A Macroeconomic Model with a Financial Sector

By Markus K. Brunnermeier and Yuliy Sannikov*

This paper studies the full equilibrium dynamics of an economy with financial frictions. Due to highly nonlinear amplification effects, the economy is prone to instability and occasionally enters volatile crisis episodes. Endogenous risk, driven by asset illiquidity, persists in crisis even for very low levels of exogenous risk. This phenomenon, which we call the volatility paradox, resolves the Kocherlakota (2000) critique. Endogenous leverage determines the distance to crisis. Securitization and derivatives contracts that improve risk sharing may lead to higher leverage and more frequent crises.

* Brunnermeier: Department of Economics, 26 Prospect Avenue, Princeton University, Princeton, NJ, 08540, USA, markus@princeton.edu, Sannikov: Department of Economics, Fisher Hall, Princeton University, Princeton, NJ 08540, sannikov@gmail.com. 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, John Heaton, Enrique Mendoza, Raf Wouters, Yili Chien, 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, SED, University of Wisconsin, IMF, Cambridge University, Cowles Foundation, Minneapolis Fed, New York Fed, University of Chicago, the Bank of Portugal Conference, the Bank of Belgium Conference, Econometric Society World Congress in Shanghai, Seoul National University, European Central Bank, UT Austin, Philadelphia Fed, NBER summer institute, and the ECB conference. We also thank Wei Cui, Ji Huang, Andrei Rachkov, and Martin Schmalz for excellent research assistance.

Economists such as Fisher (1933), Keynes (1936), and Minsky (1986) have attributed the economic downturn of the Great Depression to the failure of financial markets. Kindleberger (1993) documents that financial crises are common in history. The current financial crisis has underscored once again the importance of the financial frictions for the business cycles. These facts raise questions about financial stability. How resilient is the financial system to various shocks? At what point does the system enter a crisis regime, in the sense that market volatility, credit spreads, and financing activity change drastically? To what extent is risk exogenous, and to what extent is it generated by the interactions within the system? How does one quantify systemic risk? Does financial innovation really destabilize the financial system? How does the system respond to various policies, and how do policies affect spillovers and welfare?

The seminal contributions of Bernanke and Gertler (1989), Kiyotaki and Moore (1997) (hereafter KM), and Bernanke, Gertler and Gilchrist (1999) (hereafter BGG) uncover several important channels through which financial frictions affect the macroeconomy. First, temporary shocks can have *persistent* effects on economic activity as they affect the net worth of levered agents. Net worth takes time to rebuild. Second, financial frictions lead to the *amplification* of shocks, directly through leverage and indirectly through prices. Thus, small shocks can have potentially large effects on the economy. The amplification through prices works through adverse feedback loops, as declining net worth of levered agents leads to a drop in prices of assets concentrated in their hands, further lowering these agents' net worth.

Both BGG and KM consider the amplification and propagation of small shocks that hit the system at its deterministic steady state, and focus on linear approximations of system dynamics. We build upon the work of BGG and KM, but our work differs in important ways. We do not assume that after a shock the economy drifts back to the steady state, and instead we allow the length of the slump to be uncertain. We solve for full dynamics of the model using continuous-time methodology and find a sharp distinction between normal times and crisis episodes. We then focus on measures such as the length, severity, and frequency of crises.

As in BGG and KM, the core of our model has two types of agents: productive *experts* and less productive *households.* Because of financial frictions, the wealth of experts is important for their ability to buy physical capital and use it productively. The evolution of the wealth distribution depends on the agent's consumption decisions, as well as macro shocks that affect the agents' balance sheets. Physical capital can be traded in markets, and its equilibrium price is determined endogenously by the agents' wealth constraints. Unlike in BGG and KM, agents in our model rationally anticipate shocks. In *normal* times, the system is near the stochastic steady state: a point at which agents reach their target leverage. The stochastic steady state is defined as the balance point, to which the system tends to return after it is hit by small shocks. At this point, experts can absorb loss-inducing adverse shocks if they have sufficient time to rebuild net worth before the following shock arrives.

The most important phenomena occur when the system is knocked off balance sufficiently far away from the steady state. The full characterization of system dynamics allows us to derive a number of important implications.

First, the system's reaction to shocks is highly *nonlinear*. While the system is resilient to most shocks near the steady state, unusually large shocks are strongly amplified. Once in a crisis regime, even small shocks are subject to amplification, leading to significant endogenous risk. At the steady state, experts can absorb moderate shocks to their net worths easily by adjusting payouts, but away from the steady state payouts cannot be further reduced. Hence, near the steady state, shocks have little effect on the experts' demand for physical capital. In the crisis states away from the steady state, experts have to sell capital to cut their risk exposures. Overall, the stability of the system depends on the experts' endogenous choice of capital cushions. As it is costly to retain earnings, excess profits are paid out when experts are comfortable with their capital ratios.

Second, the system's reaction to shocks is *asymmetric*. Positive shocks at the steady state lead to larger payouts and little amplification, while large negative shocks are amplified into crisis episodes resulting in significant inefficiencies, disinvestment, and slow recovery.

Third, endogenous risk, i.e., risk self-generated by the system, dominates the volatility dynamics and affects the experts' precautionary motive. When changes in asset prices are driven by the constraints of market participants rather than fundamentals, incentives to hold cash to buy assets later at fire-sale prices increase. The precautionary motive leads to price drops in anticipation of the crisis and to higher expected return in times of increased endogenous risk.

Fourth, our model addresses the Kocherlakota (2000) critique that amplification effects in BGG and KM are quantitatively not large enough to explain the data. Unlike in BGG and KM, the extent and length of slumps is stochastic in our model, which significantly increases the amplification and persistence of adverse shocks.

Fifth, after moving through a high-volatility region, the system can get trapped for some time in a recession with low growth and misallocation of resources. The stationary distribution is $\cup$-shaped. While the system spends most of its time around the steady state, it also spends some time in the depressed regime with low growth.

In addition, a number of comparative statics arise because we endogenize the experts' payout policy. A phenomenon, which we call the *volatility paradox*, arises. Paradoxically, lower exogenous risk can lead to more extreme volatility spikes in the crisis regime. This happens because low fundamental risk leads to higher equilibrium leverage. 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-risk environments are conducive to greater buildup of systemic risk.

Financial innovation that allows experts to hedge their idiosyncratic risk can be self-defeating, as it leads to higher systemic risk. For example, securitization of home loans into mortgage-backed securities allows institutions that originate loans to unload some of the risks to other institutions. Institutions can also 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)). We find in our model that, when experts can hedge idiosyncratic risks better among one another, they take on more leverage. This makes the system less stable. Thus, while securitization is ostensibly quite beneficial, reducing costs of idiosyncratic shocks and shrinking interest rate spreads, it unintentionally leads to amplified systemic risks in equilibrium.

When intermediaries facilitate lending from households to experts, our results continue to hold. In this case, system dynamics depend on the net worth of both intermediaries and end borrowers. As in the models of Diamond (1984) and Holmström and Tirole (1997) the role of the intermediaries is to monitor end borrowers. In this process, intermediaries become exposed to macroeconomic risks.

Our model implies important lessons for *financial regulation* when financial crises lead to spillovers into the real economy. Obviously, regulation is subject to time inconsistency. For example, policies intended to ex-post recapitalize the financial sector in crisis times can lead to moral hazard in normal times. In addition, even prophylactic well-intentioned policies can have unintended consequences. For example, capital requirements, if set improperly, can easily harm welfare, as they may bind in downturns but have little effect on leverage in good times. That is, in good times, the fear of hitting a capital constraint in the future may be too weak to induce experts to build sufficient net worth buffers to overturn the destabilizing effects in downturns. Overall, our model argues in favor of countercyclical regulation that encourages financial institutions to retain earnings and build up capital buffers in good times and that relaxes constraints in downturns.

Our model makes a strong case in favor of *macro-prudential* regulation. For example, regulation that restricts payouts (such as dividends and bonus payments) should depend primarily on aggregate net worth of all intermediaries. That is, even if some of the intermediaries are well capitalized, allowing them to pay out dividends can destabilize the system if others are undercapitalized.

Literature Review

This paper builds upon several strands of literature. At the firm level, the microfoundations of financial frictions lie in papers on capital structure in the presence of informational and agency frictions, as well as papers on financial intermediation and bank runs. In the aggregate, the relevant papers study the effects of prices and collateral values, considering financial frictions in a general equilibrium context.

On the firm level, papers such as Townsend (1979), Bolton and Scharfstein (1990), and DeMarzo and Sannikov (2006) explain why violations of Modigliani-Miller assumptions lead to bounds on the agents' borrowing capacity, as well as restrictions on risk sharing. Sannikov (2012) provides a survey of capital structure implications of financial frictions. It follows that, in the aggregate, the wealth distribution among agents matters for the allocation of productive resources. In Scheinkman and Weiss (1986), the wealth distribution between two agents matters for overall economic activity. Diamond (1984) and Holmström and Tirole (1997) emphasize the monitoring role that intermediaries perform as they channel funds from lenders to borrowers. In Diamond and Dybvig (1983) and Allen and Gale (2007), intermediaries are subject to runs. He and Xiong (2012) model runs on nonfinancial firms, and Shleifer and Vishny (2010) focus on bank stability and investor sentiment. These observations microfound the balance sheet assumptions made in our paper and in the literature that studies financial frictions in the macroeconomy.[1]

In the aggregate, a number of papers also build on the idea that adverse price movements affect the borrowers' net worth and thus financial constraints. Shleifer and Vishny (1992) emphasize the importance of the *liquidating* price of capital, determined at the time when natural buyers are constrained. Shleifer and Vishny (1997) stress that insolvency risk restricts the fund managers' ability to trade against mispricing. In Geanakoplos (1997, 2003), the identity of the marginal buyer affects prices. Brunnermeier and Pedersen (2009) focus on margin constraints that depend on volatility, and Rampini and Viswanathan (2010) stress that highly productive firms go closer to their debt capacity and hence are hit harder in a downturn.

Important papers that analyze financial frictions in infinite-horizon macro settings include KM, Carlstrom and Fuerst (1997), and BGG. These papers make use of log-linear approximations to study how financial frictions amplify shocks near the steady state of the system. Other papers, such as Christiano, Eichenbaum and Evans (2005), Christiano, Motto and Rostagno (2003, 2007), Curdia and Woodford (2010), Gertler and Karadi (2011), and Gertler and Kiyotaki (2011), use these techniques to study related questions, including the impact of monetary policy on financial frictions. See Brunnermeier, Eisenbach and Sannikov (2012) for a survey of literature on economies with financial frictions.

Several papers study nonlinear effects in economies with occasionally binding constraints. In these papers, agents save away from the constraint, but nonlinearities arise near the constraint. Notably, Mendoza and Smith (2006) and Mendoza

[1]In our model, for financial frictions to have macroeconomic impact, it is crucial that financial experts cannot hedge *at least some* of aggregate risks with other agents. Otherwise, macroeconomic effects would go away. In practice, for many reasons it is difficult to identify and hedge all aggregate risks, and as the recent work of Di Tella (2012) shows, there are forms of aggregate risk that financially constrained agents choose to leave unhedged.

(2010) study discrete-time economies, in which domestic workers are constrained with respect to the fraction of equity they can sell to foreigners, as well as the amount they can borrow. Foreigners face holding costs and trading costs with respect to domestic equity, so both domestic wealth and foreign holdings of domestic equity affect system dynamics. Near the constraint, domestic workers try to sell equity to foreigners first and then sharply reduce consumption to pay off debt. Prices are very sensitive to shocks in the "sudden stop" region near the constraint. Generally, domestic agents will accumulate savings away from the constraint, placing the economy in the region where prices are not sensitive to shocks.

Like our paper, He and Krishnamurthy (2012, 2013) (hereafter HK) use continuous-time methodology to sharply characterize nonlinearities of models with occasionally binding constraints. In their endowment economy, financial experts face equity issuance constraints. Risk premia are determined by aggregate risk aversion when the outside equity constraint is slack, but they rise sharply when the constraint binds. He and Krishnamurthy (2012) calibrate a variant of the model and show that, in crisis, equity injection is a superior policy compared to interest rate cuts or asset-purchasing programs by the central bank.

While those papers and our paper share a common theme of financially constrained agents, there are important differences. First, we prove analytically a sharp result about nonlinearity, as amplification is completely absent near the steady state of our economy but becomes large away from it. Second, our model exhibits *slow recovery* from states where assets are misallocated to less productive uses, owing to financial constraints. HK and Mendoza and Smith (2006) do not study asset misallocation, focusing instead on a single aggregate production function. The system recovers much faster in HK, where risk premia can rise without a bound in crises. Third, we introduce the *volatility paradox*: the idea that the system is prone to crises even if exogenous risk is low. Fourth, we demonstrate how financial innovation can make the system less stable. Fifth, while HK focus on stabilization policies in crisis, we study prophylactic policies and their affect on overall system stability. Also, Mendoza (2010) ambitiously builds a complex model for quantitative calibration, while we opt to clearly work out the economic mechanisms on a simple model, making use of the continuous-time methods.

Several papers identify important externalities that exist because of financial frictions. These include Bhattacharya and Gale (1987), in which externalities arise in the interbank market; Gromb and Vayanos (2002), who provide welfare analysis for a setting with credit constraints; and Caballero and Krishnamurthy (2004), who study externalities an international open economy framework. On a more abstract level these effects can be traced back to the inefficiency results in general equilibrium with incomplete markets, see e.g., Stiglitz (1982) and Geanakoplos and Polemarchakis (1986). Lorenzoni (2008) and Jeanne and Korinek (2010) focus on funding constraints that depend on prices. Adrian and Brunnermeier (2010) provide a systemic risk measure and argue that financial regulation should focus on externalities.

Our paper is organized as follows. We set up our baseline model in Section I. In Section II we develop a methodology to solve the model, characterize the equilibrium that is Markov in the experts' aggregate net worth, and present a computed example. Section III discusses equilibrium dynamics and properties of asset prices. Section IV describes the volatility paradox and discusses asset liquidity and the Kocherlakota critique. Section V analyzes the effects of borrowing costs and financial innovations. Section VI discusses efficiency and regulation. Section VII concludes.

I. The Baseline Model

In an economy without financial frictions and with complete markets, the flow of funds to the most productive agents is unconstrained, and hence the distribution of wealth is irrelevant. With frictions, the wealth distribution may change with macro shocks and affect aggregate productivity. When the net worth of productive agents becomes depressed, the allocation of resources (such as capital) in the economy becomes less efficient and asset prices may decline.

In this section we develop a simple baseline model with two types of agents, in which productive agents, experts, can finance their projects only by issuing risk-free debt. This capital structure simplifies exposition, but it is not crucial for our results. As long as frictions restrict risk-sharing, aggregate shocks affect the wealth distribution across agents and thus asset prices and allocations. In Appendix A.A1, we examine other capital structures, link them to underlying agency problems, and generalize the model to include intermediaries.

Technology

We consider an economy populated by experts and households. Both types of agents can own capital, but experts are able to manage it more productively.

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

$$y_t = ak_t,$$

per unit of time, where a is a parameter. Output serves as numeraire and its price is normalized to one. New capital can be built through internal investment. When held by an expert, capital evolves according to

$$dk_t = (\Phi(\iota_t) - \delta)k_t\,dt + \sigma k_t\,dZ_t, \tag{1}$$

where ι_t is the investment rate per unit of capital (i.e., $\iota_t k_t$ is the total investment rate) and dZ_t are exogenous aggregate Brownian shocks. Function Φ, which satisfies $\Phi(0) = 0$, $\Phi'(0) = 1$, $\Phi'(\cdot) > 0$, and $\Phi''(\cdot) < 0$, represents a standard investment technology with adjustment costs. In the absence of investment, capital managed by experts depreciates at rate δ. The concavity of $\Phi(\iota)$ represents *technological illiquidity*, i.e., adjustment costs of converting output to new capital and vice versa.

Households are less productive. Capital managed by households produces output of only

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

with $\underline{a} \leq a$, and evolves according to

$$d\underline{k}_t = (\Phi(\underline{\iota}_t) - \underline{\delta})\, \underline{k}_t\, dt + \sigma \underline{k}_t\, dZ_t,$$

with $\underline{\delta} > \delta$, where $\underline{\iota}_t$ is the household investment rate per unit of capital.

The Brownian shocks dZ_t reflect the fact that one learns over time how "effective" the capital stock is.[2] That is, the shocks dZ_t capture 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 reveals not only information about current but also future expected cash flow. In this sense our model is also linked to the literature on news-driven business cycles, see, e.g., Jaimovich and Rebelo (2009).

Preferences

Experts and less productive households are risk neutral. Households have the discount 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.[3] 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]

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

In contrast, experts have the discount rate $\rho > r$, and they cannot have negative consumption. That is, cumulative consumption of an individual expert c_t must

[2] 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 $d\kappa_t = (\Phi(\iota_t/a_t) - \delta)\kappa_t dt$ where ι_t is investment per unit of *physical capital*. Effective investment ι_t/a_t is normalized by TFP to preserve the tractable scale invariance properties. The fact that investment costs increase with a_t can be justified by the fact that high TFP economies are more specialized.

[3] In an international context, one can think of a small open economy, in which foreigners finance domestic experts at a fixed global interest rate, r.

[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.

be a *nondecreasing* process, i.e., $dc_t \geq 0$. Expert utility is

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

First Best, Financial Frictions, and Capital Structure

In the economy without frictions, experts would manage capital forever. Because they are less patient than households, experts would consume their entire net worths at time 0 and finance their future capital holdings by issuing *equity* to households. The Gordon growth formula implies that the price of capital would be

$$\bar{q} = \max_{\iota} \frac{a - \iota}{r - (\Phi(\iota) - \delta)}, \tag{2}$$

so that capital earns the required return on equity, which equals the discount rate r of risk-neutral households.

If experts cannot issue equity to households, they require positive net worth to be able to absorb risks, since they cannot have negative consumption. If expert wealth ever dropped to 0, then they would not be able to hold any risky capital at all. If so, then the price of capital would permanently drop to

$$\underline{q} = \max_{\underline{\iota}} \frac{\underline{a} - \underline{\iota}}{r - (\Phi(\underline{\iota}) - \underline{\delta})},$$

the price that the households would be willing to pay if they had to hold capital forever. The difference between the first-best price $\bar{q}$ and the liquidation value $\underline{q}$ determines the *market illiquidity* of capital, which plays an important role in equilibrium.

A constraint on expert equity issuance can be justified in many ways, e.g., through the existence of an agency problem between the experts and households. There is an extensive literature in corporate finance that argues that firm insiders must have some "skin in the game" to align their incentives with those of the outside equity holders.[5] Typically, agency models imply that the expert's incentives and effort increase along with the equity stake. The incentives peak when the expert owns the entire equity stake and borrows from outside investors exclusively through risk-free debt.

While agency models place a restriction on the risk that expert net worth must absorb, they imply nothing about how the remaining cash flows are divided among outside investors. That is, the Modigliani-Miller theorem holds with respect to those cash flows. They can be divided among various securities, including risk-free debt, risky debt, equity, and hybrid securities. The choice of the securities has no effect on firm value and equilibrium. Moreover, because the assumptions

[5]See Jensen and Meckling (1976), Bolton and Scharfstein (1990), and DeMarzo and Sannikov (2006).

of Harrison and Kreps (1979) hold in our setting, there exists an analytically convenient capital structure, in which outsiders hold only equity and risk-free debt. Indeed, any other security can be perfectly replicated by continuous trading of equity and risk-free debt. More generally, an equivalent capital structure involving risky long-term debt provides an important framework for studying default in our setting. We propose an agency model and analyze its capital structure implications in Appendix A.A1.

