\sigma_\eta$ of η_t process.

Thus, point η^* is the *stochastic steady state* of our system. We draw an analogy between point η^* is our model and the *steady state* in traditional macro models, such as BGG and KM. Just like the steady state in BGG and KM, η^* is the point of global attraction of the system and, as we see from Figure 3 and as we discuss below, the volatility near η^* is low. However, unlike in traditional macro models, we do not consider the limit as noise η goes to 0 to identify the steady state, but rather look for the point where the system remains still in the absence of shocks when the agents take future volatility into account. Strictly speaking in our model, in the *deterministic steady* state where η_t ends up as $\sigma \to 0$: experts do not require any net worth to manage capital as financial frictions go away. Rather than studying how our economy responds to small shocks in the neighborhood of a stable steady state, we want to identify a region where the system stays relatively stable in response to small shocks, and see if large shocks can cause drastic changes in system dynamics. In fact, they will, and variations in system behavior are explained by *endogenous risk*.

4 Instability, Endogenous Risk, and Asset Pricing

Having solved for the full dynamics, we can address various economic questions like (i) How important is fundamental cash flow risk relative to endogenous risk created by the system? (ii) Does the economy react to large exogenous shocks differently compared to small shocks? (iii) Is the dynamical system unstable and hence the economy is subject to systemic risk? (iv) How does this affect prices of physical capital, equity and derivatives?

4.1 Amplification due to Endogenous Risk

Endogenous risk refers to changes in asset prices that are caused not by shocks to fundamentals, but rather by adjustments that institutions make in response to shocks, which may be driven by constraints or simply the precautionary motive. While exogenous fundamental shocks cause initial losses that make institutions constrained, endogenous risk is created through feedback loops that arise when experts react to initial losses. In our model, exogenous risk, σ, is assumed to be constant, whereas endogenous risk σ_t^q varies with the state of the system. Total instantaneous volatility is the sum of exogenous and endogenous risk, $\sigma + \sigma_t^q$. Total risk is also systematic in our baseline setting, since it is not diversifiable.

The amplification of shocks that creates endogenous risk depends on (i) expert leverage and (ii) feedback loops that arise as prices react to changes in expert net worth, and affect expert net worth further. Note that experts' debt is financed in short-term, while their assets are subject to aggregate market illiquidity.[18] Figure 4 illustrates the feedback mechanism of amplification, which has been identified by both BGG and KM near the steady state of their models.

Proposition 2 provides formulas that capture how leverage and feedback loops contribute to endogenous risk,

$$\sigma_t^\eta = \frac{\frac{\psi_t \tilde{\varphi} q_t}{\eta} - 1}{1 - \psi_t \tilde{\varphi} q'(\eta_t)} \sigma \quad \text{and} \quad \sigma_t^q = \frac{q'(\eta_t)}{q_t} \sigma_t^\eta \eta_t.$$

The *numerator* of σ_t^η, $\psi_t \tilde{\varphi} q_t / \eta_t - 1$, is the experts' debt to equity ratio. Without taking into account the reaction of prices to experts' net worths, this ratio captures

[18] Recall that the price impact of a single expert is zero in our setting. However, the price impact due to aggregate shocks can be large. Hence, a "liquidity mismatch index" that tries captures the mismatch between market liquidity of experts' asset and funding liquidity on the liability side has to focus on price impact of assets caused by aggregate shocks rather than idiosyncratic shocks.

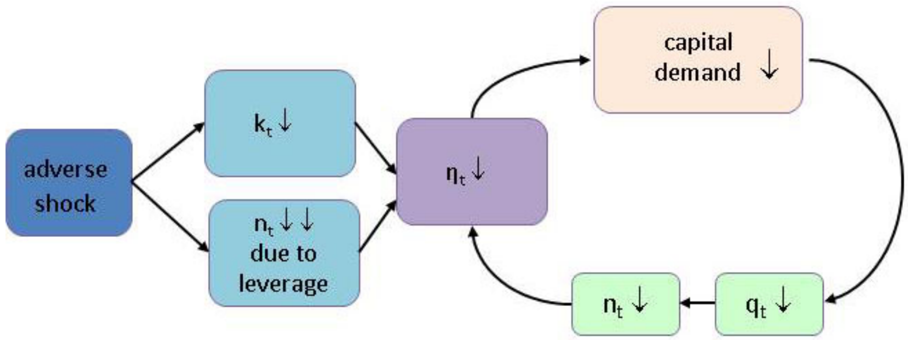

Figure 4: Adverse Feedback Loop.

the effect of an exogenous aggregate shock on η_t. An exogenous shock of dZ_t changes K_t by $dK_t = \sigma K_t\, dZ_t$, and has an immediate effect on the net worth of experts of the size $dN_t = \psi_t \tilde{\varphi} q_t \sigma K_t\, dZ_t$. The immediate effect is that the ratio η_t of net worth to total capital changes by $(\psi_t \tilde{\varphi} q_t - \eta_t)\, dZ_t$, since

$$d\left(\frac{N_t}{K_t}\right) = \frac{dN_t K_t - N_t dK_t}{(K_t)^2} = \sigma(\psi_t \tilde{\varphi} q_t - \eta_t)\, dZ_t.$$

The *denominator* of σ_t^η captures feedback effects through prices. When $q'(\eta) = 0$, even though a shock to experts' net worth's is magnified through leverage, it does not affect prices. However, when $q'(\eta) > 0$, then a drop in η_t by $\sigma(\psi_t \tilde{\varphi} q_t - \eta_t)\, dZ_t$, causes the price q_t to drop by $q'(\eta_t)\sigma(\psi_t \tilde{\varphi} q_t - \eta_t)\, dZ_t$, leading to further deterioration of the net worth of experts, which feeds back into prices, and so on. The amplification effect is nonlinear, which is captured by $1 - \psi_t \tilde{\varphi} q'(\eta_t)$ in the *denominator* of σ_t^η (and if $q'(\eta)$ were even greater than $1/(\psi_t \tilde{\varphi})$, then the feedback effect would be completely unstable, leading to infinite volatility). Note that the amplification does not arise if agents could directly contract on k_t instead of only at $k_t q_t$. Appendix Bshows that the denominator simplifies to one in this case.

Normal versus crisis times. The equilibrium in our model has no endogenous risk near the stochastic steady state, and significant endogenous risk below the steady state. This result strongly resonates what we observe in practice during normal times and crisis episodes.

Theorem 1 *For $\eta_t < \eta^*$, shocks to experts' net worth's spill over into prices and indirect dynamic amplification is given by $1/\left[1 - \psi_t \tilde{\varphi} q'(\eta_t)\right]$, while at $\eta = \eta^*$, there is no amplification since $q'(\eta^*) = 0$.*

Proof. This result follows directly from Proposition 2. ■

The reason amplification is so different in normal times and after unusual losses has to do with endogenous risk-taking. When intermediaries choose leverage, or equity buffer against the risk of their assets, they take into account the trade-off between the threat that they become constrained and the opportunity cost of funds. As a result, at target leverage intermediaries are relatively unconstrained and can easily absorb small losses. However, after large shocks, the imperative to adjust balance sheets becomes much greater, and feedback effects due to reactions to new shocks create volatility endogenously.

In our setting, endogenous leverage corresponds to the choice of the payout point η^*. Near η^*, experts are relatively unconstrained: because shocks to experts' net worth's can be easily absorbed through adjustments to payouts, they have little effect on the experts' demand for capital or on prices. In contrast, below η^* experts become constrained, and so shocks to their net worth's immediately feed into their demand for assets.

"Ergodic Instability." Due to the non-linear dynamics, the system is inherently unstable. As a consequence agents are exposed to systemic risk. As the experts' net worth falls below η^*, total price volatility $\sigma + \sigma^q$ rises sharply. The left panel of Figure 5 shows the total (systematic) volatility of the value of capital, $\sigma + \sigma_t^q$, for our computed example.

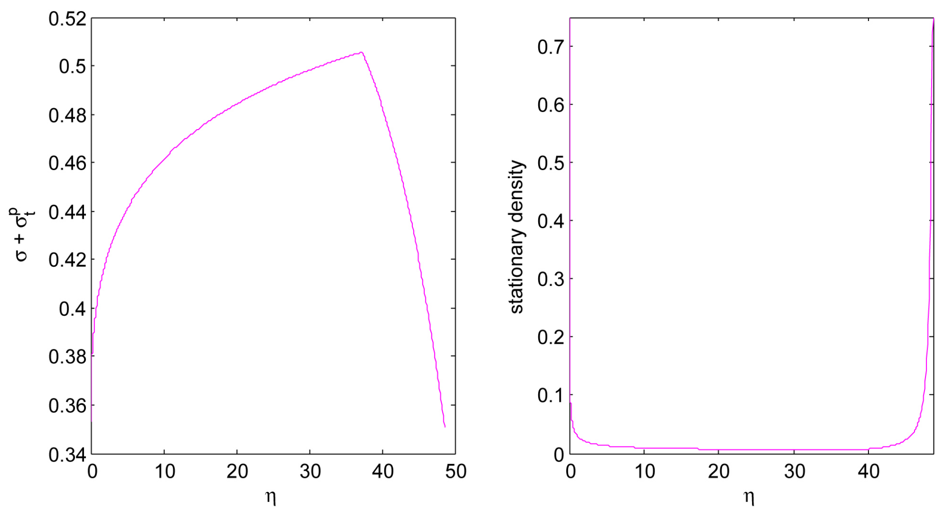

Figure 5: Systematic and systemic risk: Volatility of the value of capital and the stationary distribution of η_t.

The right panel of Figure 5 shows the stationary distribution of η_t. Starting from any point $\eta_0 \in (0, \eta^*)$ in the state space, the density of the state variable η_t converges to the stationary distribution in the long run as $t \to \infty$. Stationary density also measures the average amount of time that the variable η_t spends in the long run near each point. Proposition C1 in Appendix C provides equations that characterize this stationary distribution directly derived from $\mu_\eta(\eta)$ and $\sigma_\eta(\eta)$ depicted in Figure 3.

The key feature of the stationary distribution is that it is bimodal with high densities at the extremes. We refer to this characteristic as "ergodic instability". The system exhibits large swings, but it is still ergodic ensuring that a stationary distribution exists. More specifically, the stationary density is high near η^*, which is the attracting point of the system, but very thin in the middle region below η^* where the volatility is high. The system moves fast through regions of high volatility, and so the time spent there is very short. These excursions below the steady state are characterized by high uncertainty, and occasionally may take the system very far below the steady state. In other words, the economy is subject to break-downs – i.e. systemic risk. At the extreme low end of the state space, assets are essentially valued by unproductive households, with $q_t \sim \underline{q}$, and so the volatility is low. The system spends most of the time around the extreme points: either experts are well capitalized and financial system can deal well with small adverse shocks or it drops off quite rapidly to very low η-values, where prices and experts' net worth drop dramatically. As the economy occasionally implodes, it exhibits systemic risk, because the net worth of the highly levered expert sector is inappropriately low reflects systemic risk in our model. The (undiversifiable) systematic risk $\sigma + \sigma^q$ is also high for $\eta < \eta^*$.

Full Equilibrium Dynamics vs. Linear Approximations. Macroeconomic models with financial frictions such as BGG and KM do not fully characterize the whole dynamical system but focus on the log-linearization around the deterministic steady state. The implications of our framework differ in at least three important dimensions:

First, linear approximation near the stochastic steady state predicts a normal stationary distribution around it, suggesting a much more stable system. The fact that the stationary distribution is bimodal, as depicted on the right panel of Figure 5, suggests a more powerful amplification mechanism away from the steady state. Papers such as BGG and KM do not capture the distinction between relatively stable dynamics near the steady state, and much stronger amplification loops below the steady state. Our analysis highlights the sharp distinction between crisis and normal times, which has important implication when calibrating a macro-model.