For now, we focus on the simplest assumption that delivers the main results of this paper: experts must retain 100% of their equity and can issue only risk-free debt. If the net worth of an expert ever reaches zero, he cannot absorb any more risk, so he liquidates his assets and gets the utility of zero from then on.

Market for Capital

Individual experts and households can trade physical capital in a fully liquid market. We denote the equilibrium market price of capital in terms of output by q_t and postulate that its law of motion is of the form

$$dq_t = \mu_t^q q_t \; dt + \sigma_t^q q_t \, dZ_t. \tag{3}$$

That is, capital k_t is worth $q_t k_t$. In equilibrium q_t is determined *endogenously*, and it is bounded between $\underline{q}$ and $\bar{q}$.

Return from Holding Capital

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]

$$\frac{d(k_t q_t)}{k_t q_t} = (\Phi(\iota_t) - \delta + \mu_t^q + \sigma \sigma_t^q) \, dt + (\sigma + \sigma_t^q) \, dZ_t. \tag{4}$$

This is the experts' *capital gains rate*. The total risk of this position consists of *fundamental* risk due to news about the future productivity of capital $\sigma \, dZ_t$ and *endogenous* risk due to financial frictions in the economy, $\sigma_t^q \, dZ_t$. Capital also generates a *dividend yield* of $(a - \iota_t)/q_t$ from output remaining after internal investment. Thus, the total return that experts earn from capital (per unit of wealth invested) is

$$dr_t^k = \underbrace{\frac{a - \iota_t}{q_t} \, dt}_{\text{dividend yield}} + \underbrace{(\Phi(\iota_t) - \delta + \mu_t^q + \sigma \sigma_t^q) \, dt + (\sigma + \sigma_t^q) \, dZ_t.}_{\text{capital gains rate}} \tag{5}$$

[6]We use Ito's product rule. If $dX_t/X_t = \mu_t^X \, dt + \sigma_t^X \, dZ_t$ and $dY_t/Y_t = \mu_t^Y \, dt + \sigma_t^Y \, dZ_t$, then

$$d(X_t Y_t) = Y_t \, dX_t + X_t \, dY_t + (\sigma_t^X \sigma_t^Y)(X_t Y_t) \, dt.$$

Similarly, less productive households earn the return of

$$d\underline{r}_t^k = \underbrace{\frac{\underline{a} - \underline{\iota}_t}{q_t}\,dt}_{\text{dividend yield}} + \underbrace{(\Phi(\underline{\iota}_t) - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q)\,dt + (\sigma + \sigma_t^q)\,dZ_t}_{\text{capital gains rate}}. \tag{6}$$

Dynamic Trading and Experts' Problem

The net worth n_t of an expert who invests fraction x_t of his wealth in capital, $1 - x_t$ in the risk-free asset, and consumes dc_t evolves according to[7]

$$\frac{dn_t}{n_t} = x_t\,dr_t^k + (1 - x_t)\,r\,dt - \frac{dc_t}{n_t}. \tag{7}$$

We expect x_t to be greater than 1, i.e., experts use leverage. Less productive households provide fully elastic debt funding for the interest rate $r < \rho$ to any expert with positive net worth.[8] Any expert with positive net worth can guarantee to repay any the loan with probability one, because prices change continuously, and individual experts are small and have no price impact.

Formally, each expert solves

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

subject to the solvency constraint $n_t \geq 0, \forall t$ and the dynamic budget constraint (7).

We refer to dc_t/n_t as the consumption *rate* of an expert. Note that whenever two experts choose the same portfolio weights and consume wealth at the same rate, their expected discounted payoffs will be proportional to their net worth.

Households' Problem

Similarly, the net worth $\underline{n}_t$ of any household that invests fraction $\underline{x}_t$ of wealth in capital, $1 - \underline{x}_t$ in the risk-free asset, and consumes $d\underline{c}_t$ evolves according to

$$\frac{d\underline{n}_t}{\underline{n}_t} = \underline{x}_t\,d\underline{r}_t^k + (1 - \underline{x}_t)\,r\,dt - \frac{d\underline{c}_t}{\underline{n}_t}. \tag{8}$$

Each household solves

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

[7] Chapter 5 of Duffie (2010) offers an excellent overview of the mathematics of portfolio returns in continuous time.

[8] 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.

subject to $\underline{n}_t \geq 0$ and the dynamic budget constraint (8). Note that household consumption $d\underline{c}_t$ can be both positive and negative, unlike that of experts.

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 household consumption is allowed to be negative, ensuring that the risk-free rate is always r.[9]

Equilibrium

Informally, an equilibrium is characterized by a map from shock histories $\{Z_s, s \in [0,t]\}$, to prices q_t and asset allocations 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 $i \in \mathbb{I}$. Similarly, we denote the set of less productive households by $\mathbb{J} = (1,2]$ with index j.

Definition 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 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 $\{n_t^i, \underline{n}_t^j \geq 0\}$, capital holdings $\{k_t^i, \underline{k}_t^j \geq 0\}$, investment decisions $\{\iota_t^i, \underline{\iota}_t^j \in \mathbb{R}\}$, and consumption choices $\{dc_t^i \geq 0, d\underline{c}_t^j\}$ of individual agents $i \in \mathbb{I}$, $j \in \mathbb{J}$; such that

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

(ii) each expert $i \in \mathbb{I}$ and each household $j \in \mathbb{J}$ solve their problems given prices

(iii) markets for consumption goods[10] and capital clear, i.e.,

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

$$\text{(9)} \quad \textit{where} \quad dK_t = \left(\int_{\mathbb{I}} (\Phi(\iota_t^i) - \delta) k_t^i \, di + \int_{\mathbb{J}} (\Phi(\underline{\iota}_t^j) - \underline{\delta}) \, \underline{k}_t^j \, dj \right) dt + \sigma K_t dZ_t.$$

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

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

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

Since agents are atomistic perfectly competitive price-takers, the distribution of wealth among experts and among households in this economy does not matter. However, the wealth of experts *relative* to that of households plays a crucial role in our model, as we discuss in the next section.

II. Solving for the Equilibrium

We have to determine how the equilibrium price q_t and allocation of capital, as well as the agents' consumption decisions, depend on the history of macro shocks $\{Z_s;\ 0 \le s \le t\}$. Our procedure to solve for the equilibrium has two major steps. First, we use the agent utility maximization and market-clearing conditions to derive the properties of equilibrium *processes*. Second, we show that the equilibrium dynamics can be described by a single state variable, the experts' wealth share η_t, and derive a system of equations that determine equilibrium *variables* as functions of η_t.

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 physical capital from experts. Less productive households' purchases are speculative as they hope to sell capital back to experts at a higher price in the future. In this sense households are liquidity providers, as they provide some of the functions of the traditional financial sector in times of crises.

Internal Investment

The returns (5) and (6) that experts and households receive from capital are maximized by choosing the investment rate ι that solves

$$\max_{\iota} \Phi(\iota) - \iota/q_t.$$

The first-order condition $\Phi'(\iota) = 1/q_t$ (marginal Tobin's q) implies that the optimal investment rate is a function of the price q_t, i.e.,

$$\iota_t = \underline{\iota}_t = \iota(q_t).$$

The determination of the optimal investment rate is a completely static problem: It depends only on the current price of capital q_t. From now on, we *incorporate the optimal investment rate* in the expressions for the returns dr_t^k and $d\underline{r}_t^k$ that experts and households earn.

Households' Optimal Portfolio Choice

Denote the fraction of physical capital held by households by

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

The problem of households is straightforward because they are not financially constrained. In equilibrium they must earn the return of r, their discount rate, from risk-free lending to experts and, if $1 - \psi_t > 0$, from holding capital. If households do not hold any physical capital, i.e., $\psi_t = 1$, their expected return on capital must be *less than or equal* to r. This leads to the equilibrium condition

$$\text{(H)} \qquad \underbrace{\frac{\underline{a} - \iota(q_t)}{q_t} + \Phi(\iota(q_t)) - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q}_{E_t[d\underline{r}_t^k]/dt} \leq r, \;\; \text{with equality if } 1 - \psi_t > 0.$$

Experts' Optimal Portfolio and Consumption Choices

The experts face a significantly more complex problem, because they are financially constrained. Their problem is *dynamic*, that is, their choice of leverage depends not only on the current price levels, but also on the entire future law of motion of prices. Even though experts are risk-neutral with respect to consumption, they exhibit risk-averse behavior in our model (in aggregate) because their marginal utility of *wealth* is stochastic — it depends on *time-varying* investment opportunities. Greater leverage leads to higher profit and also greater risk. Experts who take on more risk suffer greater losses exactly when they value their funds the most: Negative shocks depress prices and create attractive investment opportunities.

We characterize the experts' optimal dynamic strategies through the Bellman equation for their value functions. Consider a feasible strategy $\{x_t, d\zeta_t\}$, which specifies leverage x_t and the consumption *rate* $d\zeta_t = dc_t/n_t$ of an expert, and denote by

$$\text{(10)} \qquad \theta_t n_t \equiv E_t \left[\int_t^\infty e^{-\rho(s-t)} dc_s \right],$$

the expert's future expected payoff under this strategy. Note that the expert's consumption $dc_t = d\zeta_t n_t$ under the strategy $\{x_t, d\zeta_t\}$ is proportional to wealth, and therefore the expert's expected payoff is also proportional to wealth. The following proposition provides necessary and sufficient conditions for the strategy $\{x_t, d\zeta_t\}$ to be optimal, given the price process $\{q_t, t \geq 0\}$.

LEMMA II.1: *Let $\{q_t, t \geq 0\}$ be a price process for which the maximal payoff* *that any expert can attain is finite.*[11] *Then the process* $\{\theta_t\}$ *satisfies (10) under the strategy* $\{x_t, d\zeta_t\}$ *if and only if*

$$\rho\theta_t n_t\, dt = n_t\, d\zeta_t + E[d(\theta_t n_t)] \tag{11}$$

when n_t *follows (7), and the transversality condition* $E[e^{-\rho t}\theta_t n_t] \to 0$ *holds.*

Moreover, this strategy is optimal if and only if

(12)

$$\rho\theta_t n_t\, dt = \max_{\hat{x}_t \geq 0, d\hat{\zeta}_t \geq 0} n_t\, d\hat{\zeta}_t + E[d(\theta_t n_t)] \quad s.t. \quad \frac{dn_t}{n_t} = \hat{x}_t\, dr_t^k + (1-\hat{x}_t)\, r\, dt - d\hat{\zeta}_t.$$

Proposition II.2 breaks down the Bellman equation (12) into specific conditions that the stochastic laws of motion of q_t and θ_t, together with the experts' optimal strategies, have to satisfy.

PROPOSITION II.2: *Consider a finite process*

$$\frac{d\theta_t}{\theta_t} = \mu_t^\theta\, dt + \sigma_t^\theta\, dZ_t.$$

Then $n_t\theta_t$ *represents the maximal future expected payoff that an expert with net worth* n_t *can attain and* $\{x_t, d\zeta_t\}$ *is an optimal strategy if and only if*

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

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

(iii) either $x_t > 0$ and
$$\underbrace{\frac{a-\iota(q_t)}{q_t} + \Phi(\iota(q_t)) - \delta + \mu_t^q + \sigma\sigma_t^q - r}_{\text{expected excess return on capital, } E_t[dr_t^k]/dt - r} = \underbrace{-\sigma_t^\theta(\sigma+\sigma_t^q)}_{\text{risk premium}}, \tag{EK}$$

or $x_t = 0$ *and* $E[dr_t^k]/dt - r \leq -\sigma_t^\theta(\sigma+\sigma_t^q)$,

(iv) and the transversality condition $E[e^{-\rho t}\theta_t n_t] \to 0$ *holds under the strategy* $\{x_t, d\zeta_t\}$.

Under (i) through (iv), θ_t represents the experts' marginal utility of wealth (not consumption), which prices assets held by experts. The left-hand side of (EK) represents the excess return on capital over the risk-free rate. The right-hand side represents the experts' risk premium, or their *precautionary motive.* We will see that in equilibrium $\sigma_t^\theta \leq 0$ while $\sigma + \sigma_t^q > 0$, so that experts suffer losses on capital exactly in the event that better investment opportunities arise, i.e., as θ_t rises. According to the second part of (EK), if endogenous risk ever made the

[11]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).

required risk premium greater than the excess return on capital, experts would choose to hold no capital in volatile times and instead lend to households at the risk-free rate, waiting to pick up assets at low prices at the bottom ("flight to quality").

As further analysis will make clear, the precautionary motive increases with aggregate leverage of experts, but disappears completely if experts invest in capital without using leverage. Therefore, the incentives of individual experts to take on risk are decreasing in the risks taken by other experts. This leads to the equilibrium choice of leverage. We conjecture, and later verify, that experts always use positive leverage in equilibrium, so that

$$\psi_t q_t K_t > N_t, \quad \text{where} \quad N_t = \int_{\mathbb{I}} n_t^i \, di.$$

It is interesting to note that because θ_t is the experts' marginal utility of wealth, at any time t they use the *stochastic discount factor* (SDF)

$$e^{-\rho s} \frac{\theta_{t+s}}{\theta_t} \tag{13}$$

to price cash flows at a future time $t+s$. That is, the price of any asset that pays a random cash flow of CF_{t+s} at time $t+s$ is

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

Market Clearing

The market for capital clears by virtue of our notation, with shares ψ_t and $1-\psi_t$ of capital allocated to experts and households. Furthermore, markets for consumption goods and risk-free assets clear because the households, whose consumption may be positive or negative, are willing to borrow and lend arbitrary amounts at the risk-free rate r.

Wealth Distribution

Due to financial frictions, the wealth distribution across agents matters. In aggregate, experts and households have wealth

$$N_t = \int_{\mathbb{I}} n_t^i \, di \qquad \text{and} \qquad q_t K_t - N_t = \int_{\mathbb{J}} \underline{n}_t^j \, dj,$$

respectively. The experts' wealth share is

$$\eta_t \equiv \frac{N_t}{q_t K_t} \in [0, 1].$$

Experts become constrained when η_t falls, leading to a larger fraction of capital $1-\psi_t$ allocated to households, a lower price of capital q_t, and a lower investment rate $\iota(q_t)$.

Our model has convenient scale-invariance properties, which imply that η_t fully determines the price level, as well as inefficiencies with respect to investment and capital allocation. That is, for any equilibrium in one economy, there is an equivalent equilibrium with the same laws of motion of η_t, q_t, θ_t, and ψ_t in any economy scaled by a factor of $\varsigma \in (0,\infty)$.

We will characterize an equilibrium that is Markov in the state variable η_t. Before we proceed, Lemma II.3 derives the equilibrium law of motion of $\eta_t = N_t/(q_t K_t)$ from the laws of motion of N_t, q_t, and K_t. In Lemma II.3, we do not assume that the equilibrium is Markov.[12]

LEMMA II.3: *The equilibrium law of motion of η_t is*

$$\frac{d\eta_t}{\eta_t} = \frac{\psi_t-\eta_t}{\eta_t}(dr_t^k - r\,dt - (\sigma+\sigma_t^q)^2\,dt) + \frac{a-\iota(q_t)}{q_t}\,dt + (1-\psi_t)(\underline{\delta}-\delta)\,dt - d\zeta_t, \tag{14}$$

where $d\zeta_t = dC_t/N_t$, with $dC_t = \int_{\mathbb{I}}(dc_t^i)\,di$, is the aggregate expert consumption rate. Moreover, if $\psi_t > 0$, then (EK) implies that we can write

$$\frac{d\eta_t}{\eta_t} = \mu_t^\eta\,dt + \sigma_t^\eta\,dZ_t - d\zeta_t, \tag{15}$$

where $\sigma_t^\eta = \dfrac{\psi_t-\eta_t}{\eta_t}(\sigma+\sigma_t^q)$ *and* $\mu_t^\eta = -\sigma_t^\eta(\sigma+\sigma_t^q+\sigma_t^\theta) + \dfrac{a-\iota(q_t)}{q_t} + (1-\psi_t)(\underline{\delta}-\delta)$.

Markov Equilibrium

In a Markov equilibrium, all processes are functions of η_t, i.e.,

$$q_t = q(\eta_t), \quad \theta_t = \theta(\eta_t) \text{ and } \psi_t = \psi(\eta_t). \tag{16}$$

If these functions are known, then we can use equation (15) to map any path of aggregate shocks $\{Z_s, s \le t\}$ into the current value of η_t and subsequently q_t, θ_t, and ψ_t.

To solve for these functions, we need to convert the equilibrium conditions into differential equations. That is, from any tuple $(\eta, q(\eta), q'(\eta), \theta(\eta), \theta'(\eta))$, we need a procedure to convert the equilibrium conditions into $(q''(\eta), \theta''(\eta))$. Proposition II.4 does this in two steps:

[12] 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 thispaper, a result like Lemma II.3 should be helpful for the proof of uniqueness.

1) Using Ito's lemma, compute the volatilities σ_t^η, σ_t^q, and σ_t^θ and find ψ_t, and
2) compute the drifts μ_t^η, μ_t^q, and μ_t^θ, and use Ito's lemma again to find $q''(\eta)$ and $\theta''(\eta)$.

Proposition 2 also describes the domain and the boundary conditions for the system.

PROPOSITION II.4: *The equilibrium domain of functions $q(\eta)$, $\theta(\eta)$, and $\psi(\eta)$ is an interval $[0, \eta^*]$. Function $q(\eta)$ is increasing, $\theta(\eta)$ is decreasing, and the boundary conditions are*

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

The experts' consumption $d\zeta_t$ is zero when $\eta_t < \eta^*$ and positive at η^*, so that η^* is a reflecting boundary of the process η_t. The following procedure can be used to compute $\psi(\eta)$, $q''(\eta)$, and $\theta''(\eta)$ from $(\eta, q(\eta), q'(\eta), \theta(\eta), \theta'(\eta))$.

1. Find $\psi \in (\eta,\ \eta + q(\eta)/q'(\eta))$ such that[13]

$$\frac{a - \underline{a}}{q(\eta)} + \underline{\delta} - \delta + (\sigma + \sigma_t^q)\sigma_t^\theta = 0, \tag{17}$$

(18)

$$\textit{where} \qquad \sigma_t^\eta \eta = \frac{(\psi - \eta)\sigma}{1 - (\psi - \eta) q'(\eta)/q(\eta)}, \quad \sigma_t^q = \frac{q'(\eta)}{q(\eta)} \sigma_t^\eta \eta \quad \text{and} \quad \sigma_t^\theta = \frac{\theta'(\eta)}{\theta(\eta)} \sigma_t^\eta \eta.$$

If $\psi > 1$, set $\psi = 1$ and recalculate (18).

2. Compute

$$\mu_t^\eta = -\sigma_t^\eta(\sigma + \sigma_t^q + \sigma_t^\theta) + \frac{a - \iota(q(\eta))}{q(\eta)} + (1 - \psi)(\underline{\delta} - \delta),$$

$$\mu_t^q = r - \frac{a - \iota(q(\eta))}{q(\eta)} - \Phi(q(\eta)) + \delta - \sigma\sigma_t^q - \sigma_t^\theta(\sigma + \sigma_t^q), \quad \mu_t^\theta = \rho - r,$$

$$q''(\eta) = \frac{2\left[\mu_t^q q(\eta) - q'(\eta)\mu_t^\eta \eta\right]}{(\sigma_t^\eta)^2 \eta^2} \qquad \text{and} \qquad \theta''(\eta) = \frac{2\left[\mu_t^\theta \theta(\eta) - \theta'(\eta)\mu_t^\eta \eta\right]}{(\sigma_t^\eta)^2 \eta^2}. \tag{19}$$

Proposition II.4 allows us to derive analytical results about equilibrium behavior and asset prices and to compute equilibria numerically. The proof is in Appendix C.

[13] The left-hand side of (17) decreases from $(a - \underline{a})/q(\eta) + \underline{\delta} - \delta > 0$ to $-\infty$ over the interval $\psi = [\eta, \eta + q(\eta)/q'(\eta)]$.

Algorithm to Solve the Equations

The numerical computation of the functions $q(\eta)$, $\theta(\eta)$, and $\psi(\eta)$ poses challenges because of the singularity at $\eta = 0$. In addition, we need to determine the endogenous endpoint η^* and match the boundary conditions at both 0 and η^*. To match the boundary conditions, it is helpful to observe that if function $\theta(\eta)$ solves the equations of Proposition II.4, then so does any function $\varsigma\theta(\eta)$, for any constant $\varsigma > 0$. Therefore, one can always adjust the level of $\theta(\eta)$ *ex post* to match the boundary condition $\theta(\eta^*) = 1$. We use the following algorithm to calculate our numerical examples.

1) Set $q(0) = \underline{q}$, $\theta(0) = 1$ and $\theta'(0) = -10^{10}$.
2) Set $q_L = 0$ and $q_H = 10^{15}$.
3) *Guess* that $q'(0) = (q_L + q_H)/2$. Use the Matlab function *ode45* to solve for $q(\eta)$ and $\theta(\eta)$ until either (a) $q(\eta)$ reaches $\bar{q}$ or (b) $\theta'(\eta)$ reaches 0 or (c) $q'(\eta)$ reaches 0, whichever happens soonest. If $q'(\eta)$ reaches 0 before any of the other events happens, then increase the *guess* of $q'(0)$ by setting $q_L = q'(0)$. Otherwise, let $q_H = q'(0)$. Repeat until convergence (e.g., 50 times).
4) If q_H was chosen in step 2 to be large enough, then in the end $\theta'(\eta)$ and $q'(\eta)$ will reach 0 at the same point η^*.
5) Divide the entire function $\theta(\eta)$ by $\theta(\eta^*)$ to match the boundary condition $\theta(\eta^*) = 1$.

The more negative the initial choice of $\theta'(0)$, the better we can approximate the boundary condition $\theta(0) = \infty$, that is, the higher the value of $\theta(0)$ becomes after we divide the entire solution by $\theta(\eta^*)$. We provide our Matlab implementation of this algorithm in the Online Appendix.

Numerical Example

Figure 1 presents functions $q(\eta)$, $\theta(\eta)$, and $\psi(\eta)$ characterized by Proposition II.4 for parameter values $\rho = 6\%$, $r = 5\%$, $a = 11\%$, $\underline{a} = 7\%$, $\delta = \underline{\delta} = 5\%$, $\sigma = 10\%$, and $\Phi(\iota) = \frac{1}{\kappa}(\sqrt{1 + 2\kappa\iota} - 1)$ with $\kappa = 2$.[14] Under these assumptions, $\underline{q} = 0.8 \quad \text{and} \quad \bar{q} = 1.2.$

As η increases, the price of capital $q(\eta)$ increases and the marginal value of expert wealth $\theta(\eta)$ declines. Experts hold all capital in the economy when they have high net worth, when $\eta_t \in [\eta^\psi, \eta^*]$, but households hold some capital, and so $\psi(\eta) < 1$, when $\eta_t < \eta^\psi$.

The map from the history of aggregate shocks dZ_t to the state variable η_t is captured by the *drift* $\mu_t^\eta \eta$ and the *volatility* $\sigma_t^\eta \eta$, depicted on the top panels of

[14] The investment technology in this example has quadratic adjustment costs: An investment of $\Phi + \frac{\kappa}{2}\Phi^2$ generates new capital at rate Φ.

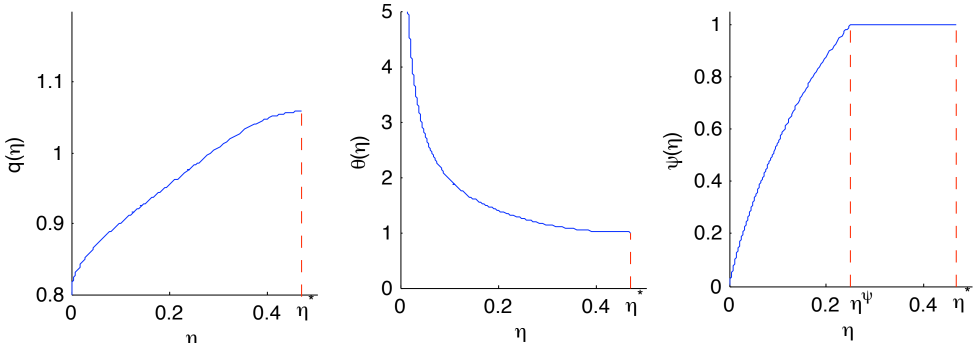

FIGURE 1. EQUILIBRIUM FUNCTIONS $q(\eta)$, $\theta(\eta)$ AND $\psi(\eta)$.

Figure 2. The drift of η_t depends the relative portfolio returns and consumption rates of experts and households. While experts are levered and earn a risk premium, households earn the risk-free return of r. The bottom panels of Figure 2 show expert leverage as well as the returns that experts and households earn from capital. Risk premia and expert leverage rise as η_t falls. The households' rate of return from capital equals r when they hold capital on $[0, \eta^\psi]$, but otherwise it is less than r.

The volatility of η_t is non-monotonic: It rises over the interval $[0, \eta^\psi]$ and falls over $[\eta^\psi, \eta^*]$. Point $\eta = 0$ is an absorbing boundary, which is never reached in equilibrium as η_t evolves like a geometric Brownian motion in the neighborhood of 0 (see Proposition III.2). Point η^* is a reflecting boundary where experts consume excess net worth.

Because η_t gravitates toward the reflecting boundary η^* in expectation, the point η^* is the *stochastic steady state* of our system. Point η^* in our model is analogous to the *deterministic steady state* in traditional macro models, such as those of BGG and KM. Similar to the steady state in these models, η^* is the point of global attraction of the system and, as we see from Figure 2 and discuss below, the volatility near η^* is low.

However, point η^* also differs from the deterministic steady state in BGG and KM in important ways. Unlike log-linearized models, our model does not set the exogenous risk σ to 0 to identify the steady state, but rather fixes the volatility of macro shocks and looks for the point where the system remains stationary in the absence of shocks. Thus, the location of η^* depends on the exogenous volatility σ. It is determined indirectly through the agents' consumption and portfolio decisions, taking shocks into account. As we discuss in Sections III and IV, the endogeneity of η^* leads to a number of important phenomena, including *nonlinearity* — the system responds very differently to small and large shocks at η^* — and the *volatility paradox* — that of the system is prone to endogenous risk even when exogenous risk σ is low.

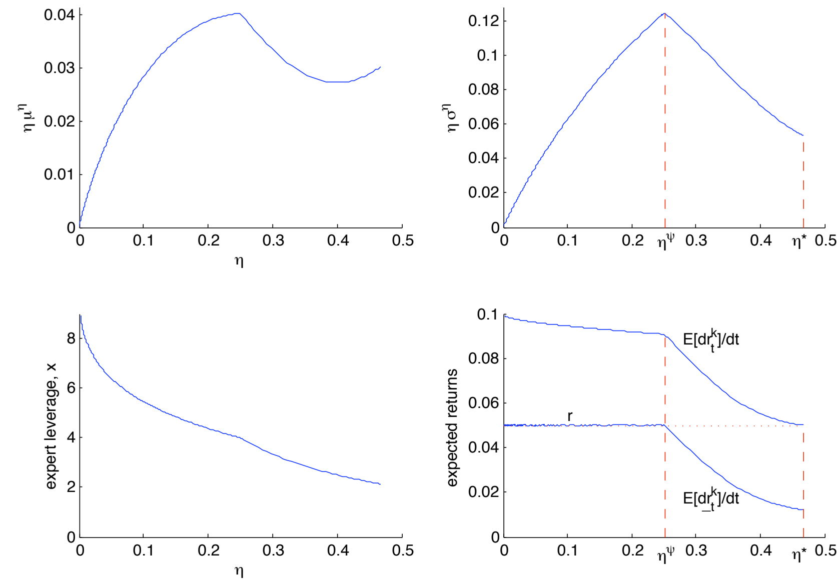

FIGURE 2. THE DRIFT AND VOLATILITY OF η_t, EXPERT LEVERAGE, AND EXPECTED ASSET RETURNS.

INEFFICIENCIES IN EQUILIBRIUM

Without financial frictions, experts would permanently manage all capital in the economy. Capital would be priced at $\bar{q}$, leading to an investment rate of $\iota(\bar{q})$. Moreover, experts would consume their net worth in a lump sum at time 0, so that the sum of utilities of all agents would be $\bar{q}K_0$. With frictions, however, there are three types of inefficiencies in our model:

(i) capital misallocation, since less productive households end up managing capital for low η_t, when $\psi_t < 1$,

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

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

Note that these inefficiencies vary with the state of the economy: They get worse when η_t drops.

Owing to these inefficiencies, the sum of utilities of all agents is less than first best utility $\bar{q}K_0$. Even at point η^* the sum of the agents' utilities is

(20)
$$\underbrace{E\left[\int_0^\infty e^{-\rho t}dC_t\right]}_{\text{expert payoff}}+\underbrace{E\left[\int_0^\infty e^{-rt}d\underline{C}_t\right]}_{\text{household payoff}} = \underbrace{\theta(\eta^*)N_0}_{\text{expert payoff}} + \underbrace{q(\eta^*)K_0 - N_0}_{\text{household payoff/wealth}} = q(\eta^*)K_0 < \bar{q}K_0,$$

since $\theta(\eta^*) = 1$ and $q(\eta^*) < \bar{q}$.

III. 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 the economy therefore subject to systemic risk?

The equilibrium exhibits *instability*, which distinguishes our analysis from the log-linearized solutions of BGG and KM. As in those papers, the price of capital in our model is subject to *endogenous risk* σ_t^q, which leads to *excess volatility*. However, unlike in BGG and KM, the amount of endogenous risk in our model varies over the cycle: It goes to zero near the steady state η^*, but it is large below the *stochastic steady state* η^*. Thus, an unusually long sequence of negative shocks throws the economy into a volatile crisis regime. More bad shocks can put the system into a depressed regime, from which it takes a long time to recover. Slow recovery implies a bimodal stationary distribution over the state space. This is in sharp contrast to papers that log linearize, predicting a much more stable system with a normal stationary distribution around 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 below the steady state. Our analysis highlights the sharp distinction between crisis and normal times.

The nonlinearities of system dynamics are robust to modeling assumptions. For example, a model with logarithmic utility would also generate low (but nonzero) amplification near the steady state, as well as high amplification below the steady state, especially at the point where experts start selling capital to households.

The differences in system dynamics near and away from the steady state have to do with the forces that determine the steady state: experts' profits and their endogenous payout/consumption decisions. The system gravitates toward a point where these two forces exactly balance each other out: the stochastic steady state. Experts accumulate net worth in crisis regimes, where volatility and *risk premia* are high. They only start paying out once their aggregate net worth recovers enough that the probability of the next crisis becomes tolerable.

Amplification Due to Endogenous Risk

Endogenous risk refers to changes in asset prices attributable not to changes in fundamentals, but rather to portfolio adjustments in response to constraints and/or precautionary motives. While exogenous fundamental shocks cause initial losses, feedback loops that arise when agents react to these losses create endogenous risk. In our model, exogenous risk σ is constant, but endogenous risk σ_t^q varies with the state of the system.

The amplification of shocks that creates endogenous risk depends on (i) expert leverage and (ii) price reactions to shocks, which feed back to the experts' net worth and lead to further adjustments. While the experts finance themselves through fully liquid short-term debt, their assets are subject to aggregate market illiquidity.[15] Figure 3 illustrates the feedback mechanism of amplification, which has been identified by both BGG and KM near the steady state of their models.

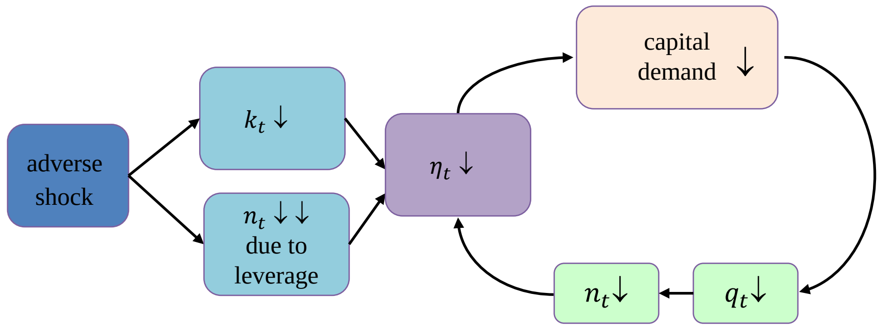

FIGURE 3. ADVERSE FEEDBACK LOOP.

The immediate effect of a shock dZ_t that reduces K_t by 1% is a drop of N_t by $\frac{\psi}{\eta}$%, and a drop of η_t by $(\frac{\psi}{\eta} - 1)$%, where $\frac{\psi}{\eta}$ is the experts' leverage ratio (assets to net worth). This drop in η_t causes the price $q(\eta_t)$ to drop by

$$\phi\% \equiv \frac{q'(\eta_t)}{q(\eta_t)} \overbrace{\left(\frac{\psi_t}{\eta_t} - 1\right) \eta_t}^{\text{total drop in } \eta} \%.$$

That is, this aftershock causes $q_t K_t$ to drop further by ϕ%, N_t further by $\frac{\psi}{\eta}\phi$% and a η_t further by $(\frac{\psi}{\eta} - 1)\phi$%. We see that the initial shock gets amplified by a factor of ϕ each time it goes through the feedback loop. If $\phi < 1$, then this loop converges with a total amplification factor of $1/(1-\phi)$ and cumulative impacts on η_t and $q(\eta_t)$ of

(21)

$$\frac{d\eta_t}{\eta_t} = \frac{\frac{\psi}{\eta} - 1}{1-\phi}\% = \frac{1}{\eta}\frac{\psi - \eta}{1 - (\psi - \eta)q'(\eta)/q(\eta)}\% \quad \text{and} \quad \frac{dq_t}{q_t} = \frac{q'(\eta)}{q(\eta)}\frac{\psi - \eta}{1 - (\psi - \eta)q'(\eta)/q(\eta)}\%$$

respectively. This leads us to formulas (18), provided by Proposition II.4, that capture how leverage and feedback loops contribute to endogenous risk.

The amplification effect of $q'(\eta)$ on the endogenous volatility σ_t^q is *nonlinear*,

[15]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 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. See Brunnermeier, Gorton and Krishnamurthy (2013).

since $q'(\eta)$ enters not only the numerator, but also the denominator of (21) and (18). If $q'(\eta)$ is so large that $\phi > 1$, then the feedback effect would be completely unstable, leading to infinite volatility.

Normal versus Crisis Times and "Ergodic Instability"

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

THEOREM III.1: *In equilibrium, at η^* the system has no amplification and $\sigma_t^q = 0$, since $q'(\eta^*) = 0$. For $\eta_t < \eta^*$, exogenous shocks spill over into prices, leading to the indirect dynamic amplification factor of $1/(1-(\psi_t - \eta_t)q'(\eta_t)/q(\eta_t))$.*

PROOF:

This result follows directly from Proposition II.4.

The left panel of Figure 4 shows the total volatility of the value of capital $\sigma + \sigma_t^q$, for our computed example. Because *endogenous* risk σ_t^q rises sharply below steady state, the system exhibits nonlinearities: Large shocks affect the system very differently from small shocks. Near the point η^ψ, increased endogenous risk and leverage lead to a high volatility of η_t, as seen in Figure 2. This leads to *systemic* risk, and the economy occasionally ends up in a depressed regime far below the steady state, where most of the capital is allocated inefficiently to households.

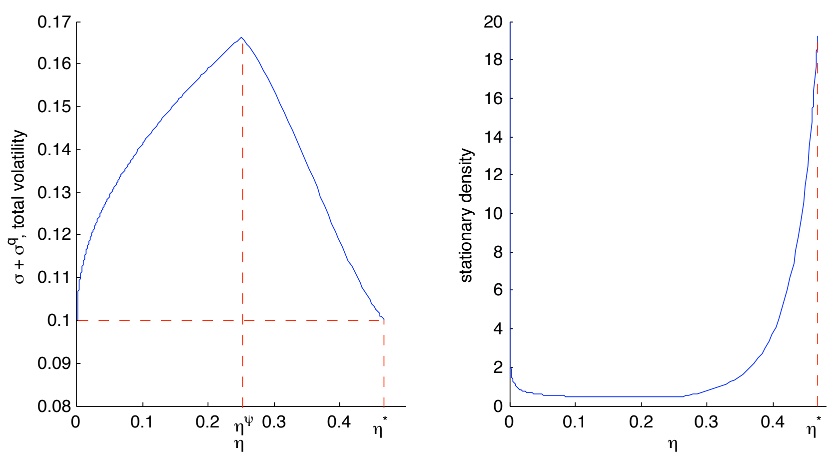

Figure 4. Systemic risk: total volatility of capital and the stationary density of η_t.

The right panel of Figure 4 shows the stationary distribution of η_t. Stationary density measures the average amount of time that the variable η_t spends in the long run at different parts of the state space. The stationary distribution can be computed from the drift and volatility of η_t using Kolmogorov forward equations (see Appendix B).

The key feature of the stationary distribution in Figure 4 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.

The stationary density is high near η^*, as it is the attracting point of the system, but very thin in the volatile middle region below η^*. The system moves fast through the regions of high volatility, and so the time spent there is short. The excursions below the stochastic steady state are characterized by high uncertainty that may take the system to depressed states near $\eta = 0$. In other words, the economy is subject to breakdowns, 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 the financial system can deal well with small adverse shocks, or the economy gets trapped for a long time near very low η-values.[16]

The following proposition formally demonstrates that the stationary density (if it exists) indeed has peaks at $\eta = 0$ and $\eta = \eta^*$. The proof in Appendix C shows that variable η_t evolves like a geometric Brownian motion in a neighborhood of 0, and it uses the Kolmogorov forward equation to characterize the stationary density near 0. The stationary distribution may fail to exist if the experts' productive advantage is small relative to the volatility of capital: In that case, the system gets trapped near $\eta = 0$ in the long run.

PROPOSITION III.2: *Denote by $\Lambda = (a - \underline{a})/\underline{q} + \underline{\delta} - \delta$ the risk premium that experts earn from capital at $\eta = 0$. As long as*

$$2(\rho - r)\sigma^2 < \Lambda^2 \tag{22}$$

the stationary density $d(\eta)$ exists and satisfies $d'(\eta^*) > 0$ and $d(\eta) \to \infty$ as $\eta \to 0$. If $2(\rho - r)\sigma^2 > \Lambda^2$ then the stationary density does not exist and in the long run η_t ends up in an arbitrarily small neighborhood of 0 with probability close to 1.

In our numerical examples, $\Lambda = 0.05$.

[16]The shape of the stationary distribution depends on the assumption that experts are able to sell capital to households. Fire sales lead to price volatility, because households have a lower valuation for capital, and slow recovery from depressed states. In contrast, in He and Krishnamurthy (2012) the stationary distribution is not U-shaped and recovery is fast. Effectively, experts get monopoly rents as only they can hold capital in crisis. Note also that the stationary distribution does not depend on amplification through prices. For example, Isohätälä, Milne and Roberston (2012) produce a U-shaped distribution in a model where agents recover slowly from a depressed region, as they scale down investments to reduce risk. In our model also, price effects are absent near $\eta = 0$, as capital is effectively priced at $\underline{q}$ by households.