Second, while log-linearized solutions can capture amplification effects of various magnitudes by placing the steady state in a particular part of the state space, these experiments may be misleading as they force the system to behave in a completely different way. Steady state can me "moved" by a choice of an *exogenous* parameter such as exogenous drainage of expert net worth in BGG. With *endogenous* payouts and a setting in which agents anticipate adverse shocks, the steady state naturally falls in the relatively unconstrained region where amplification is low, and amplification below the steady state is high.

Third, the traditional approach determines the steady state by focusing on the limiting case in which the aggregate exogenous risk σ goes to zero. A single unanticipated (zero probability) shock upsets the system that subsequently slowly drifts back to the steady state. As mentioned above, setting the exogenous risk σ to zero also alters experts behavior. In particular, they would not accumulate any net worth and the steady state would be deterministic at $\eta^* \to 0$.

4.2 Asset Pricing

Volatility and the Precautionary Motive. Endogenous risk leads to *excess volatility*, as the value of capital is not only affected by cash flow shocks σ but also by changes in the stochastic discount factor reflected leading to endogenous risk σ^q. Excess volatility increases the experts' *precautionary motive*, leading to a higher required expected return on capital. This can be seen directly from equation (EK) in Proposition 1,

$$\underbrace{\frac{a - \iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q - r}_{\text{expected excess return on capital, } E[r_t^k]-r} = \underbrace{\tilde{\varphi}(-\sigma_t^\theta)(\sigma + \sigma_t^q)}_{\text{risk premium}}. \tag{EK}$$

Endogenous risk increases the experts' incentives to hoard cash (note that $(-\sigma_t^\theta) > 0$), because cash has a greater option value when a larger fraction of price movements is explained by reasons other than changes in fundamentals.

Of course, the profit that experts can make following a price drop depends on their value functions, which are forward-looking - anticipating all future investment opportunities. According to (EK), the experts' equilibrium expected return from capital has to depend on the covariance between the experts' marginal values of net worth's θ_t and the value of capital. Capital prices have to drop in anticipation of volatile episodes, so that higher expected return balances out the experts' precautionary motive. This is our first empirical prediction.

Viewed through the stochastic discount factor (SDF) lens, Equation (EK) shows that expected return on capital is simply given by the covariance between the value of capital and the experts' stochastic discount factor. As discussed in Section 3, at time t experts value future cash flow at time $t+s$ with the SDF $e^{-\rho s}\theta_{t+s}/\theta_t$, so that an asset producing cash flow x_{t+s} at time $t+s$ has price

$$E_t\left[e^{-\rho s}\frac{\theta_{t+s}}{\theta_t}x_{t+s}\right]$$

at time t. Note that less productive households' SDF is simply e^{-rs} since they are not financially constrained. Of course, they only price capital for $\psi < 1$ and their payoff from holding the same physical capital is lower.

In models with risk averse agents, the precautionary motive is often linked to a positive "prudence coefficient" which is given by the third derivative of their utility function normalized by the second derivative. In our setting the third derivative of experts' value function (second derivative of $\theta(\eta)$) plays a similar role. It is positive, since the marginal value function, θ, is convex (see Figure 2). In short, even though experts are risk-neutral, financial frictions and the fact that $dc_t \geq 0$ make experts behave in a risk-averse and prudent manner – a feature that our setting shares with buffer stock models.

Asset Prices in Cross-Section. Excess volatility due to endogenous risk spills over across all assets held by constrained agents, making asset prices in cross-section significantly more correlated in crisis times. Erb, Harvey, and Viskanta (1994) document this increase in correlation within an international context. This phenomenon is important in practice as many risk models have failed to take this correlation effects into account in the recent housing price crash.[19]

To demonstrate this result, we have to extend the model to allow for multiple types of capital. Each type of capital k^l is hit by aggregate and type-specific shocks. Specifically, capital of type l evolves according to

$$dk_t^l = gk_t^l\,dt + \sigma k_t^l\,dZ_t + \sigma' k_t^l\,dZ_t^l,$$

where dZ_t^l is a type-specific Brownian shock uncorrelated with the aggregate shock dZ_t.

In aggregate, idiosyncratic shocks cancel out and the total amount of capital in the economy still evolves according to

$$dK_t = gK_t\,dt + \sigma K_t\,dZ_t.$$

Then, in equilibrium financial intermediaries hold fully diversified portfolios and experience only aggregate shocks. The equilibrium looks identical to one in the single-asset model, with price of capital of any kind given by q_t per unit of capital. Then

$$d(q_t k_t^l) = (\Phi(\iota_t^j) - \delta + \mu_t^q + \sigma\sigma_t^q)(k_t^l q_t)\,dt + (q_t k_t^l)(\sigma + \sigma_t^q)\,dZ_t + (q_t k_t^l)\sigma'\,dZ_t^l.$$

The correlation between assets l and l' is

$$\frac{Cov[q_t k_t^l, q_t k_t^l]}{\sqrt{Var[q_t k_t^l]Var[q_t k_t^l]}} = \frac{(\sigma + \sigma_t^q)^2}{(\sigma + \sigma_t^q)^2 + (\sigma')^2}\;.$$

Near the steady state $\eta_t = \eta^*$, there is only as much correlation between the prices of assets l and l' as there is correlation between shocks. Specifically, $\sigma_t^q = 0$ near the steady state, and so the correlation is

$$\frac{\sigma^2}{\sigma^2 + (\sigma')^2}\,.$$

Away from η^*, correlation increases as σ_t^q increases. Asset prices become most correlated in prices when σ_t^q is the largest. As $\sigma_t^q \to \infty$, the correlation tends to 1.

Of course, in practice financial institutions specialize and do not hold fully diversified portfolios. One could capture this in a model in which experts differ by specialization, with each type of expert having special skills to manage some types of capital but not others. In this case, feedback effects from shocks to one particular type of capital would depend on (i) who holds the largest quantities of this type of capital (ii) how constrained they are and (iii) who holds similar portfolios. Thus, we hypothesize that in general spillover effects depend on the network structure of financial institutions, and that shocks propagate through the strongest links and get amplified in the weakest nodes.

[19]See “Efficiency and Beyond” in The Economist, July 16, 2009.

Outside equity. Our results on excess volatility carry over to outside equity. Returns on outside equity are also negatively skewed as a negative fundamental macro shock is amplified in times of crisis. If experts cannot perfectly diversify across all forms of capital, experts outside equity is also more correlated in crisis times. However, expected returns of outside equity is not time-varying as they are priced by risk-neutral and financially unconstrained households whose stochastic discount factor is e^{-rt}. The discounted outside equity price processes follow a martingale. If households were assumed to be risk averse, these implied risk characteristics of outside equity would lead to predictability in returns in outside equity as well.

Derivatives. Since data for crisis periods are limited, it is worthwhile to look at option prices that reflect market participants' implicit probability weights of extreme events. Our result that price volatility is higher for lower η_t-values also has strong implications for option prices.

First, it provides an explanation for "volatility smirks" of options in normal times, see e.g. Bates (2000). Since the values of options monotonically increase with the volatility of the underlying stock, option prices can be used to compute the "implied volatility" from the Black-Scholes option pricing formula. One example of a "volatility smirk" is that empirically put options have a higher implied volatility when they are further out of the money. That is, the larger the price drop has to be for an option to ultimately pay off, the higher is the implied volatility or, put differently, far out of the money options are overpriced relative to at the money options. Our model naturally delivers this result as volatility in times of crises is higher.

Second, so called "dispersion trades" try to exploit the empirical pattern that the smirk effect is more pronounced for index options than for options written on individual stocks (Driessen, Maenhout, and Vilkov (2009)). Note that index options are primarily driven by macro shocks, while individual stock options are also affected by idiosyncratic shocks. The observed option price patterns arise quite naturally in our setting as the correlation across stock prices increases in crisis times. Note that in our setting options are redundant assets as their payoffs can be replicated by the underlying asset and the bond, since the volatility is a smooth function in q_t. This is in contrast to stochastic volatility models in which volatility is independently drawn and subject to a further stochastic factor for which no hedging instrument exists.

5 Volatility Paradox

Given that the economy is prone to self-generated systemic risk it is natural to ask whether a reduction in fundamental exogenous volatility would stabilize the system. In the second part of this section, we address the question whether new financial products, like derivatives, that allow experts to (better) hedge idiosyncratic risk lead to an overall reduction in risk.

5.1 Reduction in Exogenous Risk

A reduction in exogenous cash flow risk σ reduces financial frictions. Paradoxically, it can make the economy less stable. That is, it can increase the maximum volatility of experts' net worth. The reason is that a decline in cash flow volatility encourages experts to increase their leverage by reducing their net worth buffer.

Figure 6 reverts to our previous numerical example and illustrates the price of capital q_t, the volatility of the state variable σ_t^η as well as expert leverage $\psi_t q_t/(\tilde{\varphi}\eta_t)$ for three different exogenous risk values σ =.025 (blue), .05 (red), and .1 (black) (recall that $r = 5\%$, $\rho = 6\%$, $\underline{\delta} = 5\%$, $a = 1$, $\tilde{\varphi} = 1$, and the production sets of experts are degenerate, so $g(q) = 4\%$ and $\iota(q) = 0$ for all q).

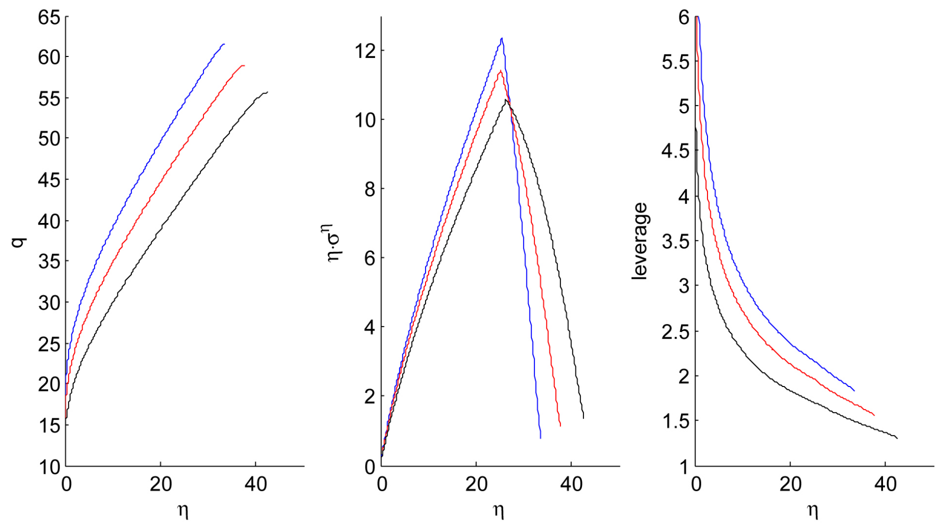

Figure 6: Equilibrium for three different levels of exogenous risk $\sigma = .025$ (blue), .05 (red), .1 (black).

As one would expect, as aggregate exogenous risk goes down, volatility near the global attractor η^* is typically declining. However, in equilibrium experts respond to lower exogenous risk by taking on higher leverage and paying out bonuses sooner (i.e. η^* is lower). Overall, this results in higher systemic risk reflected by greater amplification below steady state. This "volatility paradox" is consistent with the fact that the current crisis was preceded by a low volatility environment, referred to as the "great moderation." In other words, in the absence of financial regulation, the system is prone to instabilities even and especially when the level of aggregate risk is low.

5.2 Financial Innovation: Hedging of Idiosyncratic Jump Risk

Next, we explore the impact of financial innovations. New financial products allow experts to better share risk, and hedge idiosyncratic risks in particular. These products can also involve securitization, including pooling and trenching, credit default swaps, and various options and futures contracts. We find that financial innovation reduces idiosyncratic risk but it also emboldens experts to live with smaller net worth buffers and higher leverage, increasing systemic risk. Ironically tools intended for more efficient risk management can lead to amplification of systemic risks, making the system less stable.

To study this question we enrich our baseline setting by introducing idiosyncratic jump risk. Introducing jump risk has two further advantages. First, now debt may default and hence we can study credit spreads, the interest rate spreads between risky loans and the risk-free rate. Second, we can draw a direct comparison with the model in BGG.