Robustness of the Equilibrium Features

Our model exhibits stability in normal times, and strong amplification in crisis times, because the wealth distribution evolves *endogenously.* The steady state of the wealth distribution is determined by the relative rates of consumption of experts and households, as well as by the difference in returns that experts and households earn on their portfolios. Variable η_t reaches the stochastic steady state when experts accumulate enough wealth to absorb most shocks easily. At that point, competition among experts pushes up the price of capital and drives down the risk premia that experts earn. These factors encourage experts to consume their net worth instead of reinvesting it.

Near the stochastic steady state η^*, where experts become comfortable and risk premia come down, the price of capital is less responsive to shocks. Thus, amplification and endogenous risk are significantly lower near the steady state. In fact, in our risk-neutral model, the risk premium $-\sigma_t^\theta(\sigma + \sigma_t^q)$ and endogenous risk σ_t^q both drop to *zero* at η^*.

To confirm the robustness of these equilibrium features to risk neutrality, we solve a variation of our model, in which experts and households have *logarithmic* utilities with discount rates ρ and r. All other features of the model, including production technologies, financial frictions, and asset markets, are the same. Thus, equations (5) and (6) expressing the agents' return on capital are unchanged. The law of motion of η_t takes the same form as (14), except that the risk-free return dr_t is no longer constant.

Models with logarithmic utility are easy to solve because they lead to myopic optimal consumption and portfolio choice decisions. Specifically,

1) the optimal consumption rate of experts is $d\zeta_t = \rho\, dt$ (households, $r\, dt$), regardless of investment opportunities, and

2) the agents' optimal portfolio choice always results in the volatility of net worth equal to the Sharpe ratio of risky investment.

Proposition III.3, analogous to Proposition II.4 of the risk-neutral case, converts these equilibrium conditions into a first-order differential equation for $q(\eta)$ that can be solved to find equilibrium.

PROPOSITION III.3: *The equilibrium domain consists of subintervals* $[0, \eta^\psi)$, *where* $\psi(\eta) < 1$ *and* $[\eta^\psi, 1]$, *where* $\psi(\eta) = 1$. *Function* $q(\eta)$ *satisfies*

$$q(0) = \frac{\underline{a} - \iota(q(0))}{r} \quad \text{and} \quad q(\eta) = \frac{a - \iota(q(\eta))}{r(1-\eta) + \rho\eta} \quad \text{on} \quad [\eta^\psi, 1]. \tag{23}$$

The following procedure can be used to compute $\psi(\eta)$ *and* $q'(\eta)$ *from* $(\eta, q(\eta))$ *on* $(0, \eta^\psi)$.

1. Find ψ_t that satisfies

$$(r(1-\eta)+\rho\eta)q_t = \psi_t a + (1-\psi_t)\underline{a} - \iota(q_t) \tag{24}$$

2. Compute

$$\sigma + \sigma_t^q = \sqrt{\frac{(a-\underline{a})/q_t + \underline{\delta} - \delta}{\psi_t/\eta_t - (1-\psi_t)/(1-\eta_t)}} \quad \text{and} \quad q'(\eta) = \frac{q(\eta)}{\psi_t - \eta_t}\left(1 - \frac{\sigma}{\sigma + \sigma_t^q}\right). \tag{25}$$

The law of motion of η_t is given by

$$\sigma_t^\eta = \frac{\psi_t - \eta_t}{\eta_t}(\sigma + \sigma_t^q) \quad \text{and} \quad \mu_t^\eta = (\sigma_t^\eta)^2 + \frac{a - \iota(q_t)}{q_t} + (1-\psi_t)(\underline{\delta} - \delta) - \rho. \tag{26}$$

The proof is in Appendix C.[17]

The boundary conditions at the two endpoints 0 and η^ψ are sufficient to solve the first-order ordinary differential equation for $q(\eta)$ and determine the endogenous boundary η^ψ. Figure 5 shows a computed example for the same parameter values of ρ, r, a, $\underline{a}$, δ, $\underline{\delta}$, and κ as we used in Section II, and $\sigma = 5\%$ and 10%. The stochastic steady state η^* is now defined as the point η^* where $\mu_t^\eta = 0$. Certainly, the sharp result of the risk-neutral model that $\sigma_t^q = 0$ at the steady state η^* no longer holds exactly, as it is no longer true that $q'(\eta^*) = 0$. However, the dynamics are still characterized by temporary stability near the steady state η^* and increased volatility in crises below the steady state. Indeed, while endogenous risk is low near η^*, it spikes below η^ψ when experts start selling capital to households.[18]

These features arise because the wealth distribution depends endogenously on σ. As σ increases, risk premia rise, experts make more profit, and the steady state η^* shifts to the right into the region where experts are *less levered* (see the bottom right panel of Figure 5). This endogenous force stabilizes the steady state as exogenous risk increases. At the same time, point η^ψ where households start participating in capital markets also shifts to the right. Interestingly, the spike in volatility at point η^ψ is the highest when exogenous risk σ is the lowest.

[17]The equilibrium risk-free rate, dr_t, can be determined from the condition $E[dr_t^k - dr_t]/dt = \psi_t/\eta_t(\sigma + \sigma_t^q)^2$ (see the proof of Proposition III.3). Because $q(\eta)$ has a kink at η^ψ, we have

$$dr_t = \left(\frac{a-\iota(q_t)}{q_t} + \Phi(\iota(q_t)) - \delta + \mu_t^q + \sigma\sigma_t^q - \frac{\psi_t}{\eta_t}(\sigma + \sigma_t^q)^2\right) dt + \frac{1}{2}(\sigma_t^\eta)^2 \frac{q'(\eta_+^\psi) - q'(\eta_-^\psi)}{q(\eta^\psi)} dL_t,$$

where $\mu_t^q = (\mu_t^\eta q'(\eta) + \frac{1}{2}(\sigma_t^\eta)^2 q''(\eta))/q(\eta)$ for $\eta \neq \eta^\psi$ (and 0 at $\eta = \eta^\psi$), and L_t is the *local time* of η_t at η^ψ.

[18]The location of the steady state η^* above η^ψ depends on the assumption that ρ is not significantly larger than r. The jump in volatility at point η^ψ occurs because the price $q(\eta)$ has a kink at η^ψ, which occurs because of the mechanical relationship (24) between ψ_t and the market price in the log utility model. Technically, because of the kink, we have to write the risk-free return in the model in the form dr_t, and not $r_t\, dt$. Also, note that the market-clearing condition (24) implies that $q'(\eta) < 0$ when $\eta > \eta^\psi$.

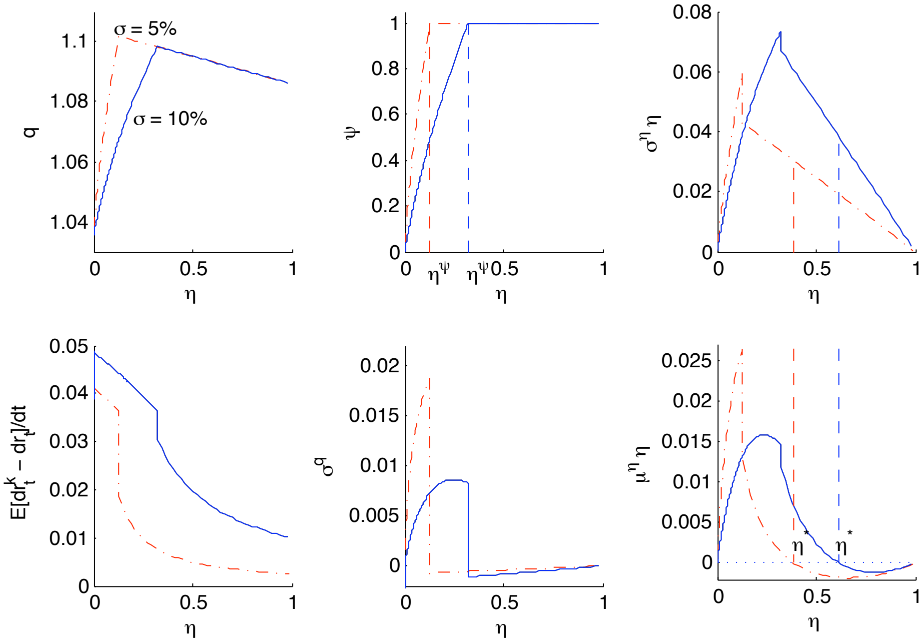

Figure 5. Equilibrium with logarithmic utility.

One may wonder about the robustness of the stationary distribution of η_t to the agents' preferences. Lemma C.1 in Appendix C shows that the hump of the stationary distribution near $\eta = 0$ exists only under some parameter values when agents have logarithmic utility (specifically, if $\rho > r + \Lambda$). Intuitively, because risk-averse experts are more cautious than risk-neutral experts, they use less leverage and the economy is less likely to get stuck near $\eta = 0$.

Correlation in Asset Prices and "Fat Tails"

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 effect into account in the recent housing price crash.[19]

We illustrate the correlation effects in our model by extending it to multiple types of capital. A similar argument about correlation has been proposed in He and Krishnamurthy (2012). Specifically, we can reinterpret equation (1),

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

[19] See "Efficiency and Beyond" in *The Economist*, July 16, 2009.

as the law of motion of fully diversified portfolios of capital held by experts, composed of specific types of capital $l \in [0,1]$ that follow

$$dk_t^l = (\Phi(\iota_t) - \delta)k_t^l\, dt + \sigma k_t^l\, dZ_t + \hat{\sigma} k_t^l\, dZ_t^l.$$

The diversifiable specific Brownian shocks dZ_t^l are uncorrelated with the aggregate shock dZ_t. Because of this, the specific shocks carry no risk premium, so all types of capital are trading at the same price q_t.

In equilibrium, the laws of motion of η_t and q_t are the same as in our basic model and depend only on the aggregate shocks dZ_t. The return on capital l is given by

$$dr_t^{k,l} = \left(\frac{a - \iota(q_t)}{q_t} + \Phi(\iota(q_t)) - \delta + \mu_t^q + \sigma\sigma_t^q\right) dt + (\sigma + \sigma_t^q)\, dZ_t + \hat{\sigma}\, dZ_t^l.$$

The correlation between assets l and l',

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

increases in the amount of endogenous risk σ_t^q. Near the steady state η^*, $\sigma_t^q = 0$ and so the correlation is $\sigma^2/(\sigma^2 + \hat{\sigma}^2)$. All the correlation near η^* is fundamental. Away from the steady state, some of the correlation becomes *endogenous*: It arises when both assets are held in portfolios of constrained agents.

The equilibrium patterns of volatility and correlation have implications for the pricing of derivatives. First, since the volatility rises in crisis times, option prices exhibit a "volatility smirk" in normal times. This observation is broadly consistent with empirical evidence (see, e.g., Bates (2000)). 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 ou- of-the-money put options are overpriced relative to at-the-money put options. 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)). 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.[20] Since data for crisis periods are limited, option prices provide valuable information about the market participants' implicit probability weights of extreme events and can be useful for model calibration.

[20]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.

IV. Volatility Paradox and the Kocherlakota Critique

Having established an equilibrium with instability away from the steady state, we now investigate whether small exogenous shocks generate significant endogenous risk. In this section we find some surprising answers. We uncover the *volatility paradox*: Endogenous risk does not go away as fundamental risk σ goes to 0. Surprisingly, the maximal level of endogenous risk σ_t^q has very low sensitivity to σ, and it may be slightly *increasing* as σ goes down. Thus, systemic risk exists even in low-volatility environments. If exogenous risk σ does not have a strong effect on maximal endogenous risk σ_t^q, then what does? The biggest determinant is liquidity: the ease with which the system can adjust to tightening financial constraints. For example, *market illiquidity*, which measures the difference between first-best value of capital $\bar{q}$ and its lowest liquidation value $\underline{q}$, plays an important role. If the liquidation value of capital $\underline{q}$ deteriorates, maximal endogenous risk σ^q in equilibrium rises significantly.

Because of the volatility paradox, the Kocherlakota critique does not apply to our setting. Kocherlakota (2000) and Cordoba and Ripoll (2004) argue amplification cannot be large in the settings of BGG and KM, when an isolated unanticipated shock knocks the log-linearized system away from the steady state. In those models, following a shock, the system is on a sure recovery path back to the deterministic steady state. When recovery is certain, the price has to fall very little to make it attractive for less productive households to buy capital, i.e., amplification is low. In contrast, in our model recovery is not certain after a shock. Rather the shock generates the forward-looking fear that the price may keep on falling all the way to $\underline{q}$. That is why *market illiquidity* of capital — which depends on the lower bound $\underline{q}$, to which prices may theoretically drop — is the key determinant of endogenous risk. Full dynamics are very different from local dynamics near the steady state. Backward induction from the boundary at $\eta_t = 0$ implies that, even as $\sigma \to 0$, endogenous risk does not disappear and the stochastic steady state of the system does not converge to the deterministic steady state. In fact, as $\sigma \to 0$, amplification (the ratio of endogenous to exogenous risk) in our model becomes infinite on almost the entire state space.

Volatility Paradox

One would expect endogenous risk to disappear as exogenous risk σ declines toward 0. In fact, it does not. As σ falls, the system becomes more prone to volatility spikes, maximal endogenous risk σ_t^q may rise, and the system still spends a significant fraction of time in crisis states where capital is misallocated. For our baseline example, Table 1 and Figure 6 demonstrate how various measures of instability persist, or even deteriorate, as σ falls.

As σ falls to 0, endogenous risk in crises persists, and amplification becomes infinite. While the level of risk at the steady state may decline, the buffer zone between the steady state η^* and crisis regimes gets thinner. At η^*, the volatility

σ	10%	5%	1%	0.2%	0.1%	as $\sigma \to 0$
Volatility of volatility $\sigma + \sigma_t^q$ near η^*	9.58%	13.75%	35.33%	127%	239%	increases
Maximal endogenous risk, $\max \sigma_t^q$	6.64%	6.69%	6.51%	6.36%	6.33%	persists
Maximal amplification, $\max \sigma_t^q/\sigma$	0.66	1.34	6.5	31.8	63.2	increases
Expected time to reach η^ψ from η^*	24.2	31.8	26.4	9.4	3.1	declines
% of time system spends when $\psi_t < 1$	14.16%	6.8%	3.35%	3.12%	3.14%	persists
% of time system spends when $\psi_t < 0.5$	6.74%	2.15%	0.47%	0.28%	0.31%.	persists

Table 1— Various measures of instability for different values of σ.

of volatility $\sigma + \sigma_t^q$ rises and the expected time to reach η^ψ falls. As a result, the time that the system spends in crisis states does not converge to zero as σ goes to 0. The shrinking distance between the steady state and crisis states has to be attributed to endogenous leverage. A decline in exogenous volatility encourages experts to increase leverage by reducing their net worth buffer through a more aggressive payout policy (see the right panel of Figure 6).[21]

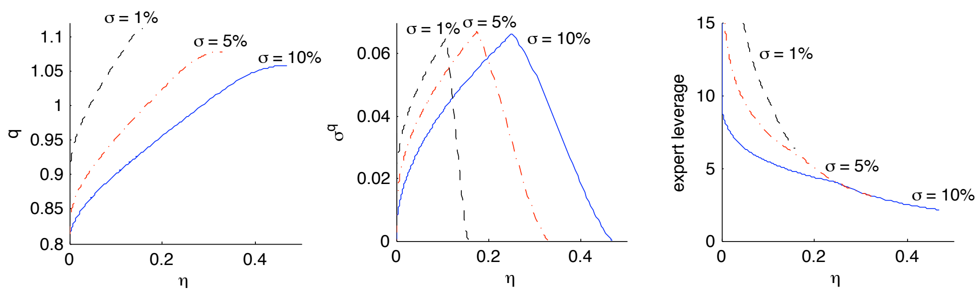

Figure 6. Equilibrium for three levels of exogenous risk, σ = 10%, 5%, and 1%.

These phenomena are fairly robust. For example, in Figure 5, in the version of our model with logarithmic utility, the maximal endogenous risk σ_t^q also rises as exogenous risk σ declines. As the following proposition shows analytically, both with risk-neutrality and logarithmic utility, in deep crisis states the equilibrium level of σ_t^η *increases* as σ declines.

PROPOSITION IV.1: *Both in the baseline risk-neutral model, and in the variation with logarithmic utility of Section III,*[22]

$$\sigma_t^\eta \to \frac{\Lambda}{\sigma} + O(\sigma), \tag{27}$$

[21]Risk shifting through asset substitution can lead to a similar effect. See Acharya and Viswanathan (2011).

[22]The term $O(\sigma)$ indicates that the difference $\lim_{\eta\to 0} \sigma_t^\eta - \Lambda/\sigma$ converges to 0 at the same rate as σ as $\sigma \to 0$.

as $\eta_t \to 0$, *where* $\Lambda = (a - \underline{a})/\underline{q} + \underline{\delta} - \delta$.

The "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, the system is prone to instabilities even when the level of aggregate risk is low.

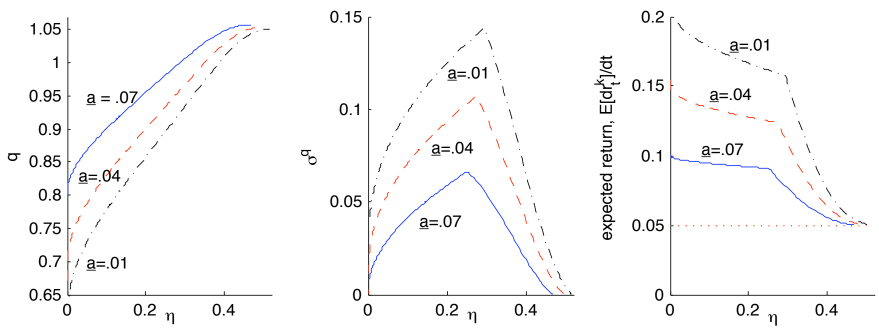

FIGURE 7. ENDOGENOUS RISK AND RISK PREMIA FOR $\underline{a} = 0.07, 0.04, 0.01$.

MARKET ILLIQUIDITY DETERMINES ENDOGENOUS RISK

In a fully dynamic model, agents assess the likelihood of quick recovery against the possibility that the system remains in depressed states for a long time. The *market liquidity* of capital — the difference between its first-best value $\bar{q}$ and its value in alternative uses $\underline{q}$ — matters a lot for the level of endogenous risk. In Figure 7, we vary the level of $\underline{a}$, which directly affects the level of $\underline{q}$. As $\underline{q}$ declines, maximal endogenous risk σ_t^q and risk premia in crisis states rise.

To sum up, small shocks can lead to unstable volatility dynamics in crisis states owing to uncertainty about the future path of the economy. This mechanism for amplification is significantly stronger than that of borrowing constraints, which has been explored in BGG and KM and was the subject of the Kocherlakota critique.

V. Borrowing Costs and Financial Innovations

In this section we explore the effect of borrowing costs on equilibrium dynamics. Borrowing costs provide an additional incentive for experts to reduce leverage, on top of the precautionary motives. We find that borrowing costs tend to stabilize the system: They lead to lower endogenous risk and lower crisis probability. Conversely, *financial innovations* that lower borrowing costs lead to higher leverage and may hurt system stability.

With borrowing costs, we are able to draw a more direct analogy to BGG, who justify borrowing costs through a costly state verification framework, and KM, who impose an exogenous borrowing constraint. Given high enough borrowing costs, the deterministic steady state of our system is no longer 0. We can discuss amplification near the deterministic steady state in a meaningful way and replicate the Kocherlakota critique. We also show once again that the Kocherlakota critique does not apply in our setting, as the dynamics near the stochastic steady state are different: It depends on the possibility of worst-case scenarios and does not rely on the assumption of certain recovery back to the steady state.