Formally, we assume that capital k_t managed by expert i evolves according to

$$dk_t = gk_t\,dt + \sigma k_t\,dZ_t + k_t\,dJ_t^i,$$

where dJ_t^i is an idiosyncratic Poisson loss process. As BGG we make the simplifying assumption that when experts get bigger, their idiosyncratic shocks are amplified proportionately, that is, there is no diversification of idiosyncratic shocks within any expert.

Losses after an idiosyncratic jump are characterized by the distribution function $F : [0,1] \to [0,1]$, which describes the percentage of capital that is *recovered* in the event of a loss. We denote the individual intensity of Poisson loss shocks by λ.

We assume that experts' balance sheets are the same as in our baseline model: experts are required to hold a fraction of $\tilde{\varphi}$ of the equity. As BGG, we adopt the costly state verification framework of Townsend (1979) to deal with the possibility that the value of the expert's total assets $q_t k_t$ drops below the value of debt, i.e. equity $n_t/\tilde{\varphi}$ becomes negative. Specifically, we assume that in such an event bankruptcy is triggered and debt holders must incur a verification cost and recover only a fraction $1-c$ of remaining capital. That is, as in BGG, we assume that the verification cost is a constant fraction $c \in (0,1)$ of the remaining capital.[20]

Default and costly state verification occur when the value of the assets $k_t q_t$ falls below the value of debt $k_t q_t - n_t/\tilde{\varphi}$ i.e. a fraction of capital less than $\vartheta_t = 1 - n_t/(\tilde{\varphi} q_t k_t)$ remains after a jump. If x is the fraction of assets recovered in the event of default, then debt holders lose

$$k_t q_t(\vartheta_t - x) + c(k_t q_t)x.$$

Denote

$$L(\vartheta) = \lambda \int_0^{\vartheta} (\vartheta - x)\,dF(x) \qquad \text{and} \qquad C(\vartheta) = \lambda \int_0^{\vartheta} cx\,dF(x).$$

[20]The basic costly state verification framework, developed by Townsend (1979) and adopted by Bernanke, Gertler, and Gilchrist (1999) is a two-period contracting framework. At date 0, the agent requires investment i from the principal, and at date 1 he receives random output y distributed on the interval The agent privately observes output y, but the principal can verify it at a cost. The optimal contract under commitment is a standard debt contract. If the agent receives $y \leq D$, the face value of debt, then he pays the principal D and there is no verification. If $y < D$, the agent cannot pay D and costly state verification (bankruptcy) is triggered, and debt holders receive all of output.

Then the expected loss rate to debt holders due to default is given by $(k_t q_t)(L(\vartheta_t) + C(\vartheta_t))$. To receive compensation for this loss, debt holders require a credit spread of

$$\frac{k_t q_t}{k_t q_t - n_t/\tilde{\varphi}} \ (L(\vartheta_t) + C(\vartheta_t)) = \frac{L(\vartheta_t) + C(\vartheta_t)}{\vartheta_t}$$

We formally justify this form of modeling default via a costly state verification framework at the end of Section 7.

As before, the equilibrium is characterized by the state variable η_t, and prices $q_t = q(\eta_t)$ and the expert's (marginal) value function $\theta_t = \theta(\eta_t)$ are functions of η_t. The net worth of an individual expert evolves according to

$$dn_t = rn_t\, dt + (k_t q_t)[(E_t[r_t^k] - r - L(\vartheta_t) - C(\vartheta_t))\, dt + dJ_t^i + \tilde{\varphi}(\sigma + \sigma_t^q)\, dZ_t] - dc_t$$

Recall that in Section 3 we defined $E_t[r_t^k] = [a - \iota(q_t)]/q_t + g(q_t) + \mu_t^q + \sigma\sigma_t^q$ as experts' expected return on unleveraged capital.

In the aggregate, experts that survive have to pay high interest rate spread partially due to $L(\vartheta_t)$, while other experts that go bankrupt experience a positive transfer since due to limited liability they do not have to cover their full losses. Hence, aggregate expert capital evolves according to

$$dN_t = rN_t\, dt + \psi_t K_t q_t[(E_t[r_t^k] - r - C(\vartheta_t))\, dt + \tilde{\varphi}(\sigma + \sigma_t^q)\, dZ_t] - dC_t,$$

where the term $L(\vartheta_t)$ drops out. The modified law of motion of $\eta_t = N_t/K_t$ is

$$d\eta_t = \mu_t^\eta \eta_t\, dt + \sigma_t^\eta \eta_t\, dZ_t - d\zeta_t$$

where

$$\mu_t^\eta = r - \psi_t g(q_t) + (1 - \psi_t)\underline{\delta} + \frac{\psi_t q_t}{\eta_t}\left(E_t[r_t^k] - r - C(\vartheta_t)\right) - \sigma\sigma_t^\eta$$

$$\sigma_t^\eta = \frac{\psi_t \tilde{\varphi} q_t}{\eta_t}(\sigma + \sigma_t^q) - \sigma, \quad \text{and} \quad d\zeta_t = \frac{dC_t}{K_t}.$$

The Bellman equation and the first-order condition with respect to k_t are now

$$\rho\theta_t n_t = \max_{k_t} \left[rn_t + (k_t q_t)(E_t[r_t^k] - r - C(\vartheta_t))\right]\theta_t + n_t \mu_t^\theta \theta_t + \sigma_t^\theta \theta_t (k_t q_t)\tilde{\varphi}(\sigma + \sigma_t^q).$$

Replacing total assets over net worth $\frac{k_t q_t}{n_t}$ with $\frac{1}{(1-\vartheta_t)\tilde{\varphi}}$ and dividing by $n_t\theta_t$, the Bellman equation can be written as

$$\rho - r = \mu_t^\theta + \max_{\vartheta} \frac{1}{\tilde{\varphi}(1-\vartheta)}\left(E_t[r_t^k] - r - C(\vartheta_t) + \tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q)\right)$$

The first-order condition with respect to ϑ is now

$$E_t[r_t^k] - r - C(\vartheta) + \tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q) + (1 - \vartheta)C'(\vartheta) = 0$$

As before, in equilibrium η_t evolves on the range $[0, \eta^*]$, with a different boundary η^*. Experts pay themselves bonuses only when η_t is at η^*.

Comparison with BGG and KM (1997). As in BGG the credit spread in our extended setting is due to idiosyncratic default risk and is higher when experts are less well capitalized, i.e. for lower η. Our model closely resembles the discrete time steady state model in BGG if we set $\varphi_t = 1$. The key distinction is however that experts are not exogenously forced to consume when they lack net worth in our setting since they endogenously decide when to pay out bonuses and are not forced to exit after facing an exogenous "exit shock". This difference is important when calibrating the model, since it implies that in our setting amplification is large in crises times, while it is muted in normal times.

By varying the verification costs and the loss distribution, our extended setting encompasses several other models in the literature. For example, the assumptions of KM imply that financial experts can borrow only up to a fraction ζ_t of the market value of assets. Thus, someone with net worth n_t can hold at most $n_t/(1-\zeta_t)$ worth of assets, by financing $n_t\zeta_t/(1-\zeta_t)$ of the assets with debt and the rest, n_t, with personal wealth. This is captured in our framework by setting the verification costs to zero up to a certain level and infinity afterwards. Alternatively, one can assume that margins are set equal to the value-at-risk (VaR) as in Brunnermeier and Pedersen (2009) and Shin (2010). In the former margins increase with endogenous price volatility. These effects can be captured in our model by letting the intensity of Poisson jumps $\lambda(\sigma_t^q)$ depend on price volatility.

Introducing Securitization. Securitization and new financial innovation allows risk-sharing within the expert sector in our model. Specifically, assume that securitization makes all shocks, both idiosyncratic J_t^i and aggregate Z_t, observable and contractible among the experts, but not between experts and households.

Denote by ω_t the risk premium on aggregate risk and by ω_t^i the risk premium on idiosyncratic risk. A hedging contract for aggregate risk adds

$$\varsigma_t(\omega_t\, dt + \, dZ_t)$$

to the law of motion of expert i's wealth, where ς_t is the overall risk exposure. A contract on idiosyncratic risk of expert i, adds

$$\varsigma_t^i(\omega_t^i\, dt + dJ_t^i)$$

to the law of motion of expert i's wealth, and may affect the verification region and verification costs. The following proposition characterizes the equilibrium when hedging within the financial sector is possible.

Proposition 3 *If hedging within the financial sector is possible, then in equilibrium experts will fully hedge idiosyncratic risk, which carries the risk premium of $\omega_t^i = 0$. Nobody hedges aggregate risk, which carries the risk premium of $\omega_t = -\sigma_t^\theta \geq 0$.Since idiosyncratic shocks are fully hedged, the equilibrium is identical to one in a setting without those shocks.*

Proof. It is easy to see that the idiosyncratic risks are fully hedged and that the risk premia are zero, since market clears when each expert optimally chooses to offload his own idiosyncratic risk, and take on a little bit of everybody's risks (which cancel out). Once idiosyncratic risks are removed, the law of motion of individual expert's capital is

$$dn_t = rn_t\,dt + (k_t q_t)[(E_t[r_t^k] - r)\,dt + \tilde{\varphi}(\sigma + \sigma_t^q)\,dZ_t] - dc_t - \varsigma_t(\omega_t\,dt + \,dZ_t)$$

where the optimal choice of ς_t must be zero in order for hedging markets to clear. The appropriate risk premium for aggregate risk can be found from the Bellman equation

$$\rho\theta_t n_t = \max_{k,\varsigma} \theta_t \left[rn_t + (kq_t)(E_t[r_t^k] - r) - \varsigma\omega_t\right] + \mu_t^\theta \theta_t n_t + \sigma_t^\theta \theta_t \left[(kq_t)\tilde{\varphi}(\sigma + \sigma_t^q) - \varsigma\right].$$

In order for $\varsigma_t = 0$ to be optimal, we need $\omega_t = -\sigma_t^\theta$. ■

Experts fully hedge out idiosyncratic shocks when securitization is allowed, they face the cost of borrowing of only r, instead of $r + C(\vartheta_t)/\vartheta_t$. Lower cost of borrowing leads to higher leverage and quicker payouts. As a result, the financial system becomes less stable. Thus, even though in principle securitization is a good thing, as it allows financial institutions to share idiosyncratic risks better and avoid bankruptcy costs, it can lead to greater leverage and the amplification of endogenous systemic risks.

6 Efficiency and Externalities

The fact that financial frictions lead to systemic instability and excess volatility does not necessarily prescribe strict financial regulation. Making the system more stable might stifle long-run economic growth. To study financial regulation one has to conduct a welfare analysis. This section makes a first small step in this direction. We start with the first-best efficient benchmark outcome that would emerge in an economy without frictions. In absence of any frictions the social planner's solution coincides with the market outcome. With frictions even a social planner faces constraints and hence his solution is only constrained efficient. The market outcome might not even be constrained efficient due to externalities. Individual market participants distort the outcome since they do not fully internalize the impact of their action on others.

6.1 First Best and Inefficiencies

First best without financial frictions. In the economy without friction, more productive experts should manage capital forever. They borrow from less productive households at the interest rate r. As a consequence, the price of capital would reach its maximal theoretical level of

$$\bar{q} = \max \frac{a - \iota}{r - \Phi(\iota) + \delta}$$

We discount the future cash flow $a - \iota$ by using the households' discount rate of r, since experts would consume their net worth in lump sum at time 0, borrow from less productive households and promise them all future cash flows. This is the maximal theoretically possible price of capital. Note that the least efficient allocation of capital is the one in which all the capital is held by less productive households. Hence, the NPV of future cash flow (and minimum price level) is $\underline{q} = \underline{a}\ (r + \underline{\delta})$.

Inefficiencies with financial frictions. With financial frictions there are inefficiencies and the allocation of capital between experts and less productive households depends on the level of η_t, i.e. how close experts are to their solvency constraint. The price of capital depends on how efficiently it is allocated, and on the speculative incentives of experts and households. With financial frictions the maximal price of capital, attained at $\eta_t = \eta^*$, is less than $\bar{q}$. At η^* the value of capital equals the sum of the NPV's of cash flows that go to experts and less productive households, valued at their respective discount rates. That is,

$$q(\eta^*)K_t = \underbrace{\theta(\eta^*)N_t}_{\text{expert NPV/wealth}} + \underbrace{q(\eta^*)K_t - N_t}_{\text{household wealth}} = \underbrace{E\left[\int_0^\infty e^{-\rho t} dC_t\right]}_{\text{expert payoff}} + \underbrace{E\left[\int_0^\infty e^{-rt} d\underline{C}_t\right]}_{\text{household payoff}} \quad (7)$$

since $\theta(\eta^*) = 1$. With financial frictions, because of three types of inefficiencies

(i) capital mis-allocation, since less productive households end up managing capital for low η_t,

(ii) under-investment, since $\iota(q_t) < \iota(\bar{q})$, and

(iii) consumption distortion, since experts postpone some of their consumption into the future.

In fact, as η_t goes down, all three forms of inefficiencies increase. When $\eta_t < \eta^*$, identity given by (7) no longer holds because $\theta(\eta_t) > 1$, i.e. experts get the expected payoff of more than one per unit of net worth. In general,

$$\begin{aligned} q_t K_t &= \underbrace{N_t}_{\text{expert wealth}} + \underbrace{(q_t K_t - N_t)}_{\text{household wealth}} \leq \underbrace{\theta_t N_t}_{\text{expert payoff}} + \underbrace{(q_t K_t - N_t)}_{\text{household payoff}} \\ &= \underbrace{E\left[\int_0^\infty e^{-\rho t} dC_t\right]}_{\text{expert payoff}} + \underbrace{E\left[\int_0^\infty e^{-rt} d\underline{C}_t\right]}_{\text{household payoff}}. \end{aligned}$$

Experts are consuming only when $\eta_t = \eta^*$, where a unit of their net worth has marginal value $\theta(\eta^*) = 1$.

6.2 Constrained Efficiency

While the first best efficient outcome is an interesting theoretical benchmark, it is unrealistic as it implicitly assumes that the social planner can overcome all financing frictions and agency problems. Therefore, in this subsection we explore what a social planner can achieve when bound by the same agency constraints as the market. To make this analysis formal, we first define the set of symmetric constrained feasible policies. Under these policies, the central planner treats all experts and households symmetrically, controls prices as well as the agents' consumption and investment choices, but must respect the financing constraint with respect to the experts' outside equity. We argue that constrained-feasible policies are incentive compatible given our microfoundation of balance sheets, which we present in Section 7. At the end of the subsection, we characterize the best outcome attainable by constrained feasible policies, and show that it coincides with the first best outcome.

Definition 2 *A symmetric constrained-feasible policy is described by a group of stochastic processes on the filtered probability space defined by the Brownian motion $\{Z_t, t \geq 0\}$: the price process $\{q_t\}$; expert investment rate $\{\iota_t\}$, fractions of equity retained $\{\varphi_t \geq \tilde{\varphi}\}$, aggregate expert capital holdings $\psi_t K_t$, consumption dC_t, and transfers $d\tau_t$; aggregate household capital holdings $(1-\psi_t)K_t$, consumption $d\underline{C}_t$, transfers $d\underline{\tau}_t$, and equity holdings $\underline{\xi}_t$, such that*

(i) representative expert net worth N_t stays nonnegative, where

$$dN_t = d\tau_t + (\psi_t K_t q_t)\left[\left(\frac{a-\iota_t}{q_t} + \Phi(\iota_t) - \delta + \mu_t^q + \sigma\sigma_t^q\right) dt + \varphi_t(\sigma + \sigma_t^q)\, dZ_t\right] - dC_t,$$

(ii) representative household net worth $\underline{N}_t$ is defined by

$$d\underline{N}_t = d\underline{\tau}_t + \underline{\xi}_t(\sigma + \sigma_t^q)dZ_t + (1-\psi_t)K_t q_t\left[\left(\frac{a}{q_t} - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q\right) dt + (\sigma + \sigma_t^q)\, dZ_t\right] - d\underline{C}_t$$

(iii) and the resource constraints are satisfied, i.e.

$$dC_t + d\underline{C}_t = (a - \iota_t)K_t\, dt,\ (1-\varphi_t)\psi_t K_t = \underline{\xi}_t, \text{ and } N_t + \underline{N}_t = q_t K_t,$$

where $dK_t/K_t = [\psi_t(\Phi(\iota_t) - \delta) - (1-\psi_t)\underline{\delta}]\, dt + \sigma\, dZ_t.$

Note that since the sum of net worth adds to the total wealth in the economy $q_t K_t$, aggregate transfers across both sectors are zero. Moreover, because of transfers we can set the risk-free rate to zero, without loss of generality. Under such a social planner policy, the net worth of an individual expert evolves according to

$$\begin{aligned} dn_t^i &= \frac{n_t^i}{N_t} d\tau_t + (1-\varphi_t)(q_t k_t^i)\left(\frac{a-\iota_t}{q_t} + \Phi(\iota_t) - \delta + \mu_t^q + \sigma\sigma_t^q\right)\, dt \\ &\quad + \varphi_t\left(\frac{a-\iota_t}{q_t}\, dt + d(q_t k_t^i)\right) - \frac{n_t^i}{N_t}\, d\underline{C}_t \end{aligned}$$

where $n_t^i/N_t\, d\tau_t$ is the expert's share of transfers, $n_t^i/N_t\, dC_t$ is his share of consumption and $k_t^i = \psi_t K_t(n_t^i/N_t)$ is the required holding of capital. The expert receives a fraction $1-\varphi_t$ of the expected return on capital (risk-free) and a fraction φ_t of the realized return on capital. There exists a process $\{\theta_t\}$ such that the experts' value functions are given by $n_t^i\theta_t$. It can be shown that under the agency model that we spell out in Section 7 to microfound contracting, any *constrained-feasible* policy remains incentive-compatible. Thus, *constrained-feasible* policies defined above respect the financing constraints that we assume throughout the paper, and solve the agency problem that microfound these financing constraints.[21]

The following proposition characterizes constrained-feasible policies that achieve the first-best allocation.

Proposition 4 *A policy is constrained-feasible and achieves first-best outcome if and only if* $dC_0 = N_0$, $d\underline{C}_0 = -N_0$, $\iota_t = \iota(\bar{q})$, $\psi_t = 1$ *for all* $t \geq 0$, $dC_t = 0$ *and* $d\underline{C}_t = (a - \iota(\bar{q}))K_t\, dt$ *for all* $t > 0$, $\underline{\xi}_t = (1 - \varphi_t)\psi_t K_t$ *for some* $\varphi_t \geq \tilde{\varphi}$ *and transfers* $d\tau_t$ *are chosen to keep the net worth's of experts nonnegative.*

Proof. The policies outlined in the proposition are constrained-feasible because the experts' net worth's stay nonnegative. They attain first-best because experts consume amounts equal to their entire net worth at time 0 and not consume thereafter, and because at all times all capital is allocated to experts who are forced to invest at the first-best rate $\iota_t = \iota(\bar{q})$. Note that after time 0, experts may receive large transfers of wealth to keep their net worth nonnegative, but they are not allowed to consume any of their net worth. ■

The main role of transfers is to ensure that the net worth of experts stays nonnegative. Interestingly, this can also be achieved by an appropriate choice of the price process q_t. That is, price stabilization policies that prevents a decline of $q_t K_t$ after an adverse shock to K_t can enhance welfare since they reduce the volatility of experts' net worth and make incentive constraint less severe. For example, by picking $\sigma^q = -\sigma$, the planner can make the experts' net worth non-random. In this case, the experts' incentive constraints can be satisfied without exposing them to risk. In short, price distortions can be a powerful device to improve upon the market outcomes that are plagued by pecuniary externalities and other inefficiencies.

6.3 Externalities

Externalities that make the market outcome constrained inefficient emerge in our setting and its generalizations. Pecuniary externalities work through prices. Individual market participants take prices as given, but as a group they affect them. While in complete frictionless market settings pecuniary externalities do not lead to inefficiencies (since a marginal change at the optimum has no welfare implication by the Envelope Theorem), in incomplete market settings this is generically not the case. Stiglitz (1982), Geanakoplos and Polemarchakis (1986), and Bhattacharya and Gale (1987) were among the first to highlight the inefficiency of a pecuniary externality.

The fire-sale externality is a pecuniary externality that arises when in crisis (i) experts are able to sell assets to another sector, e.g. vulture investors, the government

[21] If an expert diverts value from capital at rate $b_t k_t$, he obtains benefits in the amount of $\tilde{\varphi} b_t k_t \theta_t$. This action reduces the value of the firm to the expert by $\varphi_t b_t k_t \theta_t$, and creates funds of $\tilde{\varphi} b_t k_t$ outside the firm that have value $\tilde{\varphi} b_t k_t \theta_t \leq \varphi_t b_t k_t \theta_t$. Because the private costs of diverting value from the firm exceed the benefits, the expert has incentives to refrain from benefit extraction under any constrained-feasible policy.

or household sector (in our case) and (ii) the new asset buyers provide a downward-sloping demand function. When levering up ex-ante, financial experts do not take into account that in crisis, its own fire sales will depress prices that other institutions are able to sell at. This effect leads to excess leverage since they take fire-sale prices as given, i.e. a social planner would lever up less. Recent applications of this inefficiency due to pecuniary externalities within a finance context are Lorenzoni (2008) and Jeanne and Korinek (2010).

Besides the fire-sale externality, many other pecuniary externalities may exist in models of financial intermediation. These externalities arise whenever contracts are written based on prices. Examples include

- when experts can unload a fraction $1 - \varphi_t$ of risk to outside investors, there are externalities when φ_t depends on prices, for example when $\varphi_t = \Xi/q_t$. This situation arises when the expert can derive private benefits measured in output by mismanaging capital in ways that makes capital depreciate faster (see Section 7, where we formally spell out the agency problem that microfounds of balance sheets in our model)

- the terms of borrowing - the spread between the interest rate experts need to pay and the risk-free rate - may depend on prices. For example, there are externalities in the setting of Section 5, where experts face idiosyncratic jump risk

- experts may be bound by margin requirements, which may depend on both price level and price volatility

- in asset management, the willingness of investors to keep money in the fund depends on short-term returns, and thus market prices

Overall, it may be hard to quantify the effects of many of these externalities directly, because each action has rippling effects through future histories, and there can be a mix of positive and negative effects. Given that, it is best to study the overall significance of various externalities, as well as the welfare effects of possible regulatory policies, numerically on a calibrated model.

Externalities that affect the real economy and the labor sector were studied in an earlier draft Brunnermeier and Sannikov (2010) which included an extension with a labor sector.

7 Microfoundation of Balance Sheets and Intermediation

This section presents an agency model that provides a possible microfoundation for our balance sheet assumptions. First, we focus on a setting without idiosyncratic risk but introduce in addition an intermediation sector. That is, we divide expert agents into entrepreneurs and intermediaries to allow for a better interpretation of our results in the context of the financial system. Second, we extend the microfoundation to the setting with idiosyncratic jump default risk.

7.1 Intermediation Sector

Consider a continuous-time model, analogous to the one-period setting of Holmström and Tirole (1997). Experts are divided into entrepreneurs who manage capital and intermediaries who monitor entrepreneurs and facilitate lending from the less productive households. There is a double moral hazard problem. Entrepreneurs are able to divert capital at a rate of $b_t \in [0, \bar{b}]$, in order to generate private benefits. Intermediaries can reduce the private benefits that entrepreneurs derive from diverted funds. By spending capital at rate $m_t \in [0, \bar{m}]$ (per unit of capital) to monitor entrepreneurs, intermediaries lower the entrepreneurs' private benefits. Entrepreneur enjoys private benefits at the rate of $\Xi(m_t)b_t$ units of capital, where $\Xi(m_t)$ is a decreasing weakly convex function in m_t. We assume that the diversion of capital is inefficient, i.e. $\Xi(m_t) < 1$. In addintion, monitoring adds value to the project such that the quantity of capital under entrepreneur's management evolves according to

$$dk_t = (\Phi(\iota_t) - \delta + \Omega(m_t) - m_t - b_t)k_t\, dt + \sigma k_t\, dZ_t,$$

where Ω is value added through monitoring,we assume that $\Omega(m)$ is weakly concave function such that $\Omega'(m) \leq 1$.