Idiosyncratic Risk and Borrowing Costs

Next, we explore the impact of default risk on financial stability by adding idiosyncratic jump risk to our baseline setting. With this risk, defaultable debt implies a credit spread between risky loans and the risk-free rate.[23]

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

$$dk_t = (\Phi(\iota_t) - \delta)k_t\, dt + \sigma k_t\, dZ_t + k_t\, dJ_t^i, \tag{28}$$

instead of (1). The new term dJ_t^i is a *compensated* (i.e., mean zero) Poisson process with intensity λ and jump distribution $F(y)$, $y \in [-1, 0]$ (if $y = -1$, the expert's entire capital is wiped out). Jumps are independent across experts and cancel out in the aggregate, so that total capital evolves according to the same equation as in the baseline model, (9). As in BGG, the jump distribution is *the same* for all experts and does not depend on the balance sheet size.

Like BGG, we adopt the costly state verification framework of Townsend (1979) to deal with default. If a sufficiently large jump arrives, such that the expert's net worth becomes negative, lenders trigger a costly verification procedure to make sure that capital was really destroyed by a shock and not stolen. Verification costs are proportional to the balance sheet size $q_t k_{t-}$.[24]

We can solve for the equilibrium by following the same *two* steps that we took in Section II. First, we extend conditions (H), (E), and (EK) to this setting. Second, we derive the law of motion of the state variable η_t that drives the system.

Step 1. Levered experts have to compensate lenders for imperfect recovery and deadweight losses of verification in the event of default. Both of these costs, $L(x_t)$ and $\Gamma(x_t)$ respectively per dollar borrowed, are increasing in leverage x_t (with $L(x) = \Gamma(x) = 0$ if $x \leq 1$). Function $L(x_t)$ depends on the intensity and distribution of jumps, and $\Gamma(x_t)$ depends in addition on the verification costs.

[23] An interesting variation of this model allows borrowing costs to depend on volatility. This leads to a value-at-risk (VaR) constraint as in Brunnermeier and Pedersen (2009) and Shin (2010). Recently, models with such a constraint have been explored by Phelan (2012) and Adrian and Boyarchenko (2012).

[24] The basic costly state verification framework, developed by Townsend (1979) and adopted by BGG, 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 $\tilde{y}$ distributed on the interval $[0, \bar{y}]$. The agent privately observes output $\tilde{y}$, but the principal can verify it at a cost. The optimal contract is a standard debt contract with some face value D. If the agent receives $\tilde{y} \geq D$, then he pays the principal D and there is no verification. The principal commits to verify if the agent reports that $\tilde{y} < D$, and receives the entire output.

The jump term adds dJ_t^i to the experts' return on capital dr_t^k in (5), but does not affect their expected return. Experts pay interest $r + L(x_t) + \Gamma(x_t)$ on debt, and so their net worth evolves according to

$$\frac{dn_t}{n_t} = x_t\, dr_t^k + (1 - x_t)\,(r + L(x_t) + \Gamma(x_t))\, dt - dc_t/n_t. \tag{29}$$

However, n_t cannot become negative. If a jump puts n_t into *negative* territory, debt holders absorb the loss so that

$$E[dn_t/n_t] = x_t E[dr_t^k] + (1 - x_t)\,(r + \Gamma(x_t))\, dt - dc_t/n_t. \tag{30}$$

Note the absence of $(1 - x_t)L(x_t)$, the expected loss rate of debt holders due to imperfect recovery. That is, because debt holders have to earn the expected return of r, they charge a higher interest rate such that, effectively, the experts bear the deadweight costs of verification, $\Gamma(x_t)$. In other words, experts' *expected* cost of borrowing is $r + \Gamma(x_t)$. Thus, the experts' Bellman equation (12) is transformed to

$$\rho = \mu_\theta + \max_x \left(xE[dr_t^k]/dt + (1 - x)(r + \Gamma(x)) + x\sigma_\theta(\sigma + \sigma_t^q) \right) \tag{EEK}$$

This equation replaces (E) and (EK) in Proposition II.2, and it implies (E) and (EK) if $\Gamma(x) = 0$, i.e., there are no verification costs. In equilibrium, $x_t = \psi_t/\eta_t$ should solve the maximization problem in (EEK). As before, $\theta_t \geq 1$ and experts consume only when $\theta_t = 1$. The household optimal portfolio choice condition (H) remains the same.

Step 2. Aggregating the experts' net worth, equation (30) implies that

$$dN_t = \psi_t q_t K_t\, dr_t^k - (\psi_t q_t K_t - N_t)(r + \Gamma(\psi_t/\eta_t))\, dt - dC_t.$$

With the extra term $\Gamma(\psi_t/\eta_t)$, an analogue of Lemma II.3 leads to the formula

$$\frac{d\eta_t}{\eta_t} = \frac{\psi_t - \eta_t}{\eta_t}(dr_t^k - r dt - \Gamma(\psi_t/\eta_t)dt - (\sigma + \sigma_t^q)^2 dt) + \frac{a - \iota(q_t)}{q_t} dt + (1 - \psi_t)(\underline{\delta} - \delta)dt - d\zeta_t. \tag{31}$$

Figure 8 illustrates how expected verification costs of the form $\Gamma(x) = \max\{\xi(x - 1), 0\}$, with $\xi = 0$, 0.01 and 0.02, affect equilibrium in the example of Section II. The effects of borrowing frictions on equilibrium dynamics may seem surprising at first. One may guess that these frictions, which make it harder for experts to get funding, particularly in downturns, cause amplification effects to become more severe.

In fact, the opposite is true: While borrowing frictions depress prices and investment, they actually lead to a *more stable* equilibrium. The amount of endogenous risk σ_t^q drops significantly as expert leverage decreases and prices in

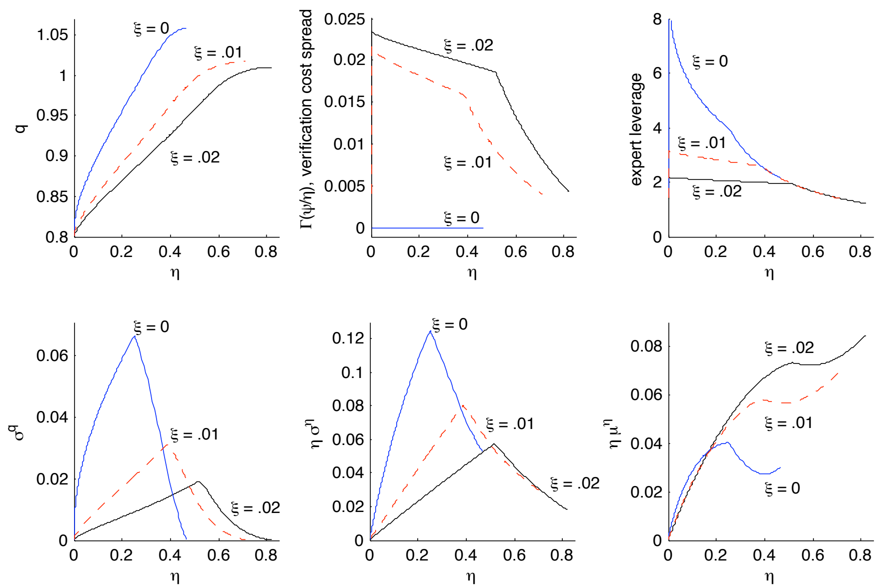

FIGURE 8. EQUILIBRIUM WITH VERIFICATION SPREAD $\Gamma(x) = \xi(x-1)$, $\xi = 0$, .01 AND .02.

booms become lower. Phelan (2012) has also recently obtained a related result that a stricter explicit leverage constraint leads to lower endogenous risk in crises in equilibrium.

Conversely, instruments that reduce borrowing costs make the equilibrium *less* stable.

DESTABILIZING FINANCIAL INNOVATION

Next, we explore the impact of financial innovations that allow experts to share risk better and, in particular, hedge idiosyncratic risks. These products can involve *securitization*, including pooling and tranching, credit default swaps, and various options and futures contracts. We find that risk sharing among experts reduces inefficiencies from idiosyncratic risk on one hand, but on the other hand emboldens them to maintain smaller net worth buffers and attain higher leverage. This leads to an increase of systemic risk. Ironically, tools intended for more efficient risk management at the individual level may make the system less stable overall.

Formally, assume that all shocks, including idiosyncratic jumps dJ_t^i, are observable and contractible among experts, but not between experts and households. Then experts can trade insurance contracts that cover jump losses dJ_t^i on expert i's capital. Experts can also trade contracts on the aggregate risk dZ_t.

PROPOSITION V.1: *If experts can contract on all shocks among each other, then the equilibrium in a setting with idiosyncratic shocks is equivalent to that in* *the baseline setting. Experts fully hedge their idiosyncratic risks, which carry the risk premium of zero.*

SKETCH OF PROOF:

Idiosyncratic risk of any expert i carries the risk premium of zero because it can be fully diversified among other experts. Given that, experts choose to fully insure their idiosyncratic risks, so their debt becomes risk-free. With borrowing frictions eliminated, the laws of motion of η_t and functions $q(\eta)$, $\theta(\eta)$, and $\psi(\eta)$ are the same as in the baseline setting with $\Gamma(x) = 0$. Contracts on aggregate risk among experts do not change the equilibrium, as they do not alter the total aggregate risk exposure of the expert sector.

When experts can trade contracts on idiosyncratic shocks, then they face the cost of borrowing of only r, and equilibrium dynamics end up being the same as in our baseline model. Thus, in the example of Figure 8, for any function $\Gamma(x)$ the equilibrium becomes transformed to that described by the solid plot, which corresponds to the parameter $\xi = 0$.

Instruments that help experts share risks eliminate the deadweight losses of costly state verification in this model. These instruments also lead to greater systemic risk, because experts endogenously increase leverage by lowering their net worth buffers. If instability harms the economy, e.g., due to (not yet modeled) spillovers to other sectors (such as the labor sector), then financial innovations may be detrimental to welfare. However, financial innovations are always welfare-enhancing if accompanied by regulations that encourage experts to maintain adequate net worth buffers (see Section VI).

The link between financial innovations and aggregate leverage has also been illustrated concurrently by Gennaioli, Shleifer and Vishny (2012), who build a two-period model in which agents ignore the possibility of certain bad events. In particular, they interpret securitization as one important form of risk-sharing.

Comparison with BGG and KM

Deterministic steady state η^0 is defined as the point at which the system remains stationary in equilibrium in an economy without shocks, i.e., when $\sigma = 0$. Point η^0 may be different from the stochastic steady state η^* even when σ is close to 0. That is, there may be a discontinuity at $\sigma = 0$. For example, in our baseline model without verification costs, $\eta^0 = 0$, i.e., the deterministic steady state is degenerate, but $\lim_{\sigma \to 0} \eta^* > 0$. Point η^* is determined taking into account that a sequence of bad shocks may plunge the economy into a depressed region. This can happen as a result of high endogenous risk, even when exogenous risk σ is low. In contrast, η^0 is determined as the balance point where earnings offset payouts in a risk-free economy.

To compare deterministic and stochastic steady states, we use the following proposition, which characterizes the location of η^0 when $\eta^0 > 0$.

PROPOSITION V.2: *Leverage* $x^0 = 1/\eta^0$ *at the non-degenerate* deterministic steady state *of the model with idiosyncratic shocks is characterized by equation*

$$\rho - r = x^0(x^0 - 1)\Gamma'(x^0) + \Gamma(x^0), \tag{32}$$

and the price of capital is characterized by

$$\max_\iota \frac{a-\iota}{q^0} + \Phi(\iota) - \delta = r + \Gamma(x^0) + (x^0 - 1)\Gamma'(x^0). \tag{33}$$

PROOF:

When $\sigma = 0$, then $\sigma_t^q = \sigma_t^\theta = 0$ and system dynamics are *deterministic*. Taking the first-order condition with respect to x in the Bellman equation (EEK), we get

$$E[dr_t^k]/dt = r + \Gamma(x) + (x-1)\Gamma'(x), \tag{34}$$

where $x = \psi/\eta$ in equilibrium. Furthermore, at η^0 we have $\mu_t^\theta = 0$, since it is an absorbing state of the system. Using (34) and $\mu_t^\theta = 0$, the Bellman equation (EEK) at $\eta = \eta^0$ implies (32). Finally, since $\mu^q = 0$ at η^0, the left-hand side of (33) is the expected return on capital, and so (34) implies (33).

Table 2 presents the location of stochastic and deterministic steady states for several model parameters.

	η^* for $\sigma = 10\%$	$\sigma = 5\%$	$\sigma = 1\%$	$\sigma = 0.1\%$	$\sigma = 0.01\%$	η^0
$\xi = 0$	0.4678	0.3291	0.1569	0.0994	0.0928	0
$\xi = 0.01$	0.7135	0.7086	0.7072	0.7071	0.7071	0.7071

TABLE 2— STOCHASTIC STEADY STATES VS. DETERMINISTIC STEADY STATE.

The location of the deterministic steady state depends on the borrowing costs. Higher borrowing costs lead experts to accumulate more net worth, to reach lower target leverage. Amplification near the deterministic steady state works as follows. An unanticipated negative shock increases the experts' cost of borrowing and makes it harder to hold capital. The price of capital in response to the shock has to drop sufficiently to generate enough demand for capital.

Kocherlakota (2000) argues that this mechanism cannot lead to large amplification. Since the economy recovers to the steady state for sure, and the price is known to rise to its original level, it takes only a small drop in price to make capital attractive to hold again.

In contrast, amplification can be large in the fully dynamic equilibrium, particularly when exogenous risk σ is low (see Table 1). The economy is not known to recover for sure: The price may drop all the way to $\underline{q}$. It is because of endogenous risk that the stochastic steady state η^* does not converge to 0 as $\sigma \to 0$ even when there are no verification costs (see Table 2). Experts keep net worth buffers against endogenous risk even when exogenous risk is low.[25]

VI. Efficiency and Macroprudential Policies

Systemic instability and endogenous risk, created by financial frictions, do not necessarily prescribe strict financial regulation. Making the system more stable might stifle economic growth. To study financial regulation, one has to conduct a welfare analysis. This section makes a first small step in this direction.

Within our model, the welfare of experts is $\eta_t \theta(\eta_t) q(\eta_t) K_t$ and that of households $(1 - \eta_t) q(\eta_t) K_t$. As discussed at the end of Section 3, the total welfare is less than the first best of $\bar{q} K_t$ (because of capital misallocation, under-investment, and consumption distortions). Endogenous risk during crises exacerbates many of these inefficiencies. In addition, regulators may be concerned about inefficiencies outside our model, such as spillovers from the financial system to the labor sector. If so, then the regulator may be concerned with the percentage of time the system spends in states of capital misallocation (see Table 1).

In this section we investigate the effects of policies on the equilibrium outcome. First, we show that there are policies that attain the first-best efficient outcome. These policies generally require large transfers to and from the financial system, or large open-market operations with financial assets. Second, we show that policies, in which the regulator assumes only the tail risk, can improve welfare significantly, particularly when exogenous risk is small and potential endogenous risk is large. Third, we argue that small policy mistakes can have a huge effect on the equilibrium outcome. Fourth, we investigate the effects of other natural policies — capital requirements and restrictions on dividends — and show that they may lead to unintended consequences.

Efficient Policies When Planner Can Control Consumption

A social planner can achieve the first-best outcome while respecting the same financing frictions with respect to equity issuance that individual experts face. To formalize this result, we define a set of constrained-feasible policies, under which

[25] Otherwise, there are only *cosmetic* differences between our model and those of BGG and KM, as many features of those models can be captured by various versions of our framework. Instead of introducing credit spreads, KM instead assume that the experts' leverage cannot exceed $\bar{x}$. This can be captured in our model by setting $\Gamma(x) = 0$ on $[0, \bar{x}]$ and ∞ on $(\bar{x}, \infty)$. This assumption leads to a deterministic steady state of $\eta^0 = 1/\bar{x}$, at which experts lever up to the constraint. Shocks have asymmetric effects near the deterministic steady state in KM: While negative shocks lead to amplification, positive shocks lead to payouts. This also happens in our baseline model. In contrast, the model of BGG is more similar to the version of our model with log utility (see Section III), as it reacts symmetrically to small shocks near the stochastic steady state. The difference in discount rates of experts and households in that model plays the same role as the exogenous exit rate of experts in BGG: it prevents the wealth share of the expert sector from converging to 1.

the central planner controls prices and the agents' consumption and investment choices, but treats all experts and households symmetrically.

Definition A symmetric constrained-feasible policy is described by stochastic processes on the filtered probability space defined by the Brownian motion $\{Z_t, t \geq 0\}$: The price process q_t, investment rates ι_t and $\underline{\iota}_t$, capital allocations $\psi_t K_t$ and $(1-\psi_t)K_t$, consumptions dC_t and $d\underline{C}_t$, and transfers $d\tau_t$ such that

(i) representative expert net worth $dN_t = d\tau_t + \psi_t K_t q_t \, dr_t^k - dC_t$ stays nonnegative,[26]

(ii) representative household net worth is defined by $\underline{N}_t = q_t K_t - N_t$, and

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

$$\frac{dC_t + d\underline{C}_t}{K_t} = (\psi_t(a-\iota_t)+(1-\psi_t)(\underline{a}-\underline{\iota}_t))dt, \quad \frac{dK_t}{K_t} = (\psi_t(\Phi(\iota_t)-\delta)+(1-\psi_t)(\Phi(\underline{\iota}_t)-\underline{\delta}))dt+\sigma dZ_t.$$

Note that since the sum of net worth equals the total wealth in the economy $q_t K_t$, aggregate transfers across both sectors are zero. The following proposition characterizes constrained-feasible policies that achieve the first-best allocation.

PROPOSITION VI.1: *Constrained-feasible policies that achieve a first-best outcome are those that satisfy $\psi_t = 1$ and $\iota_t = \iota(\bar{q})$ for all $t \geq 0$, and $dC_t = 0$ for all $t > 0$, although experts may consume positively at time 0, and transfers $d\tau_t$ are chosen to keep the net worth of experts nonnegative.*[27]

PROOF:

The policies outlined in Proposition VI.1 are constrained-feasible because the experts' net worth stays nonnegative. They attain first-best outcomes because experts consume *only* at time 0, since all capital is always allocated to experts and they are forced to invest at the first-best rate of $\iota_t = \iota(\bar{q})$.

Efficient Policies with Open-Market Operations

It turns out that it is possible to attain first best without controlling the agents' consumption and investment decisions, and even without direct transfers of wealth. A social planner can recapitalize experts by creating an insurance

[26] Because of transfers, without loss of generality we set the risk-free rate to zero.

[27] One may wonder whether the suggested policies preserve incentive compatibility. According to our microfoundation of balance sheets in Appendix A.A1, experts must retain full equity stakes in their projects because otherwise they would divert some of the capital and use it in another firm, while original outside equity holders suffer losses. Under any policy of Proposition VI.1, such a deviation would not enhance the expert's utility because the social planner controls the experts' consumption and sets it to zero for all $t > 0$. Even if experts could secretly *consume* diverted funds, the maximum amount that an expert would be able to divert under the policy of Proposition VI.1, before his net worth becomes negative, would be 0.

asset that experts can use to hedge risks. The idea comes from Brunnermeier and Sannikov (2012), where the insurance asset used to recapitalize intermediaries is the long-term bond.

Suppose that the aggregate value of the insurance asset P_t follows

$$dP_t = rP_t\,dt - \Sigma_t^P\,dZ_t + \Sigma_t^P\sigma_t^\theta\,dt - dD_t, \tag{35}$$

where $-\Sigma_t^P dZ_t$ is the aggregate risk of the asset and $\Sigma_t^P\sigma_t^\theta$ is the experts' insurance premium. The social planner can create the cash flows dD_t through open-market operations, i.e., by issuing or repurchasing the asset in exchange for output, or through dividends (if $dD_t > 0$). These cash flows can be financed by taxing or by selling the opposite end of the trade to households.[28]

As long as the process P_t satisfies the transversality condition $\lim_{t\to 0} E[e^{-rt}P_t] = 0$, P_t correctly reflects the experts' valuation of the asset with cash flows dD_t. Therefore, by controlling cash flows, the planner can endow the insurance asset with any risk profile Σ_t^P.