If the market price of capital is denoted by q_t, then the value $k_t q_t$ of an investment in capital evolves according to

$$\frac{dV_t}{V_t} = \left(\frac{a - \iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q + \Omega(m_t) - m_t - b_t\right) dt + (\sigma + \sigma_t^q)\, dZ_t,$$

There is a continuum of entrepreneurs and intermediaries who act in a competitive and anonymous market. The market for capital is fully liquid and individual market participants have no price impact.

In order to solve the agency problems, entrepreneurs, intermediaries and less productive household investors sign a contract that specifies how the value produced by the investment and production activities of the entrepreneur are divided among the three groups. As in our baseline model, we make the following two assumptions about contracting:

A The allocation of profit is determined by the total value of capital, and shocks to k_t or q_t separately are not contractible

B Lockups are not allowed - at any moment of time any party can break the contractual relationship and the value of assets is divided among the parties the same way independently of who breaks the relationship

Condition A creates an amplification channel in which market prices affect the agents' net worth. This assumption is consistent with what we see in the real world, as well as with the models of Kiyotaki and Moore (1997) and Bernanke, Gertler, and Gilchrist (1999). We assume that contracting directly on k_t is difficult because we view k_t not as something objective and static like the number of machines, but rather something much more forward looking, like the expected NPV of assets under a particular management strategy. Moreover, even though in our model there is a one-to-one correspondence between k_t and output, in a more general model this relationship could be different for different types of projects, and could depend on the private information of the expert. Furthermore, output can be manipulated, e.g. by underinvestment.

Condition B is related to daily settlement of derivatives contracts through the mark-to-market process. It subsumes a degree of anonymity, so that once the relationship breaks, parties never meet again and the outcome of the relationship that just ended affects future relationships only through net worth. This assumption prevents commitment to long-term contracts, such as in the setting of Myerson (2010). However, the restriction is natural, and in many settings it is possible to implement an optimal long-term contract through short-term contracts with continuous marking-to-market, as shown in Fudenberg, Holmström, and Milgrom (1990).[22]

As in our baseline model, Conditions A and B and anonymity implies that the payoff of any agent is proportional to his net worth. Denote by θ_t^e the value that an entrepreneur gets per unit of net worth, and by θ_t^i the value that an intermediary gets per unit of net worth. Furthermore, denote by φ_t^e the fraction of asset risk that the entrepreneur bears and by φ_t^i the fraction that the intermediary bears. Then φ_t^e, φ_t^i and $1-\varphi_t^e-\varphi_t^i$ are the equity investments of the entrepreneurs, intermediaries and households respectively. Then changes in the value of the project V_t affect the net worth's of the participating entrepreneur, intermediary and less productive household according to

$$n_t^e + n_t^i + n_t^h = V_t, \qquad dn_t^e = \varphi_t^e dV_t + f_t^e\, dt,$$

$$dn_t^i = \varphi_t^i\, dV_t + f_t^i\, dt \qquad \text{and} \qquad dn_t^h = \varphi_t^h\, dV_t - (f_t^e + f_t^i)\, dt,$$

where contractually specified variables f_t^e and f_t^i adjust for the cost of capital committed by each party, as well as management fees of the entrepreneurs and intermediaries. This system of profit-sharing captures the most general set of contracts allowed under Conditions A and B. Figure 7 depicts this more general financing structure, in which entrepreneurs hold capital receive funds from intermediaries. Financial intermediaries issue debt b_t claims as well as outside equity towards less productive households, and households may also hold entrepreneur equity directly.

Let us discuss the relationship between equity shares φ_t^e and φ_t^i and the entrepreneur incentives to divert funds for private benefits as well as intermediary incentives to monitor. By extracting private benefit $b_t \geq 0$, the entrepreneur decreases the value of the assets at rate b_t, causing an impact of

$$-\varphi_t^e b_t \theta_t^e$$

on the entrepreneur's payoff. The monitoring intensity of the intermediary is observable by the entrepreneur, but not by the less productive household investors (and, therefore,

[22] In our setting the result of Fudenberg, Holmström, and Milgrom (1990) does not hold because of incomplete contracting, as long-term contracts can subtly make the allocation of profits dependent on q_t and k_t separately, instead of the total value.

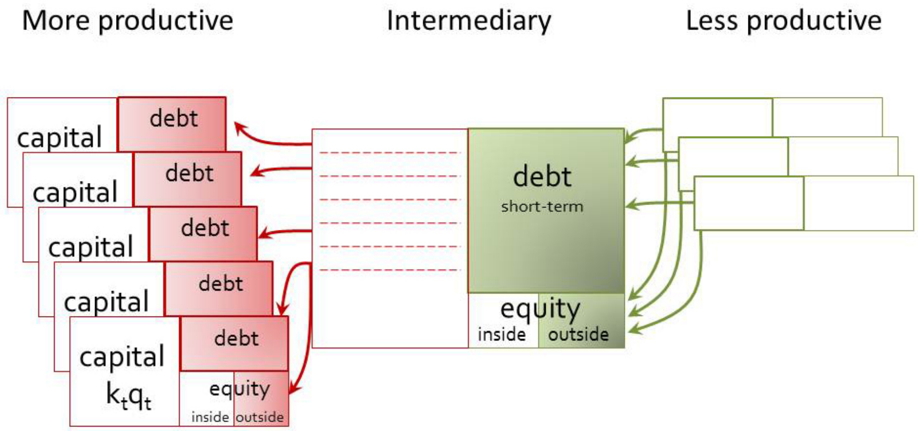

Figure 7: Balance sheets structures of entrepreneurs and financial intermediaries.

it is not contractible). Of the diverted funds, the entrepreneur receives $\Xi(m_t)b_t$, and $(1-\Xi(m_t))\,b_t$ is wasted. The entrepreneur can use diverted funds to receive the payoff of

$$\Xi(m_t)b_t\theta_t^e.$$

It is incentive-compatible for the entrepreneur to not pursue private benefits if and only if

$$\Xi(m_t)b_t\theta_t^e - \varphi_t^e b_t\theta_t^e \le 0 \Rightarrow \varphi_t^e \ge \Xi(m_t) \equiv \varphi^e(m_t).$$

Otherwise, the entrepreneur diverts capital at the maximal rate, i.e. $b_t = \bar{b}$. When $\Xi(m_t) > 0$, the incentive constraint also implies a solvency constraint, since it is possible to reward and punish the entrepreneur only as long as his net worth is positive. For high enough monitoring effort m_t such that $\Xi(m_t) = 0$, the intermediary completely eliminates the entrepreneur's agency problem through monitoring, and it becomes possible to employ insolvent entrepreneurs with zero net worth.

The monitoring incentives of the intermediary are more complicated. The intermediary knows that for low monitoring intensities such that $\Xi(m_t) > \varphi_t^e$, the entrepreneur will divert funds at rate and for higher monitoring intensities such that $\Xi(m_t) \le \varphi_t^e$, the entrepreneur will not divert funds. Since $\Omega'(m) \le 1$ for all m, the marginal costs of monitoring always exceed the direct marginal benefits on the productivity of capital. Thus, the intermediary always chooses the lowest level of monitoring that induces a given action of the entrepreneur: either monitoring intensity 0 that allows benefit extraction at a rate of $\bar{b}$, or the lowest monitoring intensity, such that $\Xi(m') = \varphi_t^e$, to prevent benefit extraction altogether. The intermediary has incentives to prevent benefit extraction if his share of equity φ_t^i satisfies

$$\varphi_t^i\,\Omega(m') - m' \ge \varphi_t^i\,(\Omega(0) - \bar{b}) \quad \Leftrightarrow \quad \varphi_t^i \ge \varphi^i(m') \equiv \frac{m'}{\Omega(m') - \Omega(0) + \bar{b}}.$$

Thus, the class of contracts in which the entrepreneur refrains from benefit extraction is characterized by pairs

$$\left\{(\varphi_t^e, \varphi_t^i) \text{ such that } \varphi_t^e = \varphi^e(m) \text{ and } \varphi_t^i \ge \varphi^i(m),\ \text{for some } m \in [0, \bar{m}]\right\}.$$

Whenever $\varphi_t^i < \varphi^i(m)$ for m defined by $\varphi_t^e = \varphi_e^i(m)$, the intermediary monitors with zero intensity and the entrepreneur pursues benefit extraction, so the value of assets follows

$$\frac{dV_t}{V_t} = \left(\frac{a - \iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q - \bar{b}\right)\, dt + (\sigma + \sigma_t^q)\, dZ_t.$$

We assume that $\bar{b}$ is large enough, so that capital management under entrepreneurs who extract private benefits is always less efficient than under the less productive households.

The equilibrium conditions that determine θ_t^e, θ_t^i, q_t, ψ_t, the monitoring intensity m_t and risk allocation $(\varphi_t^e, \varphi_t^i)$ are

$$\max_{m_t \geq 0} \frac{a - \iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q + \Omega(m_t) - m_t - r + (\varphi^e(m)\sigma_t^{\theta,e} + \varphi^i(m)\sigma_t^{\theta,i})(\sigma + \sigma_t^q) = 0$$

$$\mu_t^{\theta,e} = \rho^e - r, \quad \mu_t^{\theta,i} = \rho^i - r, \quad \varphi_t^e = \varphi^e(m_t), \quad \varphi_t^i = \varphi^i(m_t),$$

and

$$\frac{a}{q_t} - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q \geq r, \text{ with equality if } 1 - \psi_t > 0,$$

where ρ^e and ρ^i are the discount rates of entrepreneurs and intermediaries.

Together with the two conditions that arise directly from the Bellman equation, these *seven* conditions determine *seven* equilibrium processes.

The equilibrium can be characterized with state variables $\eta_t^e = N_t^e/K_t$, and $\eta_t^i = N_t^i/K_t$, where N_t^e and N_t^i are aggregate net worth's of entrepreneur and intermediary sectors. The evolution of net worth's over time is given by the equations

$$dN_t^e = rN_t^e\, dt + \psi_t\varphi_t^e(\sigma + \sigma_t^q)K_t(-\sigma_t^{\theta,e}\, dt + \, dZ_t) - dC_t^e,$$

and

$$dN_t^i = rN_t^i\, dt + \psi_t\varphi_t^i(\sigma + \sigma_t^q)K_t(-\sigma_t^{\theta,i}\, dt + \, dZ_t) - dC_t^i,$$

and aggregate capital evolves according to

$$dK_t = (\psi_t(g(q_t) + \Omega(m_t) - m_t) - (1 - \psi_t)\underline{\delta})\, K_t\, dt + \sigma K_t\, dZ_t.$$

Reduced Form with One State Variable. Next, we consider three special cases in which the general two-state-variable model reduces to our baseline model with one state variable. The first case is when the net worth of entrepreneurs is $N_t^e = 0$. The second case is when the net worth of intermediaries is $N_t^i = 0$. The third case is when the net worth of entrepreneurs and intermediaries are perfect substitutes, so only the total net worth $N_t = N_t^e + N_t^i$ matters.