PROPOSITION VI.2: *If the planner sets $\Sigma_t^P = (\psi_t - \eta_t)(\sigma + \sigma_t^q)q_tK_t$, then the volatility of N_t is $\sigma+\sigma_t^q$ and the volatility of η_t is 0. In the Markov equilibrium, the dynamics are deterministic with $\sigma_t^q = \sigma_t^\theta = 0$. Experts hold all capital, consume their entire net worth at time 0, and maintain infinite leverage thereafter. The price of capital is $\bar{q}$.*

PROOF:

Experts' aggregate net worth N_t is exposed to risk $\psi_t(\sigma + \sigma_t^q)q_tK_t\,dZ_t$ from capital. After the insurance asset hedges some of this risk, $(\sigma + \sigma_t^q)N_t\,dZ_t$ is left. Ito's lemma implies that the volatility of η_t is 0, and $\sigma_t^q = \sigma_t^\theta = 0$ trivially. Because experts do not require any risk premium, they must hold all capital (as they earn a higher return than households do) and earn a return of r on capital. To maintain the transversality conditions, the price of capital must be $\bar{q}$.

The policy of Proposition VI.2 achieves the first-best efficient outcome because experts hold all capital, consume their entire net worth at time 0, and invest efficiently. The policy may be confusing owing to degeneracy, but it is possible to get arbitrarily close to first best with non-degenerate policies. Consider what happens if the regulator also imposes the constraint that any expert's portfolio allocation to capital x_t cannot exceed $\bar{x} > 1$. Then, with the policy of Proposition VI.2 in place, the equilibrium dynamics are deterministic. The deterministic and stochastic steady states coincide at $\eta^* = \eta^0 = 1/\bar{x}$. At that point, experts must be indifferent among all payout times, so they must earn the return of

$$\rho = \bar{x}\left(\frac{a - \iota(q^*)}{q^*} + \Phi(\iota(q^*)) - \delta\right) + (1 - \bar{x})r \tag{36}$$

[28]In fact, if $\sigma_t^\theta < 0$ and $\Sigma_t^P > 0$, then the regulator can make a profit on the trade by collecting the insurance premium from experts.

on their portfolios. Equation (36) determines the price level q^* at the steady state. As the regulator relaxes the leverage bound $\bar{x}$, the outcome converges to first best: η^* converges to 0, the return on capital converges to r, and q^* converges to $\bar{q}$.

Tail Risk Insurance

While the policy of Proposition VI.2 attains the first-best efficient outcome, it requires the planner to be involved in asset markets continuously and to a large extent. It turns out that when exogenous risk is low, but endogenous risk can be potentially high, then the planner can come close to efficiency by providing experts with only tail risk insurance.

For example, consider a policy that makes transfers $d\tau_t \geq 0$ to experts only when η_t hits a lower bound $\underline{\eta} > 0$, in such a way that the process η_t is reflecting at $\underline{\eta} > 0$. Transfers are proportional to the experts' net worths.[29] Figure 9 shows the effects of the policy, for $\underline{\eta} = 0.002$, on the example in Section 3 with $\sigma = 0.1\%$. The two left panels show that the policy raises the price of capital toward the first-best price of $\bar{q} = 1.2$ and lowers endogenous risk significantly. The two right panels show that while the cost of insurance to households is low, the policy significantly improves welfare. Also, the market value of tail risk insurance to experts is significantly higher than its cost to households. In addition, the policy prevents inefficient reallocation of capital to households (not shown).

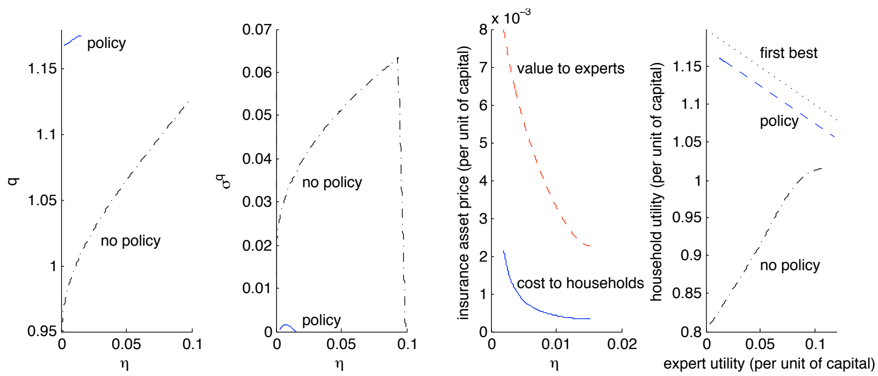

Figure 9. Equilibrium with and without insurance.

In sum, the policies considered thus far are redistributive: They redistribute wealth between experts and households based on information that these agents

[29]Under this policy, the equilibrium is characterized by the differential equations of Proposition II.4, but with boundary conditions $q'(\eta^*) = \theta'(\eta^*) = 0$, $\theta(\eta^*) = 1$, $q'(\underline{\eta}) = 0$, and $\theta'(\underline{\eta}) = \theta(\underline{\eta})/\underline{\eta}$. The last boundary condition is least obvious; it ensures that the experts' value function $\theta(\eta_t)n_t$ does not jump when the expert receives his share of the transfer $d\tau_t$ at $\eta_t = \underline{\eta}$.

cannot include directly in contracts that define the experts' capital structure. It is worthwhile to make a few remarks about these policies.

- *Effectiveness:* Tail risk insurance is effective (i.e., creates a large improvement in welfare at low cost) only if it targets endogenous risk in environments with low fundamental (exogenous) risk. If we replicate the example in Figure 9 with a higher level of exogenous risk $\sigma = 1\%$, the policy is a lot less effective: It raises $q(\eta^*)$ from 1.111 to only 1.117.
- *Policy mistakes:* The equilibrium can be unforgiving to small policy mistakes. For example, tail risk insurance would fail to reduce endogenous risk if it recapitalizes experts only partially when η_t falls below $\underline{\eta}$. If so, then the effect of the policy would be similar to a reduction in exogenous risk σ, and endogenous risk would persist as we observed in Section IV. Likewise, any other policy based on open-market operations (35) has to provide complete insurance in order to be effective.
- *Moral hazard:* Tail risk insurance does not create significant moral hazard if it makes transfers to experts proportionate to net worth. If so, most benefits go to cautious experts, those who accumulated more net worth and took lower leverage. The increase in leverage at the steady state due to the policy reflects primarily the significant reduction in endogenous risk, rather than the anticipation of insurance.

Other Policies

Many other policies have been proposed or implemented with the goal of improving financial stability. Some, such as equity infusions, asset purchases, or funding subsidies by the central bank (see Gertler and Kiyotaki (2011)), are aimed at recapitalizing financial institutions in crises. Others are aimed at controlling the overall risk within the financial system.

When considering policies, it is important to understand how they affect the entire equilibrium. For that purpose, our framework provides a useful laboratory to study the effects of policies on financial stability, on endogenous risk, on the amount of time the system spends in crises, and on welfare. With the help of our model, one can often identify unintended consequences of policies. While a comprehensive theoretical study of policies is beyond the scope of this paper, here we present several observations from our numerical experiments.[30]

- *Leverage constraints:* We considered a constraint that prohibits experts from taking on leverage greater than $\bar{x}(\eta)$. Generally, experts respond to

[30]Note that, in our setting, experts cannot simply recapitalize themselves by issuing extra equity. Such recapitalization goes against the agency microfoundations of balance sheets that we describe in Appendix A.A1. However, experts could, in principle, hedge aggregate risks. Di Tella (2012) shows that this would fully eliminate macroeconomic fluctuations, so instability in our model does depend on some aggregate risks being unhedgeable.

leverage constraints by accumulating net worth. On one hand, an increase in η^* can potentially stabilize the system and improve welfare. On the other hand, leverage constraints also can create many inefficiencies through capital misallocation and depressed prices that lead to underinvestment.

Our numerical experiments suggest that it is not easy to create welfare improvements through leverage constraints. While small improvements through carefully targeted policies are possible, crude bounds on leverage are often counterproductive and reduce welfare.

- *Restrictions on dividends:* We also considered policies that force experts to retain earnings and allow them to make payouts (proportionately to net worth) only when η_t hits a critical level $\bar{\eta}$. While a small restriction on payouts tends to improve welfare slightly within the model, large restrictions harm welfare. At the same time, there are a number of other consequences (desirable and undesirable.) As experts are forced to retain more net worth, the price of capital rises and may even become greater than $\bar{q}$ at the steady state. The marginal value of expert net worth $\theta(\eta)$ falls and becomes non-monotonic near $\bar{\eta}$. As a result, risk premia near $\bar{\eta}$ become negative.[31] Crisis episodes become less frequent, but more severe. The maximal endogenous risk in crises increases since prices have more room to fall.

VII. Conclusions

Events during the great liquidity and credit crunch of 2007-10 have highlighted the importance of financing frictions for macroeconomics. Unlike many existing papers in this field, our analysis is not restricted to local effects around the steady state. Importantly, we show that endogenous risk due to adverse feedback loops is significantly larger away from the steady state. This leads to nonlinearities: Small shocks keep the economy near the stable steady state, but large shocks put the economy in the unstable crisis regime characterized by liquidity spirals. The economy is prone to instability regardless of the level of aggregate risk because leverage and risk-taking are endogenous. As aggregate risk goes down, equilibrium leverage goes up, and amplification loops in crisis regimes become more severe: a volatility paradox. Owing to the volatility paradox, the Kocherlakota critique does not apply in our model: In fact, amplification in crises can be unbounded in low-volatility environments. In an environment with idiosyncratic and aggregate risks, equilibrium leverage also increases with diversification and with financial instruments that facilitate the hedging of idiosyncratic risks. Thus, paradoxically, tools designed to better manage risks may increase systemic risk.

[31] As η_t gets close to $\bar{\eta}$, $\theta(\eta_t) < 1$ and experts wish they could pay out funds. At $\bar{\eta}$, experts are able to pay out some funds, and so $\theta(\eta_t)$ increases. However, payouts are restricted to being just sufficient for η_t to reflect at $\bar{\eta}$. The equilibrium is still characterized by the equations of Proposition II.4, but with the boundary condition $\theta(\bar{\eta}) + \bar{\eta}\theta'(\bar{\eta}) = 1$ instead of $\theta'(\eta^*) = 0$ and $\theta(\eta^*) = 1$. The new boundary condition ensures that the experts' value functions drop by the amount of payout at point $\bar{\eta}$, when they are allowed to consume.

Policy interventions can make crisis episodes less likely, although many seemingly reasonable policies can harm welfare. Policies for crisis episodes alone, such as those aimed at recapitalizing the financial system, can increase risk-taking incentives ex ante. However, the effects of moral hazard are mitigated if these policies benefit strong institutions more than the weak. Surprisingly, simple restrictions on leverage may do more harm than good, as they bind only in downturns and may have little impact on behavior in booms. Policies encouraging financial institutions to retain earnings longer in booms do reduce the frequency of crises, but may raise endogenous risk (by stimulating asset prices in booms) and slow the recovery.

REFERENCES

Acharya, Viral, and S. Viswanathan. 2011. "Leverage, Moral Hazard and Liquidity." *The Journal of Finance*, 66(1): 99–138.

Adrian, Tobias, and Markus K. Brunnermeier. 2010. "CoVaR."

Adrian, Tobias, and Nina Boyarchenko. 2012. "Intermediary Leverage Cycles and Financial Stability." Working Paper, Federal Reserve Bank of New York.

Allen, Franklin, and Douglas Gale. 2007. *Understanding Financial Crises.* Oxford University Press.

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

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics.* Vol. 1, , ed. John B. Taylor and Michael Woodford, Chapter 21, 1341–1393. Elsevier.

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

Bhattacharya, Sudipto, and Douglas Gale. 1987. "Preference Shocks, Liquidity, and Central Bank Policy." In *New Approaches to Monetary Economics.* , ed. William A. Barnett and Kenneth J. Singleton, 69–88. Cambridge University Press.

Biais, Bruno, Thomas Mariotti, Guillaume Plantin, and Jean-Charles Rochet. 2007. "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *The Review of Economic Studies*, 74(2): 345–390.

Bolton, Patrick, and David S. Scharfstein. 1990. "A Theory of Predation Based on Agency Problems in Financial Contracting." *American Economic Review*, 80(1): 93–106.

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

Brunnermeier, Markus K., and Yuliy Sannikov. 2012. "The I Theory of Money." Working Paper, Princeton University.

Brunnermeier, Markus K., Gary Gorton, and Arvind Krishnamurthy. 2013. "Liquidity Mismatch." In *Risk Topography: Systemic Risk and Macro Modeling.* , ed. Markus K. Brunnermeier and Arvind Krishnamurthy. Chicago University Press.

Brunnermeier, Markus K., Thomas Eisenbach, and Yuliy Sannikov. 2012. "Macroeconomics with Financial Frictions: A Survey." In *Advances in Economics and Econometrics: Tenth World Congress of the Econometric Society.* Vol. 2, , ed. Daron Acemoglu, Manuel Arellano and Eddie Dekel, 3–94. Cambridge University Press.

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

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

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

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

Christiano, Lawrence, Roberto Motto, and Massimo Rostagno. 2007. "Shocks, Structures or Monetary Policies? The Euro Area and US After 2001." *Journal of Economic Dynamics and Control*, 32(8): 2476–2506.

Cordoba, Juan-Carlos, and Marla Ripoll. 2004. "Credit Cycles Redux." *International Economic Review*, 45(4): 1011–1046.

Curdia, Vasco, and Michael Woodford. 2010. "Credit Spreads and Monetary Policy." *Journal of Money, Credit & Banking*, 42(S1): 47–74.

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

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

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

Di Tella, Sebastian. 2012. "Uncertainty Shocks and Balance Sheet Recessions." Working Paper, MIT.

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

Duffie, Darrel. 2010. *Dynamic Asset Pricing Theory.* Princeton University Press.

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

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

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

Geanakoplos, John. 1997. "Promises Promises." In *The Economy as an Evolving Complex System II.* , ed. W. B. Arthur, S. Durlauf and D. Lane, 285–320. Addison-Wesley.

Geanakoplos, John. 2003. "Liquidity, Defaults, and Crashes." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Vol. 2 of *Econometric Society Monographs*, 170–205. Cambridge University Press.

Geanakoplos, John, and Heraklis 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.* Vol. 3, , ed. Walter P. Heller, Ross M. Starr and David A. Starrett, 65–95. Cambridge University Press.

Gennaioli, Nicola, Andrei Shleifer, and Robert W. Vishny. 2012. "Neglected Risks, Financial Innovation and Financial Fragility." *Journal of Financial Economics*, 104: 452 – 468.

Gertler, Mark, and Nobuhiro Kiyotaki. 2011. "Financial Intermediation and Credit Policy in Business Cycle Analysis." In *Handbook of Monetary Economics.* , ed. Michael Woodford and Benjamin M. Friedman. North Holland Elsevier.

Gertler, Mark, and Peter Karadi. 2011. "A Model of Unconventional Monetary Policy." *Journal of Monetary Economics*, 58(1): 17–34.

Ghosh, Arka P. 2010. "Backward and Forward Equations for Diffusion Processes." In *Wiley Encyclopedia of Operations Research and Management Science (EORMS).* , ed. James J. Cochran, Louis A. Cox, Pinar Keskinocak, Jeffrey P. Kharoufeh and J. Cole Smith. Wiley.

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

Harrison, Michael J., and David M. Kreps. 1979. "Martinglaes and Arbitrage in Multiperiod Securities Markets." *Journal of Economic Theory*, 20: 381–408.

He, Zhiguo, and Arvind Krishnamurthy. 2012. "A Model of Capital and Crises." *Review of Economic Studies*, 79(2): 735–777.

He, Zhiguo, and Arvind Krishnamurthy. 2013. "Intermediary Asset Prices." *American Economic Review*, 103(2): 1–43.

He, Zhiguo, and Wei Xiong. 2012. "Dynamic Debt Runs." *Review of Financial Studies*, 25: 1799–1843.

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

Isohätälä, Jukka, Alistair Milne, and Donald Roberston. 2012. "A Model of Investment Subject to Financing Constraints." Working Paper.

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

Jeanne, Olivier, and Anton Korinek. 2010. "Managing Credit Booms and Busts: A Pigouvian Taxation Approach." Working Paper, University of Maryland.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3(4): 305 – 360.

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

Kindleberger, Charles. 1993. *A Financial History of Western Europe.* . 2nd ed., Oxford University Press.

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

Kocherlakota, Narayana R. 2000. "Creating Business Cycles Through Credit Constraints." *Federal Reserve Bank of Minneapolis Quarterly Review*, 24(3): 2–10.

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

Meh, Césaire, and Kevin Moran. 2010. "The Role of Bank Capital in the Propagation of Shocks." *Journal of Economic Dynamics & Control*, 34(3): 555–576.

Mendoza, Enrique G. 2010. "Sudden Stops, Financial Crisis, and Leverage." *American Economic Review*, 100: 1941–1966.

Mendoza, Enrique G., and Katherine A. Smith. 2006. "Quantitative Implications of a Debt-Deflation Theory of Sudden Stops and Asset Prices." *Journal of International Economics*, 70(1): 82–114.

Minsky, Hyman. 1986. *Stabilizing an Unstable Economy.* Yale University Press.

Myerson, Roger B. 2010. "A Model of Moral Hazard Credit Cycles." Working Paper, University of Chicago.

Phelan, Gregory. 2012. "Financial Intermediation, Leverage, and Macroeconomic Instability." Working Paper, Yale University.

Rampini, Adriano, and S. Viswanathan. 2010. "Collateral, Risk Management, and the Distribution of Debt Capacity." *Journal of Finance*, 65: 2293–2322.

Sannikov, Yuliy. 2012. "Dynamic Security Design and Corporate Financing." In *Handbook of the Economics of Finance.* Vol. 2. Elsevier.

Scheinkman, Jose A., and Laurence Weiss. 1986. "Borrowing Constraints and Aggregate Economic Activity." *Econometrica*, 54(1): 23–45.

Shin, Hyun. 2010. *Risk and Liquidity.* Oxford University Press.

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

Shleifer, Andrei, and Robert W. Vishny. 1997. "The Limits of Arbitrage." *The Journal of Finance*, 52(1): 35–55.

Shleifer, Andrei, and Robert W. Vishny. 2010. "Unstable Banking." *Journal of Financial Economics*, 97(3): 306–318.

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

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

A Macroeconomic Model with a Financial Sector
Markus K. Brunnermeier and Yuliy Sannikov
Online Appendix

A1. Microfoundation of Balance Sheets and Intermediation

This section describes the connection between balance sheets in our model and agency problems. Extensive corporate finance literature (see Townsend (1979), Bolton and Scharfstein (1990), DeMarzo and Sannikov (2006), Biais et al. (2007), or Sannikov (2012) for a survey of these models) suggests that agency frictions increase when the agent's net worth falls. In a macroeconomic setting, this logic points to the aggregate net worth of end borrowers, as well as that of intermediaries.

Incentive provision requires the agent to have some "skin in the game" in the projects he manages. When projects are risky, it follows that the agent must absorb some of project risk through net worth. Some of the risks may be identified and hedged, reducing the agent's risk exposure. However, whenever some aggregate risk exposures of constrained agents cannot be hedged, macroeconomic fluctuations due to financial frictions arise, as these residual risks have aggregate impact on the net-worth-constrained agents.