First, if $N_t^e = 0$, then the only way to discipline entrepreneurs to refrain from benefit extraction is through "complete" monitoring. Without loss of generality suppose that $\bar{m}$ is the minimal monitoring intensity such that $\Xi(m) = 0$ and let us normalize $\Omega(\bar{m}) = \bar{m}$ (we can always change the entire function Ω and the depreciation rate δ by the same constant to make sure that $\Omega(\bar{m}) = \bar{m}$). Then under the monitoring intensity $\bar{m}$, capital managed by entrepreneurs evolves according to the equation

$$\frac{dk_t}{k_t} = (\Phi(\iota_t) - \delta + \Omega(\bar{m}) - \bar{m})\ dt + \sigma\ dZ_t = (\Phi(\iota_t) - \delta)\ dt + \sigma\ dZ_t,$$

as in the baseline model. The minimal fraction of equity that gives the intermediaries incentives to eliminate benefit extraction by entrepreneurs is

$$\tilde{\varphi} \equiv \varphi^i(\bar{m}) = \frac{\bar{m}}{(\Omega(\bar{m}) - \Omega(0) + \bar{b})}.$$

Thus, with the constraint $\varphi_t^i \geq \tilde{\varphi}$, the setting is reduced to our baseline model, as summarized by the following proposition:

Proposition 5 *If $N_0^e = 0$, $\Xi(m) = 0$ and $\Omega(\bar{m}) = \bar{m}$ then the model with entrepreneurs and intermediaries reduces to our baseline model with the constraint that $\varphi_t^i \geq \tilde{\varphi}$ and $\rho = \rho^i$.*

Proof. Under the assumptions of Proposition 3, it is possible to prevent entrepreneurs from extracting benefits only through monitoring. Moreover, because $\Omega'(m) \leq 1$, monitoring level $m = \bar{m}$ is optimal. The equilibrium conditions are reduced to

$$\frac{a - \iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q + \Omega(m_t) - m_t - r = -\tilde{\varphi}\sigma_t^{\theta,i}(\sigma + \sigma_t^q),$$

$$\mu_t^{\theta,i} = \rho^i - r, \qquad \text{and}$$

$$\underline{a}/q_t - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q \geq r, \text{ with equality if } 1 - \psi_t > 0,$$

as in our baseline model. The intermediaries take the role of expert, and value functions of entrepreneurs are no longer relevant because they have zero net worth. ■

Second, the general model also reduces to our baseline model if we assume (as in BGG or KM) that there are no intermediaries who can monitor entrepreneurs, i.e $N_t^i = 0$. Normalizing $\Omega(0) = 0$ for the monitoring intensity $m = 0$, capital managed by entrepreneurs evolves according to

$$\frac{dk_t}{k_t} = (\Phi(\iota_t) - \delta + \Omega(0) - 0)\ dt + \sigma\ dZ_t = (\Phi(\iota_t) - \delta)\ dt + \sigma\ dZ_t,$$

as in the baseline model. The minimal fraction of equity that gives the entrepreneurs incentives to refrain from benefit extraction is

$$\varphi_t^e \geq \tilde{\varphi} = \Xi(0).$$

Third, the model also collapses to a one-state variable model if the net worth of intermediaries and entrepreneurs are perfect substitutes. In this case, we can merge entrepreneurial and intermediary sectors to an "expert sector". The sum of entrepreneurial and intermediary net worths N^e and N^i, divided by capital K_t, forms the single state variable. This case emerges, for example, when entrepreneurs and intermediaries have the same discount rate $\rho = \rho^e = \rho^i$, the direct benefits of monitoring offset the monitoring costs, i.e. $\Omega(m) = m$, and when for all m, $\varphi^e(m) + \varphi^i(m) = \tilde{\varphi}$. In this case, the financing constraint is that entrepreneurs and intermediaries must together hold at least a fraction $\tilde{\varphi}$ of entrepreneur equity, so it is their total net worth that determines their ability to absorb risk.

7.2 Contracting with Idiosyncratic Losses and Costly State Verification

The reduced-form model with a single state variable can be extended to allow for idiosyncratic shocks to capital. Below we discuss the extension under the specific assumption that there are no intermediaries, i.e. $N_t^i = 0$. The other two cases, with $N_t^e = 0$ and with N_t^e and N_t^i being perfect substitutes, also admit similar extensions with details dependent on the case at hand.[23]

Assume as in Section 5.2 that in the absence of benefit extraction, capital managed by expert (entrepreneur) $i \in I$ evolves according to

$$dk_t = (\Phi(\iota_t) - \delta)\, k_t\, dt + \sigma k_t\, dZ_t + k_t\, dJ_t^i,$$

where dJ_t^i is a loss process with intensity λ, and the distribution the percentage of capital *recovered* in the event of a loss given by the function $F : [0,1] \to [0,1]$. We assume that dJ_t^i is a compensated process, i.e. in the absence of jumps it has a positive drift of

$$dJ_t^i = \left(\lambda \int_0^1 (1-x) dF(x)\right)\, dt,$$

so that $E[dJ_t^i] = 0$.

The entrepreneur can extract benefits continuously or via discrete-jumps. Benefit extraction is described by a non-decreasing process $\{B_t, t \geq 0\}$, which alters the law of motion of capital to

$$dk_t = (\Phi(\iota) - \delta)\, k_t\, dt + \sigma k_t\, dZ_t + k_t\, dJ_t^i - dB_t,$$

and gives entrepreneur benefits at the rate of $\tilde{\varphi} dB_t$ units of capital. The jumps in B_t are bounded by k_{t-}, i.e. the total amount of capital under entrepreneur management just before time t.

This specification of the agency problem between the entrepreneur and less productive household investors corresponds to our earlier formulation when the rate of extraction is finite, i.e. $dB_t = b_t\, dt$. However, now the entrepreneur may also extract benefits discontinuously, including in quantities that reduce the value of capital under management below the value of debt.

We assume that there is a monitoring technology that can be employed in the event of discrete drops in capital. In particular, if a monitoring action is triggered by outside investors when capital drops from k_{t-} to k_t at time t, then investors

(i) learn whether a drop in capital was caused partially by entrepreneur's benefit extraction at time t and in what amount

(ii) recover all capital that was diverted by the entrepreneur at time t

[23] For the case when $N_t^e = 0$, we have to allow intermediaries to extract benefits via discrete jumps. For the case when both entrepreneurs and intermediaries have positive net worth, we need to make the assumption that entrepreneurs can extract benefits only occasionally, during opportunities that arrive at a Poisson rate.

(iii) pay the monitoring cost of $c(k_{t-}+dJ_t^i)$ that is proportional to the amount of capital recovered $k_{t-}+dJ_t^i$.

If the drop in capital from k_{t-} to k_t at time t was partially caused by benefit extraction, then $k_{t-}+dJ_t^i > k_t$. In this event (usually $dJ_t^i = 0$ - unless the entrepreneur can anticipate Poisson losses, it is a probability-zero event for a discrete extraction of benefits to coincide with a discrete drop in asset value), the entrepreneur is not able extract any benefits, as diverted capital $k_{t-}+dJ_t^i - k_t$ is returned to investors through the monitoring process.

About contracting, we maintain the same assumptions as before about what happens in the event that no monitoring action is triggered, i.e. (A) the contract determines how the total market value of assets is divided between the entrepreneur and outside investors, and (B) at any moment either party can break the relationship and walk away with its share of assets. In particular, contracting on k_t or q_t separately is not possible. In addition, the contract specifies conditions under which a change (k_t, q_t) in triggers a monitoring action. In this event, the contract specifies how the remaining assets, net of monitoring costs, are divided among the contracting parties conditional on the amount of capital that was diverted at time t. As in Townsend (1979) and BGG, we assume that the monitoring action is not randomized, i.e. it is completely determined by the asset value history.

The following proposition shows that with Poisson losses and costly state verification, there exists an optimal contract of the same form as we considered previously, i.e. the entrepreneur holds a fraction $\tilde{\varphi}$ of equity, with the modification that a monitoring action is triggered in the event that the value of the assets falls below the value of debt.

Proposition 6 *Assume that $\sigma_t^q \geq 0$ and $\sigma_t^\theta \leq 0$. Then there is an optimal contract in which the entrepreneur holds a fraction $\tilde{\varphi}$ of the firm's equity, and operates as long as his net worth is positive. Default with costly state verification is triggered in the event that the value of assets drops by more than $n_{t-}/\tilde{\varphi}$, and in the event of default the entrepreneur does not get any of the value recovered through the verification process.*

Proof. First, it is suboptimal to employ monitoring the event that the value of the firm's assets falls by less than $n_{t-}/\tilde{\varphi}$ due to a jump. It is possible to guarantee that jumps of size $n_{t-}/\tilde{\varphi}$ or less are never caused by benefits extraction by subtracting value $\tilde{\varphi}k_t|dJ_t^i|$ from the expert's inside equity stake in the event that such a jump occurs. Such an incentive mechanism is costless, since the expert is risk-neutral with respect to jump risks as they are uncorrelated with the experts' marginal value of net worth θ_t. At the same time, monitoring carries the deadweight loss of a verification cost. Also, monitoring is not an effective way to prevent continuous diversion of private benefits, because outside investors have to pay a positive cost of monitoring in response to a possible infinitesimal deviation (recall that we disallow randomized monitoring).

Second, monitoring has to be employed in the event that the value of the firm's assets falls by more than $n_{t-}/\tilde{\varphi}$, because it is the only way to prevent benefit extraction in such large quantities (other than simply keeping the value of the assets below $n_{t-}/\tilde{\varphi}$). Without loss of generality, we can consider contracts that leave the expert with zero net worth if he is caught diverting such large amounts for private benefit. In the event that a loss of size more than $n_{t-}/\tilde{\varphi}$ is verified to have occurred without benefit extraction, recovered capital can be split arbitrarily between the expert and outside investors in an optimal contract. Because the expert is risk-neutral with respect to idiosyncratic risks uncorrelated with aggregate shocks, without loss of generality we can assume that all recovered capital goes to outside debt holders.[24]

In this case, to compensate outside investors for monitoring costs and for the expected value lost in possible default (i.e. event when costly state verification is triggered), expert's net worth has to evolve according to

$$\begin{aligned} dn_t &= rn_t\, dt + (k_t q_t)\Bigg[\left(\frac{a-\iota(q_t)}{q_t} + g(q_t) + \mu_t^q + \sigma\sigma_t^q - r - L(\vartheta_t) - C(\vartheta_t)\right) dt \\ &\quad + dJ_t^i + \tilde{\varphi}(\sigma+\sigma_t^q)\, dZ_t\Bigg] - dc_t, \end{aligned}$$

where $\vartheta_t = 1 - n_t/(\tilde{\varphi} q_t k_t)$ is the expert's debt to total asset ratio (leverage). We set $\varphi_t = \tilde{\varphi}$ to minimize the expert's exposure to aggregate risk, because we assumed that $\sigma_t^q \geq 0$ and $\sigma_t^\theta \leq 0$. ■

8 Conclusions

Events during the great liquidity and credit crunch in 2007-10 have highlighted the importance of financing frictions for macroeconomics. Unlike many existing papers in macroeconomics, our analysis is not restricted to local effects around the steady state. Importantly, we show that non-linear effects in form of adverse feedback loops and liquidity spirals are significantly larger further away from the steady state. Especially volatility effects and behavior due to precautionary motives cause these large effects. In addition to exogenous risk, swings in the economy are amplified by endogenous risk that arises due to financial frictions. This leads to interesting asset pricing implications with time-varying risk premia even though all agents in the economy are risk-neutral.

[24]There are other optimal contracts, for example the expert could be fully insured against drops in asset value that are verified to involve no benefit extraction. Of course, the expert would have to pay a 'premium' for such insurance in the event that there were no jump losses.

Appendix A: Proofs

Proof of Lemma 1. Consider two experts A and B with net worth's n_t^A and n_t^B, respectively. Denote by u_t^A and u_t^B the maximal expected utilities that these experts can get in equilibrium from time t onwards. We need to show that $u_t^A/n_t^A = u_t^B/n_t^B$. Suppose not, e.g. $u_t^A/n_t^A > u_t^B/n_t^B$. Denote by $k_s^A, dc_s^A, \varphi_s^A; s \geq t$ the optimal dynamic strategy of expert A, which attains utility u_t^A, i.e.

$$u_t^A = E_t \left[\int_t^\infty e^{-\rho(s-t)} dc_{t+s}^A \right].$$

Because the strategy is feasible, the process

$$dn_s^A = rn_s^A ds + (k_s^A q_s) \left[\left(\frac{a - \iota(q_s)}{q_s} + g(q_s) + \mu_s^q + \sigma\sigma_s^q - r \right) ds + \varphi_s^A (\sigma + \sigma_s^q) dZ_s \right] - dc_s^A.$$

stays nonnegative. Let $\varsigma = n_t^B/n_t^A$, and consider the strategy $\varsigma k_s^A, \varsigma dc_s^A, \varphi_s^A; s \geq t$ of expert B. This strategy is also feasible, because it leads to a non-negative wealth process $n_t^B = \varsigma n_t^A$, and it delivers the expected utility of ςu_t^A to expert B. Thus, $u_t^B \geq \varsigma u_t^A$, leading to a contradiction.