Our baseline model assumes the simplest form of balance sheets, in which constrained agents (experts) absorb all risk and issue just risk-free debt. Qualitatively, however, our results still hold if experts can issue some outside equity and even hedge some of their risks, as long as they *cannot* hedge *all* the risks. Quantitatively, the assumption regarding equity issuance matters: If experts can issue more equity or hedge more risks, then they can operate efficiently with much lower net worths. This does not necessarily lead to a more stable system because, as we saw in Section 5, the steady state in our model is endogenous. Agents who can function with lower wealth accumulate lower net worth buffers. Thus, we expect that our baseline model with simple balance sheets captures many characteristics of equilibria of more general models.

To illustrate the connection between balance sheets and agency models, first, we discuss the agency problem with direct lending from investors to a single agent. Second, we illustrate agency problems that arise with intermediaries. In this case, the net worth of intermediaries matters as well. At the end of this section, we discuss contracting with idiosyncratic jump risk, which is relevant for Section 5.

Agency Frictions between an Expert and Households

Assume that experts are able to divert capital returns at rate $b_t \in [0,\infty)$. Diversion is inefficient: Of the funds b_t diverted, an expert is able to recover only a portion $h(b_t) \in [0, b_t]$, where $h(0) = 0$, $h' \leq 1$, $h'' \leq 0$. Net of diverted funds, capital generates the return of

$$dr_t^k - b_t\,dt$$

and the expert receives an income flow of $h(b_t)\,dt$.

If capital is partially financed by outside equity held by households, then households receive the return of

$$dr_t^k - b_t\,dt - f_t\,dt,$$

where f_t is the fee paid to the expert. When the expert holds a fraction Δ_t of equity, then per dollar invested in capital he gets

$$\Delta_t(dr_t^k - b_t\,dt) + (1-\Delta_t)f_t\,dt + h(b_t)\,dt.$$

The incentives with respect to diversion are summarized by the first-order condition $\Delta_t = h'(b_t)$, which leads to a weakly decreasing function $b(\Delta)$ with $b(1) = 0$. In equilibrium, the fee f_t is chosen so that household investors get the expected required return of r on their investment, i.e., $f_t = E[dr_t^k]/\,dt - b_t - r$. As a result, the return on the expert's equity stake in capital (including the benefits of diversion) is

$$\frac{\Delta_t(dr_t^k - b_t\,dt) + (1-\Delta_t)f_t\,dt + h(b_t)\,dt}{\Delta_t} = \frac{E[dr_t^k] - r\,dt}{\Delta_t} + r\,dt + (\sigma+\sigma_t^q)dZ_t - \frac{b(\Delta_t) - h(b(\Delta_t))}{\Delta_t}dt,$$

where $\frac{b(\Delta_t)-h(b(\Delta_t))}{\Delta_t}$ is the deadweight loss rate due to the agency problem. The law of motion of the expert's net worth in this setting is of the form

(A1)
$$\frac{dn_t}{n_t} = x_t\left(\frac{E[dr_t^k] - r\,dt}{\Delta_t} + r\,dt + (\sigma+\sigma_t^q)\,dZ_t - \frac{b(\Delta_t) - h(b(\Delta_t))}{\Delta_t}\,dt\right) + (1-x_t)r\,dt - d\zeta_t,$$

where x_t is the portfolio allocation to inside equity and $d\zeta_t$ is the expert's consumption rate. It is convenient to view equation (A1) as capturing the issuance of equity and risk-free debt. However, it is possible to reinterpret this capital structure in many other ways, since securities such as risky debt can be replicated by continuous trading in the firm's stock and risk-free debt.

Equation (EK) generalizes to

$$\max_{\Delta} \frac{E[dr_t^k]/dt - r - (b(\Delta) - h(b(\Delta)))}{\Delta} = -\sigma_t^\theta(\sigma + \sigma_t^q)$$

and determines optimal equity issuance. Our results suggest that, as risk premia $E[dr_t^k]/dt - r$ rise in downturns, the experts' equity retention Δ decreases and deadweight losses increase.[32]

Our baseline setting is a special case of this formulation, in which there are no costs to the diversion of funds, i.e., $h(b) = b$ for all $b \geq 0$. In this case, the agency problem can be solved only by setting $\Delta = 1$, i.e., experts can finance

[32] One natural way to interpret this is through a capital structure that involves risky debt, as it becomes riskier (more equity-like) after experts suffer losses. Many agency problems become worse when experts are "under water."

themselves only through risk-free debt. Our analysis can be generalized easily, but for expositional purposes we keep our baseline model as simple as possible.

We would like to be clear about our assumptions regarding the space of acceptable contracts, which specify how observable cash flows are divided between the expert and household investors. We make the following two restrictions on the contracting space:

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, and

B lockups are not allowed; at any moment of time, any party can break the contractual relationship. The value of assets is divided among the parties the same way, independently of who breaks the relationship.

Condition B simplifies analysis, as it allows us to focus only on expert net worth rather than a summary of the expert's individual past performance history. It assumes 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 condition prevents commitment to long-term contracts, such as in the setting of Myerson (2010). However, in many settings this restriction alone does not rule out optimal contracts: Fudenberg, Holmström and Milgrom (1990) show that it is possible to implement the optimal long-term contract through short-term contracts with continuous marking-to-market.

Condition A requires that contracts have to be written on the total return of capital and that innovations in k_t, q_t, or the aggregate risk dZ_t cannot be hedged separately. This assumption is clearly restrictive, but it creates a convenient and simple way to capture important phenomena that we observe in practice. Specifically, condition A creates an amplification channel, in which market prices affect the agents' net worth, and is consistent with the models of Kiyotaki and Moore (1997) and Bernanke, Gertler and Gilchrist (1999). Informally, 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 differ across projects, depend on the expert's private information, and be manipulable, e.g., through underinvestment.

More generally, we could assume that aggregate shocks dZ_t to the experts' balance sheets can be hedged partially. As long as it is impossible to design a perfect hedge and to perfectly share all aggregate risks with households, the model will generate economic fluctuations driven by the shocks to the net worth of the constrained agents. Thus, to generate economic fluctuations, we make assumptions that would otherwise allow agents to write optimal contracts, but place restrictions on hedging. Experts can still choose their risk exposure Δ_t, but cannot hedge aggregate shocks dZ_t.

Intermediary Sector

It is possible to reinterpret our model to discuss the capitalization of intermediaries as well as end borrowers.

One natural model of intermediation involves a double moral-hazard problem motivated by Holmström and Tirole (1997). Let us, like Meh and Moran (2010) separate experts into two classes of agents: entrepreneurs who manage capital under the productive technology, and intermediaries who can channel funds from households to entrepreneurs. Through costly monitoring actions that are unobservable by outside investors, intermediaries are able to reduce the benefits that entrepreneurs get from the diversion of funds. Specifically, the entrepreneurs' marginal benefit of fund diversion $\frac{\partial}{\partial b}h(b_t|m_t)$ is continuously decreasing with the proportional cost of monitoring $m_t \geq 0$, i.e. $\frac{\partial^2}{\partial b \partial m}h(b_t|m_t) < 0$. Thus, for a fixed equity stake Δ_t of the entrepreneur, higher monitoring intensity m_t leads to a lower diversion rate $b_t = b(\Delta_t|m_t)$. Assuming that $\frac{\partial^2}{\partial b^2}h(\cdot|m_t) < 0$, the entrepreneur's optimal diversion rate b_t is uniquely determined by the first-order condition $\frac{\partial}{\partial b}h(b_t|m_t) = \Delta_t$ and is continuously decreasing in Δ_t.

Intermediaries have no incentives to exert costly monitoring effort unless they themselves have a stake in the entrepreneur's project. An intermediary who holds a fraction Δ_t^I of the entrepreneur's equity optimally chooses the monitoring intensity m_t that solves

$$\min_m \; \Delta_t^I b(\Delta_t|m) + m.$$

The solution to this problem determines how the rates of monitoring $m(\Delta_t, \Delta_t^I)$ and cash flow diversion $b(\Delta_t, \Delta_t^I)$ depend on the allocations of equity to the entrepreneur and the intermediary.

By reducing the entrepreneurs' agency problem through monitoring, intermediaries are able to increase the amount of financing available to entrepreneurs. However, intermediation itself requires risk-taking, as the intermediaries need to absorb the risk in their equity stake Δ_t^I through their net worth. Thus, the aggregate net worth of intermediaries becomes related to the amount of financing available to entrepreneurs. Figure A1 depicts the interlinked balance sheets of entrepreneurs, intermediaries, and households. Fraction $\Delta_t + \Delta_t^I$ of entrepreneur risk gets absorbed by the entrepreneur and intermediary net worths, while fraction $1 - \Delta_t - \Delta_t^I$ is held by households.

The marginal values of entrepreneur and intermediary net worths, θ_t and θ_t^I, can easily differ in this economy. If so, then the capital-pricing equation (EK) generalizes to

$$\max_{\Delta, \Delta^I} E[dr_t^k]/dt - r - (b(\Delta, \Delta^I) - h(b(\Delta, \Delta^I))) - m(\Delta, \Delta^I) + (\Delta \sigma_t^\theta + \Delta^I \sigma_t^{\theta, I})(\sigma + \sigma_t^q) = 0.$$

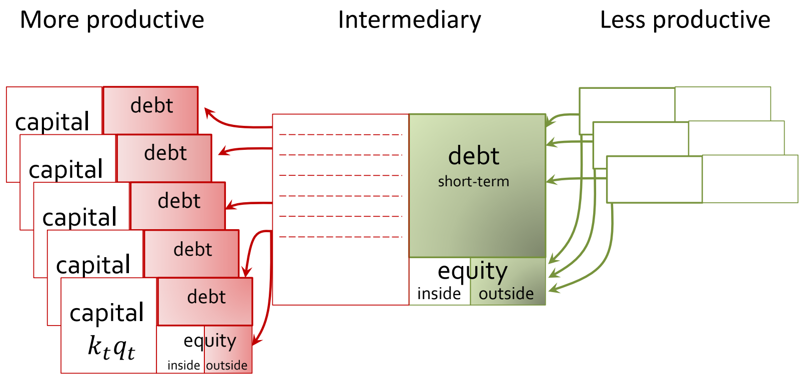

Figure A1. Balance sheet structures of entrepreneurs and financial intermediaries.

Equilibrium dynamics in this economy depend on two state variables, the shares of net worth that belong to the entrepreneurs η_t and intermediaries η_t^I. Generally, these are imperfect substitutes, as intermediaries can reduce the entrepreneurs' required risk exposure by taking on risk and monitoring. However, several special cases can be reduced to a single state variable. For example, if entrepreneurs and intermediaries can write contracts on aggregate shocks among themselves (but not with households), then the two groups of agents have identical risk premia (i.e., $\sigma_t^\theta = \sigma_t^{\theta,I}$) and the sum $\eta_t + \eta_t^I$ determines the equilibrium dynamics.

Contracting with Idiosyncratic Losses and Costly State Verification

Next, we discuss contracting in an environment of Section 5, where experts may suffer idiosyncratic loss shocks. For simplicity, we focus on the simplest form of the agency problem without intermediaries, in which $h(b) = b$ for all $b \geq 0$. As discussed earlier in the Appendix, in our *baseline* model this assumption leads to a simple capital structure, in which experts can borrow only through risk-free debt.

Assume, as in Section 5, that, in the absence of benefit extraction, capital managed by expert $i \in \mathbb{I}$ 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 *compensated* loss process with intensity λ and jump distribution $F(y)$, $y \in [-1, 0]$. Then, in the absence of jumps, J_t^i has a positive drift of

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

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 changes 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 dB_t units of capital. The jumps in B_t are bounded by k_{t-}, the total amount of capital under the entrepreneur's management just before time t.

Unlike in our earlier specification of the agency problem, in which the entrepreneur's *rate* of benefit extraction $b_t = \frac{dB_t}{(q_t k_t)\, dt}$ must be finite, now the entrepreneur can also extract benefits discontinuously, including in quantities that reduce the value of capital under management below the value of debt.

We assume a verification technology that can be employed in the event of discrete drops in capital. In particular, if a verification 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 the entrepreneur's benefit extraction at time t and in what amount,

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

(iii) pay a cost of $(q_t k_{t-}) c(dJ_t^i)$, that is proportional to the value of the investment prior to verification[33].

If verification reveals that the drop in capital at time t was partially caused by benefit extraction, i.e., $k_{t-}(1 + dJ_t^i) > k_t$, then the entrepreneur cannot extract any benefits, as diverted capital $k_{t-}(1 + dJ_t^i) - k_t$ is returned to the investors.

We maintain the same assumptions as before about the form of the contract in the absence of verification, 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 sudden drop in the market value of the expert's assets $q_t k_t$ triggers a verification action. In this event, the contract specifies how the remaining assets, net of verification costs, are divided among the contracting parties conditional on the amount of capital that was diverted at time t. We assume that the monitoring action is not randomized, i.e., it is completely determined by the asset value history.

PROPOSITION A.1: *With idiosyncratic jump risk, it is optimal to trigger verification only in the event that the market value of the expert's assets $q_t k_t$ falls*

[33] The assumption that the verification cost depends only on the amount of capital recovered, regardless of the diverted amount, is without loss of generality since on the equilibrium path the entrepreneur does not divert funds.

below the value of debt. In the event of verification, it is optimal for debt holders to receive the value of the remaining assets net of verification costs.

PROOF:

Because jumps are idiosyncratic, they carry no risk premium. Therefore, it is better to deter fund diversion that does not bankrupt the expert by requiring him to absorb jump risk through equity rather than triggering costly state verification (which leads to a deadweight loss). However, verification is required to deter the expert from diverting more funds than his net worth at a single moment of time.

The division of value between debt holders and the expert in the event of verification matters for the expert's incentives only if it is in fact revealed that the expert diverted cash. If no cash was diverted (i.e. it is clear that the loss was caused by an exogenous jump), the division of value between debt holders and the expert can be arbitrary (as long as the expected return of debt holders, net of verification costs, is r) since idiosyncratic jump risk carries no risk premium. Without loss of generality we can assume that debt holders receive the entire remaining value in case of verification.

Proposition A.1 implies that with idiosyncratic jump risk, debt is no longer risk-free.

Stationary Distribution and Time to Reach

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

$$dX_t = \mu^x(X_t)\,dt + \sigma^x(X_t)\,dZ_t. \tag{B1}$$

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 Kolmogorov forward equation (see, e.g., Ghosh (2010)):

$$\frac{\partial}{\partial t} d\,(x,t) = -\frac{\partial}{\partial x}\left(\mu^x\,(x)\,d\,(x,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big(\sigma^x\,(x)^2\,d\,(x,t)\Big)\,. \tag{B2}$$

A *stationary density* is a function that solves (B2) and does not change with time, i.e., $\frac{\partial d(x,t)}{\partial t} = 0$ on the left-hand side of (B2). If so, then integration over x yields the first-order ordinary differential equation

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

where the constant of integration F is the "flow" of the density in the positive direction. If one of the endpoints of the interval $[x_L, x_R]$ is reflecting (as η^* in our model), then the flow is $F = 0$.

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

$$D'(x) = 2\frac{\mu^x(x)}{\sigma^x(x)^2}D(x). \tag{B3}$$

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

With an absorbing boundary, the process X_t eventually ends up absorbed (and so the stationary distribution is degenerate) unless the law of motion (B1) prevents X_t from hitting the absorbing boundary with probability one. A non-degenerate stationary density, with an absorbing boundary at x_L, exists 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 $\mu_t^x(x)$ is positive near x_L (i.e., it pushes X_t away from the boundary x_L) and strong enough working against the volatility that may move X_t toward 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 $\int_0^{x_0} \frac{2\mu^x(x)}{\sigma^x(x)^2}\, dx = \int_0^{x_0} \frac{2\mu}{\sigma^2 x}\, dx = \infty$.

The following proposition characterizes the expected amount of time it takes to reach any point $x \leq x_R$ starting from x_R.

PROPOSITION B.1: *Suppose that X_t follows (B1) and x_R is a reflecting boundary. Then the expected amount of time $g(x)$ that it takes to reach $x \leq x_R$ from x_R solves equation*

$$1 - g'(x)\mu^x(x) - \frac{\sigma^x(x)^2}{2}g''(x) = 0 \tag{B4}$$

with boundary conditions $g(x_R) = 0$ and $g'(x_R) = 0$.

PROOF:

Denote by $f_{x_0}(y)$ the expected amount of time it takes to reach a point x_0 starting from $y \geq x_0$. Then, to reach x_0 from x_R (expected time $f_{x_0}(x_R) = g(x_0)$), one has to reach $x \in (x_0, x_R)$ first (expected time $g(x)$) and then reach x_0 from x (additional expected time $f_{x_0}(x)$). Therefore,[34]

$$g(x) = g(x_0) - f_{x_0}(x). \tag{B5}$$

[34] Equation (B5) implies that, in expectation to reach any point x_0 starting from $x > x_0$, it takes time $g(x_0) - g(x)$.

Since $t + f_{x_0}(X_t)$ is a martingale, it follows that f_{x_0} satisfies the ordinary differential equation

$$1 + f'_{x_0}(x)\mu^x(x) + \frac{\sigma^x(x)^2}{2} f''_{x_0}(x) = 0.$$

Since $g'(x) = -f'_{x_0}(x)$ and $g''(x) = -f''_{x_0}(x)$, it follows that g must satisfy (B4). It takes no time to reach x_R from x_R, so $g(x_R) = 0$. Furthermore, since x_R is a reflecting barrier, due to volatility, $g'(x_R) = 0$.

Proofs

PROOF OF LEMMA II.1:

Let us show that if the process θ_t satisfies (11) and the transversality condition holds, then θ_t represents the expert's continuation payoff, i.e., satisfies (10). Consider the process

$$\Theta_t = \int_0^t e^{-\rho s} n_s \, d\zeta_s + e^{-\rho t}\theta_t n_t.$$

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

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

If (11) holds, then $E[d\Theta_t] = 0$, so Θ_t is a martingale under the strategy $\{x_t, d\zeta_t\}$. Therefore,

$$\theta_0 n_0 = \Theta_0 = E[\Theta_t] = E\left[\int_0^t e^{-\rho s} n_s \, d\zeta_s\right] + E\left[e^{-\rho t}\theta_t n_t\right].$$

Taking the limit $t \to \infty$ and using the transversality condition, we find that

$$\theta_0 n_0 = E\left[\int_0^\infty e^{-\rho s} n_s \, d\zeta_s\right],$$

and the same calculation can be done for any other time t instead of 0.

Conversely, if θ_t satisfies (10), then Θ_t is a martingale since

$$\Theta_t = E_t\left[\int_0^\infty e^{-\rho s} n_s \, d\zeta_s\right].$$

Therefore, the drift of Θ_t must be zero, and so (11) holds.

Next, let us show that the strategy $\{x_t, d\zeta_t\}$ is optimal if and only if the Bellman equation (12) holds. Under any alternative strategy $\{\hat{x}_t, d\hat{\zeta}_t\}$, define

$$\hat{\Theta}_t = \int_0^t e^{-\rho s} n_s \, d\hat{\zeta}_s + e^{-\rho t}\theta_t n_t, \quad \text{so that} \quad d\hat{\Theta}_t = e^{-\rho t}(n_t \, d\hat{\zeta}_t - \rho\theta_t n_t \, dt + d(\theta_t n_t)).$$

If the Bellman equation (12) holds, then $\hat{\Theta}_t$ is a supermartingale under an arbitrary alternative strategy, so

$$\theta_0 n_0 = \hat{\Theta}_0 \geq E[\hat{\Theta}_t] \geq E\left[\int_0^t e^{-\rho s} n_s \, d\hat{\zeta}_s\right].$$

Taking the limit $t \to \infty$, we find that $\theta_0 n_0$ is an upper bound on the expert's payoff from an arbitrary strategy.