Therefore, for all experts their expected utility under the optimal trading strategy is proportional to wealth. It follows that $\theta_t = u_t^A/n_t^A = u_t^B/n_t^B$. ■

Proof of Lemma 2. First, assume that the process $\theta_t, t \geq 0$ represents marginal value of experts' net worth. Let us show that then θ_t must satisfy the Bellman equation (5), which characterizes the experts' optimal strategies, and the transversality condition. Let $\{k_t \geq 0, dc_t \geq 0, \varphi_t \geq \tilde{\varphi}\}$ be an arbitrary admissible strategy (i.e. does not violate the solvency constraint). We argue that the process

$$\Theta_t = \int_0^t e^{-\rho s} dc_s + e^{-\rho t} \theta_t n_t$$

is always a supermartingale; and it is a martingale if the strategy $\{k_t, dc_t, \varphi_t\}$ is optimal. Note that the maximal payoff that an expert can obtain at time t is

$$\theta_t n_t \geq E_t \left[\int_t^{t+s} e^{-\rho(s'-t)} \, dc_{s'} + e^{-\rho s} \theta_{t+s} n_{t+s} \right],$$

where equality is attained if the agent follows an optimal strategy from time t to $t+s$, since $\theta_{t+s} n_{t+s}$ is the maximal payoff that the agent can attain from time $t+s$ onwards. Therefore,

$$\Theta_t = \int_0^t e^{-\rho s'} dc_{s'} + e^{-\rho t} \theta_t n_t \geq E_t \left[\int_0^{t+s} e^{-\rho s'} dc_{s'} + e^{-\rho s} n_{t+s} \theta_{t+s} \right] = E_t[\Theta_{t+s}]$$

with equality if the agent follows the optimal strategy.

Differentiating Θ_t with respect to t using Ito's lemma, we find

$$d\Theta_t = e^{-\rho t}(dc_t - \rho\theta_t n_t \, dt + d(\theta_t n_t))$$

For the optimal strategy we have $E[dc_t - \rho\theta_t n_t \, dt + d(\theta_t n_t)] = 0$ (since Θ_t is a martingale), and for any arbitrary strategy we have $E[dc_t - \rho\theta_t n_t dt + d(\theta_t n_t)] \leq 0$ (since $\Theta_t$is a supermartingale). Therefore, the optimal strategy of any expert is characterized by the Bellman equation (5). To verify that the transversality condition holds under an optimal strategy k_t, c_t, φ_t, note that (a) the expected payoff of an expert with net worth n_t is given by

$$\theta_0 n_0 = E\left[\int_0^\infty e^{-\rho s} dc_s\right] = \lim_{t\to\infty} E\left[\int_0^t e^{-\rho s} dc_s\right],$$

where expectation value and limit can be interchanged by the Monotone Convergence Theorem (because $dc_s \geq 0$), and (b) for all t,

$$\theta_0 n_0 = E\left[\int_0^t e^{-\rho s} dc_s + e^{-\rho t}\theta_t n_t\right].$$

Taking $t \to \infty$ in the latter formula, and combining with the former, we get the transversality condition.

Conversely, let us show that if a process θ_t satisfies the Bellman equation and the transversality condition holds, then θ_t represents the experts' marginal value of net worth and characterizes their optimal strategies. Note that, as we just demonstrated, equation (5) implies that the process Θ_t is always a supermartingale, and a martingale for any strategy $\{k_t, dc_t, \varphi_t\}$ that attains the maximum in equation (5). Thus, any expert who follows such a strategy attains the payoff of

$$E\left[\int_0^\infty e^{-\rho s} \, dc_s\right] = \lim_{t\to\infty} E\left[\int_0^t e^{-\rho s} \, dc_s\right] = \lim_{t\to\infty}(\theta_0 n_0 - E[e^{-\rho t}\theta_t n_t]) = \theta_0 n_0$$

where the last equality follows from the transversality condition.

Any alternative strategy achieves utility

$$\lim_{t\to\infty} E\left[\int_0^t e^{-\rho s} dc_s\right] \leq \lim_{t\to\infty}(\theta_0 n_0 - E[e^{-\rho t}\theta_t n_t]) \leq \theta_0 n_0$$

where the last inequality holds because $\theta_t n_t \geq 0$. We conclude that $\theta_0 n_0$ is the maximal utility that any expert with net worth n_0 can attain, and that the optimal strategy must solve the maximization problem in the Bellman equation. ■

Proof of Lemma 3. Aggregating over all experts, the law of motion of N_t is

$$dN_t = rN_t \, dt + \psi_t(K_t q_t)[(E_t[r_t^k] - r) \, dt + \varphi_t(\sigma + \sigma_t^q) \, dZ_t] - dC_t,$$

where C_t is are aggregate payouts, and the law of motion of K_t is

$$dK_t = (\psi_t g(q_t) - (1 - \psi_t)\underline{\delta})K_t \, dt + \sigma K_t \, dZ_t$$

$$d(1/K_t) = -(\psi_t g(q_t) - (1 - \psi_t)\underline{\delta})(1/K_t) \, dt + \sigma^2(1/K_t) \, dt - \sigma(1/K_t) \, dZ_t.$$

Combining the two equations, and using Ito's lemma, we get

$$\begin{aligned} d\eta_t &= \left(r - \psi_t g(q_t) + (1-\psi_t)\underline{\delta} + \sigma^2\right)\eta_t\, dt + \psi_t q_t (E_t[r_t^k] - r)\, dt - \psi_t \varphi_t q_t \sigma(\sigma + \sigma_t^q)\, dt \\ &\quad + (\psi_t q_t \varphi_t (\sigma + \sigma_t^q) - \sigma\eta_t)\, dZ_t - d\zeta_t \end{aligned}$$

Substituting $\sigma_t^\eta = \psi_t \varphi_t q_t (\sigma + \sigma_t^q)/\eta_t - \sigma$ into the expression for the drift of η_t, we get (6). Furthermore, if $\sigma_t^q \geq 0$, $\sigma_t^\theta \leq 0$ and $\psi_t > 0$, then Proposition 1 implies that $\varphi_t = \tilde{\varphi}$, $E[r_t^k] - r = -\tilde{\varphi}\sigma_t^\theta(\sigma + \sigma_t^q)$, and so

$$\mu_t^\eta = r - \psi_t g(q_t) + (1-\psi_t)\underline{\delta} - \sigma_t^\eta(\sigma + \sigma_t^\theta) - \sigma\sigma_t^\theta$$

■

Appendix B: Contracting on k_t

Appendix B analyzes the case in which contracting directly on k_t is possible instead of $k_t q_t$. An expert manages capital that follows

$$dk_t = (\Phi(\iota_t) - \delta - b_t)\, k_t\, dt + \sigma k_t\, dZ_t.$$

where b_t is the rate of private benefit extraction, and produces output $(a - \iota_t)dt$. Furthermore, suppose that the expert can get the marginal benefit of $\tilde{\varphi} \leq 1$ units of capital per unit diverted. Denote by q_t the market price of capital, by θ_t, the value of expert funds per dollar, by $\iota(q_t)$ the optimal level of investment and by $g(q_t) = \Phi\left(\iota(q_t)\right) - \delta$ the implied growth rate. What is the optimal contract, if k_t rather than $k_t q_t$ is used as the measure of performance? In this section we follow the literature on dynamic contracting to derive the implications of contracting directly on k_t, e.g. see DeMarzo and Sannikov (2006).

Consider contracts based on the agent's net worth as a state variable. The "official" net worth follows

$$dn_t = rn_t\, dt + \beta_t\left(dk_t - g(q_t)k_t\, dt\right) - \sigma_t^\theta \beta_t \sigma k_t\, dt, \tag{8}$$

and the agent also gets funds at rate $\tilde{\varphi} b_t q_t$ if he extracts benefits $b_t \geq 0$. The incentive constraint is

$$\beta_t \geq \tilde{\varphi} q_t$$

since the expert gets $\tilde{\varphi} q_t$ units of net worth (that can be used elsewhere to gain the utility of $\tilde{\varphi} q_t \theta_t$) for one unit of capital diverted. Note that the stochastic as well as the deterministic portion of the law of motion of n_t depends directly on k_t, so households need to observe k_t directly in order to write a contract that rewards the expert according to equation (8).

Note that $e^{-\rho t}\theta_t n_t$ is a martingale when the expert refrains from extracting benefits and does not consume. We have

$$\begin{aligned} d(\theta_t n_t) &= \theta_t(rn_t\, dt + \beta_t \sigma k_t\, dZ_t - \sigma_t^\theta \beta_t \sigma k_t\, dt) + (\mu_t^\theta\, dt + \sigma_t^\theta\, dZ_t)\theta_t n_t + \sigma_t^\theta \theta_t \beta_t \sigma k_t\, dt \\ &= \theta_t(rn_t\, dt + \beta_t \sigma k_t\, dZ_t) + ((\rho - r)\, dt + \sigma_t^\theta\, dZ_t)\, \theta_t n_t \\ &= \rho(\theta_t n_t)\, dt + \text{volatility term,} \end{aligned}$$

where we use as in Section 3 the property that $\mu_t^\theta = (\rho - r)$.

Next, we study the price of capital, q_t. We derive a pricing equation for capital by setting the expected return that households earn from investing in capital to r.

If contracting is based on k_t only, then households hire experts to manage their capital, but households themselves take on the price risk. The market price of capital still depends on the experts' risk-taking capacity. The return that households earn on their capital holdings k_t is given by

$$\begin{aligned}(k_t q_t)\, \underline{r}_t^k &= (a - \iota(q_t))\, k_t\ dt + d\,(q_t k_t) - \beta_t k_t \sigma\ dZ_t + \beta_t \sigma_t^\theta \sigma k_t\ dt \\ &= (a - \iota\,(q_t))\, k_t\ dt + (q_t k_t)\, [(\mu_t^q + g\,(q_t) + \sigma\sigma_t^q)\ dt + (\sigma + \sigma_t^q)\ dZ_t] \\ &\quad - \beta_t k_t \sigma\ dZ_t + \beta_t \sigma_t^\theta \sigma k_t\ dt\end{aligned}$$

If $\sigma_t^\theta < 0$, then households optimally set $\beta_t = \tilde{\varphi} q_t$ to minimize the costs of compensating experts for risk. In expectation $\underline{r}_t^k$ should equal r, so we need

$$\frac{a - \iota\,(q_t)}{q_t} + \mu_t^q + g(q_t) + \sigma\sigma_t^q - r + \tilde{\varphi}\sigma_t^\theta \sigma = 0.$$

This equation is different from the pricing equation (EK) because the risk premium is based only on exogenous risk (for which households must compensate the experts that manage their capital).

Also, the law of motion of η_t will be different. Combining the law of motion of n_t and the condition that the households must get an expected return of r, we get the equation

$$dn_t = r n_t\ dt + (k_t q_t) \left[\left(\frac{a - \iota(q_t)}{q_t} + \mu_t^q + g(q_t) + \sigma\sigma_t^q - r \right) dt + \tilde{\varphi}\sigma\ dZ_t \right] - dc_t,$$

which does not have the endogenous risk term. As a result,

$$dN_t = r N_t\ dt + \psi_t\,(K_t q_t) \left[\left(\frac{a - \iota(q_t)}{q_t} + \mu_t^q + g(q_t) + \sigma\sigma_t^q - r \right) dt + \tilde{\varphi}\sigma\ dZ_t \right] - dC_t.$$

Because $dK_t/K_t = (\psi_t g\,(q_t) - (1 - \psi_t)\, \underline{\delta})\ dt + \sigma\ dZ_t$, and so

$$d\,(1/K_t)\,/\,(1/K_t) = -\,(\psi_t g(q_t) - (1 - \psi_t)\,\underline{\delta})\ dt + \sigma^2\ dt - \sigma\ dZ_t,$$

we get

$$\begin{aligned} d\eta_t &= \left(r - \psi_t g\,(q_t) + (1 - \psi_t)\,\underline{\delta} + \sigma^2\right) \eta_t\, dt \\ &\quad + \psi_t q_t \left(\frac{a - \iota(q_t)}{q_t} + \mu_t^q + g(q_t) + \sigma\sigma_t^q - r \right)\ dt \\ &\quad - \psi_t \tilde{\varphi} q_t \sigma^2\ dt + (\psi_t \tilde{\varphi} q_t - \eta_t)\, \sigma\, dZ_t - d\zeta_t. \end{aligned}$$

The volatilities of η_t and q_t are found to be

$$\sigma_t^\eta = \left(\frac{\psi_t \tilde{\varphi} q_t}{\eta_t} - 1 \right) \sigma \quad \text{and} \quad \sigma_t^q = \frac{q'(\eta_t)}{q_t} \sigma_t^\eta \eta_t,$$

so there is still amplification through leverage, but no more feedback effect through prices.

Appendix C. Stationary Distribution

Suppose that X_t is a stochastic process that evolve on the state space $[x_L, x_R]$ according to the equation

$$dX_t = \mu^x(X_t)\,dt + \sigma^x(X_t)\,dZ_t \tag{9}$$

If at time $t = 0$, X_t is distributed according to the density $d(x, 0)$, then the density of X_t at all future dates $t \geq 0$ is described by the forward Kolmogorov equations:

$$\frac{\partial}{\partial t} d(x,t) = -\frac{\partial}{\partial x}\left(\mu^x(x)\, d(x,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\sigma^x(x)^2\, d(x,t)\right).$$

If one of the endpoints is a reflecting barrier, then the boundary condition at that point is

$$-\mu^x(x) d(x,t) + \frac{1}{2}\frac{\partial}{\partial x}(\sigma^x(x)^2 d(x,t)) = 0.$$

A *stationary density* stays fixed over time under the law of motion of the process, so the left-hand side of the Kolmogorov forward equation is $\frac{\partial d(x,t)}{\partial t} = 0$. If one of the endpoints of the interval $[x_L, x_R]$ is reflecting, then integrating with respect to x and using the boundary condition at the reflecting barrier to pin down the integration constant, we find that the stationary density is characterized by the first-order ordinary differential equation

$$-\mu^x(x) d(x) + \frac{1}{2}\frac{\partial}{\partial x}(\sigma^x(x)^2 d(x)) = 0.$$

To compute the stationary density numerically, it is convenient to work with the function $D(x) = \sigma^x(x)^2 d(x)$, which satisfies the ODE

$$D'(x) = 2\frac{\mu^x(x)}{\sigma^x(x)^2} D(x). \tag{10}$$

Then $d(x)$ can be found from $D(x)$ using $d(x) = \frac{D(x)}{\sigma^x(x)^2}$.

With absorbing boundaries, the process eventually ends up absorbed (and so the stationary distribution is degenerate) unless the law of motion prevents (9) it from hitting the boundary with probability one. A non-degenerate stationary density exists with an absorbing boundary at x_L if the boundary condition $D(x_L) = 0$ can be satisfied together with $D(x_0) > 0$ for $x_0 > x_L$. For this to happen, we need

$$\log D(x) = \log D(x_0) - \int_x^{x_0} \frac{2\mu^x(x')}{\sigma^x(x')^2}\,dx' \to -\infty, \text{ as } x \to x_L$$

i.e $\int_{x_L}^{x_0} \frac{2\mu^x(x)}{\sigma^x(x)^2}\,dx = \infty$. This condition is satisfied whenever the drift that pushes X_t away from the boundary x_L (so we need $\mu^x(x) > 0$) is strong enough working against the volatility that may move X_t towards x_L. For example, if X_t behaves as a geometric Brownian motion near the boundary $x_L = 0$, i.e. $\mu^x(x) = \mu x$ and $\sigma^x(x) = \sigma x$, with $\mu > 0$, then $\int_0^{x_0} \frac{2\mu^x(x)}{\sigma^x(x)^2}\,dx = \int_0^{x_0} \frac{2\mu}{\sigma^2 x}\,dx = \infty$.

References

Adrian, T., and M. K. Brunnermeier (2010): "CoVaR," *working paper*.

Allen, F., and D. Gale (2000): "Financial Contagion," *Journal of Political Economy*, 108(1), pp.1–33.

Allen, F., and D. Gale (2007): *Understanding Financial Crisis*. Oxford University Press.

Bates, D. S. (2000): "Post-'87 Crash Fears in the S&P Futures Market," *Journal of Econometrics*, 94, 181–238.

Bernanke, B., M. Gertler, and S. Gilchrist (1999): "The Financial Accelerator in a Quantitative Business Cycle Framework," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, chap. 21, pp. 1341–1393. Elsevier.

Bernanke, B. S., and M. Gertler (1989): "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 79(1), 14–31.

Bhattacharya, S., and D. Gale (1987): "Preference Shocks, Liquidity, and Central Bank Policy," in *New Approaches to Monetary Economics*, ed. by W. A. Barnett, and K. J. Singleton, pp. 69–88. Cambridge University Press.

Brunnermeier, Markus, K. (2009): "Deciphering the Liquidity and Credit Crunch 2007-08," *Journal of of Economic Perspectives*, 23(1), 77–100.

Brunnermeier, M. K., G. Gorton, and A. Krishnamurthy (2010): "Risk Topography," working paper.

Brunnermeier, M. K., and L. Pedersen (2009): "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22(6), pp.2201–2238.

Brunnermeier, M. K., and Y. Sannikov (2010): "A Macroeconomic Model with a Financial Sector," mimeo.

Caballero, R., and A. Krishnamurthy (2004): "Smoothing Sudden Stops," *Journal of Economic Theory*, 119(4), pp.104–127.

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.

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

Christiano, L., R. Motto, and M. Rostagno (2003): "The Great Depression and the Friedman-Schwartz Hypothesis," *Journal of Money, Credit & Banking*, 35(6), pp.1119–1197.

——— (2007): "Shocks, Structures or Monetary Policies? The Euro Area and US After 2001," *Journal of Economic Dyanmics and Control*, 32(8), 2476–2506.

Cooper, G. (2008): *The Origin of Financial Crisis : Central Banks, Credit Bubbles, and the Efficient Market Fallacy*. Vintage.

Curdia, V., and M. Woodford (2009): "Credit Spreads and Monetary Policy," working paper.

DeMarzo, P. M., and Y. Sannikov (2006): "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *The Journal of Finance*, 61(6), 2681–2724.

Diamond, D. W. (1984): "Financial Intermediation and Delegated Monitoring," *The Review of Economic Studies*, 51(3), 393–414.

Diamond, D. W., and P. H. Dybvig (1983): "Bank Runs, Deposit Insurance, and Liquidity," *The Journal of Political Economy*, 91(3), 401–419.

Driessen, J., P. Maenhout, and G. Vilkov (2009): "The Price of Correlation Risk: Evidence from Equity Options," *Journal of Finance*, 64(3), 1377–1406.

Erb, C. B., C. R. Harvey, and T. E. Viskanta (1994): "Forecasting International Equity Correlations," *Financial Analysts Journal*, 50(6), 32–45.

Fisher, I. (1933): "The Debt-Deflation Theory of Great Depressions," *Econometrica*, 1(4), 337–357.

Fudenberg, D., B. Holmström, and P. Milgrom (1990): "Short-term Contracts and Long-term Agency Relationships," *Journal of Economic Theory*, 51(1), 1–31.

Garleanu, N., and L. Pedersen (2010): "Margin-Based Asset Pricing and Deviations from the Law of One Price," working paper.

Geanakoplos, J. (1997): "Promises Promises," in *The Economy as an Evolving Complex System II*, ed. by W. B. Arthur, S. Durlauf, and D. Lane, pp. 285–320. Addison-Wesley, Reading, MA.

——— (2003): "Liquidity, Defaults, and Crashes," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 2*, vol. 2 of *Econometric Society Monographs*, pp. 170–205. Cambridge University Press.

Geanakoplos, J., and H. Polemarchakis (1986): "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations when the Asset Market is Incomplete," in *Uncertainty, Information and Communication: Essays in Honor of Kenneth J. Arrow*, ed. by W. P. Heller, R. M. Starr, and D. A. Starrett, vol. 3, pp. 65–95. Cambridge University Press.

Gertler, M., and P. Karadi (2009): "A Model of Unconventional Monetary Policy," working paper.

Gertler, M., and N. Kiyotaki (2011): "Financial Intermediation and Credit Policy in Business Cycle Ana," in *Handbook of Monetary Economics*. North Holland Elsevier.

Gromb, D., and D. Vayanos (2002): "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs," *Journal of Financial Economics*, 66(2-3), 361 – 407.

He, Z., I. G. Khang, and A. Krishnamurthy (2010): "Balance Sheet Adjustments," *IMF Economic Review*, 58, 118–156, working paper.

He, Z., and A. Krishnamurthy (2010a): "Intermediary Asset Prices," working paper.

——— (2010b): "A Model of Capital Crisis," working paper.

He, Z., and W. Xiong (2009): "Dynamic Debt Runs," working paper.

Holmström, B., and J. Tirole (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *The Quarterly Journal of Economics*, 112(3), pp.663–691.

Jaimovich, N., and S. Rebelo (2009): "Can News about the Future Drive the Business Cycle," *American Economic*, 99(4), 1097–1118.

Jeanne, O., and A. Korinek (2010): "Managing Credit Booms and Busts: A Pigouvian Taxation Approach," working paper.

Keynes, John, M. (1936): *The General Theory of Employment, Interest and Money*. Macmillan, London.

Kindleberger, C. (1993): *A Financial History of Western Europe*. Oxfod University Press, 2nd edn.

Kiyotaki, N., and J. Moore (1997): "Credit Cycles," *Journal of Political Economy*, 105(2), pp.211–248.

Ljungqvist, L., and T. J. Sargent (2004): *Recursive Macroeconomic Theory*. MIT Press.

Lorenzoni, G. (2008): "Inefficient Credit Booms," *The Review of Economic Studies*, 75(3), 809–833.

Mendoza, E. G. (2010): "Sudden Stops, Financial Crisis, and Leverage," *American Economic Rev*, 100, 1941–1966.

Minsky, H. (1986): *Stabilizing and Unstable Economy*. Yale University Press.

Myerson, R. B. (2010): "A Model of Moral Hazard Credit Cycles," Working Paper, University of Chicago.

Rampini, A., and S. Viswanathan (2011): "Collateral, Risk Management, and the Distribution of Debt Capacity," *Journal of Finance*, forthcoming.

Sannikov, Y. (2008): "A Continuous- Time Version of the Principal: Agent Problem," *The Review of Economic Studies*, 75(3), 957–984.

Shin, H. (2010): *Risk and Liquidity*. Oxford University Press, Oxford, UK, Clarendon Lecture.

Shleifer, A., and R. W. Vishny (1992): "Liquidation Values and Debt Capacity: A Market Equilibrium Approach," *The Journal of Finance*, 47(4), 1343–1366.

——— (1997): "The Limits of Arbitrage," *The Journal of Finance*, 52(1), 35–55.

——— (2010): "Unstable Banking," *Journal of Financial Economics*, 97(3), 306 – 318, The 2007-8 financial crisis: Lessons from corporate finance.

Stiglitz, J. E. (1982): "The Inefficiency of the Stock Market Equilibrium," *The Review of Economic Studies*, 49(2), 241–261.

Stokey, N., and R. Lucas (1989): *Recursive Methods in Economics Dynamics*. Harvard University Press.

Townsend, R. M. (1979): "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21(2), 265–293.

Zawadowski, A. (2009): "Entangled Financial System," working paper.