Conversely, if the Bellman equation (12) fails, then there exists a strategy $\{\hat{x}_t, d\hat{\zeta}_t\}$ such that

$$n_t \, d\hat{\zeta}_t - \rho\theta_t n_t \, dt + E[d(\theta_t n_t)] \geq 0,$$

with a strict inequality on the set of positive measures. Then, for large enough t,

$$\theta_0 n_0 = \hat{\Theta}_0 < E[\hat{\Theta}_t]$$

and so the expert's expected payoff from following the strategy $\{\hat{x}_t, d\hat{\zeta}_t\}$ until time t, and $\{x_t, d\zeta_t\}$ thereafter, exceeds that from following $\{x_t, d\zeta_t\}$ throughout.

PROOF OF PROPOSITION II.2:

Using the laws of motion of θ_t and n_t as well as Ito's lemma, we can transform the Bellman equation (12) into

$$\rho\theta_t n_t dt = \max_{\hat{x}_t \geq 0, d\hat{\zeta}_t \geq 0} (1-\theta_t) n_t d\hat{\zeta}_t + r\theta_t n_t dt + n_t E_t[d\theta_t] + \hat{x}_t \theta_t n_t \left(E_t[dr_t^k] - r \, dt + \sigma_t^\theta(\sigma + \sigma_t^q) \, dt\right).$$

Assume that $n_t\theta_t$ represents the expert's maximal expected future payoff, so that by Lemma II.1 the Bellman equation holds, and let us justify (i) through (iii). The Bellman equation cannot hold unless $1 \leq \theta_t$ and $E_t[dr_t^k]/dt - r + \sigma_t^\theta(\sigma + \sigma_t^q) \leq 0$, since otherwise the right-hand side of the Bellman equation can be made arbitrarily large. If so, then the choices $d\hat{\zeta}_t = 0$ and $\hat{x}_t = 0$ maximize the right-hand side, which becomes equal to $r\theta_t n_t \, dt + \theta_t n_t \mu_t^\theta \, dt$. Thus,

$$\rho\theta_t n_t \, dt = r\theta_t n_t \, dt + \theta_t n_t \mu_t^\theta \, dt \quad \Rightarrow \quad \text{(E)}.$$

Furthermore, any $d\hat{\zeta}_t > 0$ maximizes the right-hand side only if $\theta_t = 1$, and $\hat{x}_t > 0$ does only if $E_t[dr_t^k]/dt - r + \sigma_t^\theta(\sigma + \sigma_t^q) \leq 0$. This proves (i) through (iii). Finally, Lemma II.1 implies that the transversality condition must hold for any strategy that attains value $n_t\theta_t$, proving (iv).

Conversely, it is easy to show that if (i) through (iii) hold, then the Bellman equation also holds and the strategy $\{x_t, d\zeta_t\}$ satisfies (11). Thus, by Lemma II.1, the strategy $\{x_t, d\zeta_t\}$ is optimal and attains value $\theta_t n_t$.

PROOF OF LEMMA II.3:

Aggregating over all experts, the law of motion of N_t is

$$dN_t = rN_t\,dt + \psi_t q_t K_t(dr_t^k - r\,dt) - dC_t, \tag{C1}$$

where C_t are aggregate payouts. Furthermore, note that $d(q_t K_t)/(q_t K_t)$ are the capital gains earned by a world portfolio of capital, with weight ψ_t on expert capital and $1-\psi_t$ on household capital. Thus, from (5) and (6),

$$\frac{d(q_t K_t)}{q_t K_t} = \underbrace{dr_t^k - \frac{a-\iota(q_t)}{q_t}\,dt}_{\text{expert capital gains}} - \underbrace{(1-\psi_t)(\underline{\delta}-\delta)\,dt}_{\text{adjustment for household−held capital}},$$

since household capital gains are less than those of experts by $\underline{\delta}-\delta$. Using Ito's lemma,

$$\frac{d(1/(q_t K_t))}{1/(q_t K_t)} = -dr_t^k + \frac{a-\iota(q_t)}{q_t}\,dt + (1-\psi_t)(\underline{\delta}-\delta)\,dt + (\sigma+\sigma_t^q)^2\,dt.$$

Combining this equation with (C1) and using Ito's lemma, we get

$$d\eta_t = (dN_t)\frac{1}{q_t K_t} + N_t\,d\left(\frac{1}{q_t K_t}\right) + \psi_t q_t K_t(\sigma+\sigma_t^q)\frac{-1}{q_t K_t}(\sigma+\sigma_t^q)\,dt =$$

$$(\psi_t-\eta_t)(dr_t^k - r\,dt - (\sigma+\sigma_t^q)^2\,dt) + \eta_t\left(\frac{a-\iota(q_t)}{q_t} + (1-\psi_t)(\underline{\delta}-\delta)\right)dt - \eta_t d\zeta_t,$$

where $d\zeta_t = dC_t/N_t$. If $\psi_t > 0$, then Proposition II.2 implies that $E[dr_t^k] - r\,dt = -\sigma_t^\theta(\sigma+\sigma_t^q)\,dt$, and the law of motion of η_t can be written as in (15).

PROOF OF PROPOSITION II.4:

First, we derive expressions for the volatilities of η_t, q_t, and θ_t. Using (15), the law of motion of η_t, and Ito's lemma, the volatility of q_t is given by

$$\sigma_t^q q(\eta) = q'(\eta)\underbrace{(\psi-\eta)(\sigma+\sigma_t^q)}_{\sigma_t^\eta \eta} \quad\Rightarrow\quad \sigma_t^q = \frac{q'(\eta)}{q(\eta)}\underbrace{\frac{(\psi-\eta)\sigma}{1-\frac{q'(\eta)}{q(\eta)}(\psi-\eta)}}_{\sigma_t^\eta \eta}.$$

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

$$\frac{a-\underline{a}}{q(\eta)} + \underline{\delta} - \delta + (\sigma+\sigma_t^q)\sigma_t^\theta \geq 0, \tag{C2}$$

with equality if $\psi < 1$. Moreover, when $q(\eta), q'(\eta), \theta(\eta) > 0$ and $\theta'(\eta) < 0$, then $\sigma_t^q, \sigma_t^\eta > 0$ are increasing in ψ while $\sigma_t^\theta < 0$ is decreasing in ψ. Thus, the left-hand side of (C2) is decreasing from $\frac{a-\underline{a}}{q(\eta)} + \underline{\delta} - \delta$ at $\psi = \eta$ to $-\infty$ at $\psi = \eta + \frac{q(\eta)}{q'(\eta)}$, justifying our procedure for determining ψ.

We get μ_t^η from (15), μ_t^q from (EK), and μ_t^θ from (E). The expressions for $q''(\eta)$ and $\theta''(\eta)$ then follow directly from Ito's lemma and (15), the law of motion of η_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 $q(0) = \underline{q}$. The households are willing to pay this price for capital if they have to hold it forever. 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.[35] 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

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

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^*.

PROOF OF PROPOSITION III.2:

Since $q'(\eta^*) = \theta'(\eta^*) = 0$, the drift and volatility of η at η^* are given by

$$\mu_t^\eta(\eta^*)\eta^* = (1-\eta^*)\sigma^2 + \frac{a - \iota(q(\eta^*))}{q(\eta^*)}\eta^* > 0 \quad \text{and} \quad \sigma_t^\eta(\eta^*)\eta^* = (1-\eta^*)\sigma.$$

Hence, $D'(\eta^*) = 2\mu_t^\eta(\eta^*)\eta^*/(\sigma_t^\eta(\eta^*)\eta^*)^2 D(\eta^*) > 0$, where $D(\eta) = d(\eta)(\sigma_t^\eta(\eta)\eta)^2$. Furthermore, because in the neighborhood of η^*,

$$\sigma_t^\eta(\eta)\eta = \frac{(1-\eta)\sigma}{1-(1-\eta)q'(\eta)/q(\eta)}$$

is *decreasing* in η, it follows that the density $d(\eta)$ must be *increasing* in η.

[35]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 $\left(\sigma_t^\eta\right)^2 \epsilon$. 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 $\epsilon \to 0$.

The dynamics near $\eta = 0$ are more difficult to characterize because of the singularity there. We will do that by conjecturing, and then verifying, that asymptotically as $\eta \to 0$,

$$\mu_t^\eta = \hat{\mu} + o(1) \quad \text{and} \quad \sigma_t^\eta = \hat{\sigma} + o(1),$$

i.e., η_t evolves as a geometric Brownian motion, and that

$$\psi(\eta) = C_\psi \eta + o(\eta), \quad q(\eta) = \underline{q} + C_q \eta^\alpha + o(\eta^\alpha), \quad \text{and} \quad \theta(\eta) = C_\theta \eta^{-\beta} + o(\eta^{-\beta})$$

for some constants $C_\psi > 1$, C_q, $C_\theta > 0$, $\alpha, \beta \in (0, 1)$. We need to verify that the equilibrium equations hold, up to terms of smaller order. Using the equations of Proposition II.4, we have

$$\sigma_t^\eta = \frac{(C_\psi - 1)\sigma + o(1)}{1 - O(\eta^\alpha)} \;\Rightarrow\; \hat{\sigma} = (C_\psi - 1)\sigma,$$

$$\sigma_t^q = \frac{\alpha C_q \eta^\alpha}{\underline{q}} \hat{\sigma} + o(\eta^\alpha) = o(1), \quad \sigma_t^\theta = -\beta\hat{\sigma} + o(1),$$

$$\text{(C3)} \qquad (17) \quad \Rightarrow \quad \beta\hat{\sigma}\sigma = \Lambda \quad \Rightarrow \quad \hat{\sigma} = (C_\psi - 1)\sigma = \frac{\Lambda}{\beta\sigma} \qquad \text{and}$$

$$\hat{\mu} = -\hat{\sigma}(\sigma - \beta\hat{\sigma}) + \frac{a - \iota(\underline{q})}{\underline{q}} + \underline{\delta} - \delta = -\frac{\Lambda}{\beta\sigma}\left(\sigma - \frac{\Lambda}{\sigma}\right) + \frac{\underline{a} - \iota(\underline{q})}{\underline{q}} + \Lambda.$$

We can determine μ_t^q from the household valuation equation

$$\frac{\underline{a} - \iota(q_t)}{q_t} + \Phi(\iota(q_t)) - \underline{\delta} + \mu_t^q + \sigma\sigma_t^q = r$$

instead of that of experts, because we already took into account (17). By the envelope theorem,

$$\frac{\underline{a} - \iota(q(\eta))}{q(\eta)} + \Phi(q(\eta)) - \underline{\delta} = \underbrace{\frac{\underline{a} - \iota(\underline{q})}{\underline{q}} + \Phi(\underline{q}) - \underline{\delta}}_{r} - \frac{\underline{a} - \iota(\underline{q})}{\underline{q}^2}(C_q\eta^\alpha + o(\eta^\alpha)) + o(\eta^\alpha).$$

Therefore,

$$\mu_t^q = \frac{\underline{a} - \iota(\underline{q})}{\underline{q}^2} C_q \eta^\alpha - \frac{\alpha C_q \eta^\alpha}{\underline{q}} \hat{\sigma}\sigma + o(\eta^\alpha).$$

Our conjecture is valid if equations

$$\mu^q q(\eta) = q'(\eta)\mu_t^\eta \eta + \frac{1}{2} q''(\eta)(\sigma_t^\eta \eta)^2 \quad \text{and} \quad \mu^\theta \theta(\eta) = \theta'(\eta)\mu_t^\eta \eta + \frac{1}{2}\theta''(\eta)(\sigma_t^\eta \eta)^2$$

hold up to higher-order terms of $o(\eta^\alpha)$ and $o(\eta^{-\beta})$, respectively. Ignoring those terms, we need

$$\frac{\underline{a}-\iota(\underline{q})}{\underline{q}} C_q \eta^\alpha - \alpha C_q \eta^\alpha \hat{\sigma}\sigma = \alpha C_q \eta^\alpha \hat{\mu} + \frac{1}{2}\alpha(\alpha-1) C_q \eta^\alpha \hat{\sigma}^2 \quad \text{and}$$

$$(\rho - r) C_\theta \eta^{-\beta} = -\beta C_\theta \eta^{-\beta} \hat{\mu} + \frac{1}{2}\beta(\beta+1) C_\theta \eta^{-\beta} \hat{\sigma}^2. \tag{C4}$$

These equations lead to

$$\frac{\underline{a}-\iota(\underline{q})}{\underline{q}} - \alpha\frac{\Lambda}{\beta} = \alpha\left(-\frac{\Lambda}{\beta\sigma}\left(\sigma - \frac{\Lambda}{\sigma}\right) + \frac{\underline{a}-\iota(\underline{q})}{\underline{q}} + \Lambda\right) + \frac{1}{2}\alpha(\alpha-1)\frac{\Lambda^2}{\beta^2\sigma^2} \quad \text{and}$$

$$\rho - r = -\beta\left(-\frac{\Lambda}{\beta\sigma}\left(\sigma - \frac{\Lambda}{\sigma}\right) + \frac{\underline{a}-\iota(\underline{q})}{\underline{q}} + \Lambda\right) + \frac{1}{2}\beta(\beta+1)\frac{\Lambda^2}{\beta^2\sigma^2} \quad \Rightarrow$$

$$\alpha\left(\frac{\Lambda^2}{\beta\sigma^2} + \Lambda\right) + \frac{\underline{a}-\iota(\underline{q})}{\underline{q}}(\alpha-1) + \frac{1}{2}\alpha(\alpha-1)\frac{\Lambda^2}{\beta^2\sigma^2} = 0 \quad \text{and} \tag{C5}$$

$$\rho - r = \Lambda - \beta\left(\frac{\underline{a}-\iota(\underline{q})}{\underline{q}} + \Lambda\right) + \frac{(1-\beta)\Lambda^2}{2\beta\sigma^2} \tag{C6}$$

We can solve for β, C_ψ, and α in the following order. First, equation (C6) has a solution $\beta \in (0,1)$. To see this, note that, as $\beta \to 0$ from above, the right-hand side of (C6) converges to infinity. For $\beta = 1$, the right-hand side becomes

$$-\frac{\underline{a}-\iota(\underline{q})}{\underline{q}} < 0.$$

We have $\underline{a}-\iota(\underline{q}) > 0$, since the net rate of output that households receive at $\eta = 0$ must be positive. Second, equation (C3) determines the value of $C_\psi > 1$ for any $\beta > 0$. Lastly, equation (C5) has a solution $\alpha \in (0,1)$. To see this, note that the left-hand side is negative when $\alpha = 0$ and positive when $\alpha = 1$.

This confirms our conjecture about the asymptotic form of the equilibria near $\eta = 0$. Arbitrary values of constants C_q and C_θ are consistent with these asymptotic dynamics. The value of C_q has to be chosen to ensure that functions $q(\eta)$ and $\theta(\eta)$ reach slope 0 at the same point η^*, and the C_θ, to ensure that $\theta(\eta^*) = 1$.

We are now ready to characterize the asymptotic form of the stationary distribution near $\eta = 0$. We have $D'(\eta) = 2\hat{\mu}/\hat{\sigma}^2 \, D(\eta)/\eta$, so

$$D(\eta) = C_D \, \eta^{2\hat{\mu}/\hat{\sigma}^2} \quad \text{and} \quad d(\eta) = D(\eta)/(\hat{\sigma}\eta)^2 = C_d \, \eta^{2\hat{\mu}/\hat{\sigma}^2 - 2}. \tag{C7}$$

Equation (C4) implies that

$$\frac{2(\rho - r)}{\hat{\sigma}^2} = -\beta \frac{2\hat{\mu}}{\hat{\sigma}^2} + \beta(\beta + 1) \quad \Rightarrow \quad \frac{2\hat{\mu}}{\hat{\sigma}^2} - 2 = \beta - 1 - \frac{2(\rho - r)}{\Lambda^2}\sigma^2\beta.$$

We see that $2\hat{\mu}/(\hat{\sigma}^2) - 2 < 0$, and so $d(\eta) = C_d \, \eta^{2\hat{\mu}/\hat{\sigma}^2 - 2} \to \infty$ as $\eta \to 0$. Furthermore, if

$$\frac{2\hat{\mu}}{\hat{\sigma}^2} - 2 > -1 \quad \Leftrightarrow \quad 1 - \frac{2(\rho - r)}{\Lambda^2}\sigma^2 > 0 \quad \Leftrightarrow \quad 2(\rho - r)\sigma^2 < \Lambda^2,$$

then the stationary density exists and has a hump near $\eta = 0$. Otherwise, if $2(\rho - r)\sigma^2 \geq \Lambda^2$, then the integral of $d(\eta)$ is infinity, implying that the stationary density does not exist and in the long run η_t ends up in an arbitrarily small neighborhood of 0 with probability close to 1.

PROOF OF PROPOSITION III.3:

Boundary conditions (23) as well as equation (24) follow from the market-clearing condition for consumption goods,

$$r(q_t K_t - N_t) + \rho N_t = (\psi a + (1 - \psi)\underline{a} - \iota(q_t))K_t. \tag{C8}$$

Furthermore, since the volatilities of expert and household net worths are $\frac{\psi_t}{\eta_t}(\sigma + \sigma_t^q)$ and $\frac{1-\psi_t}{1-\eta_t}(\sigma + \sigma_t^q)$, respectively, the portfolio optimization conditions imply that

$$E[dr_t^k - dr_t]/dt = \frac{\psi_t}{\eta_t}(\sigma + \sigma_t^q)^2 \quad \text{and} \quad E[d\underline{r}_t^k - dr_t]/dt \leq \frac{1 - \psi_t}{1 - \eta_t}(\sigma + \sigma_t^q)^2, \quad \text{with equality if } \psi_t < 1.$$

As $E[dr_t^k - d\underline{r}_t^k]/dt = (a - \underline{a})/q_t + \underline{\delta} - \delta$, these two conditions together imply that, when $\psi_t < 1$,

$$\frac{a - \underline{a}}{q_t} + \underline{\delta} - \delta = \left(\frac{\psi_t}{\eta_t} - \frac{1 - \psi_t}{1 - \eta_t}\right)(\sigma + \sigma_t^q)^2. \tag{C9}$$

This leads to the first equation in (25). The second equation in (25) holds because, as in the risk-neutral model,

$$\sigma + \sigma_t^q = \frac{\sigma}{1 - (\psi_t - \eta_t)q'(\eta)/q(\eta)}.$$

Equations (26) hold because, by Lemma II.3,

$$\frac{d\eta_t}{\eta_t} = \frac{\psi_t - \eta_t}{\eta_t}(dr_t^k - r\,dt - (\sigma + \sigma_t^q)^2\,dt) + \frac{a - \iota(q_t)}{q_t}\,dt + (1 - \psi_t)(\underline{\delta} - \delta)\,dt - \rho\,dt.$$

LEMMA C.1: *Under the logarithmic utility model, the stationary density exists if*

$$2\sigma^2(\Lambda + r - \rho) + \Lambda^2 > 0$$

and has a hump at 0 if also $\rho > r + \Lambda$, where $\Lambda = (a - \underline{a})/q(0) + \underline{\delta} - \delta$.

PROOF:

Note that asymptotically $\sigma_t^q \to 0$ as $\eta \to 0$. Thus, from equation (C9),

$$\psi_t = \eta_t \frac{\Lambda}{\sigma^2} + o(\eta_t).$$

Therefore, equation (26) implies that

$$\sigma_t^\eta = \frac{\Lambda}{\sigma} + o(1) \quad \text{and} \quad \mu_t^\eta = (\sigma_t^\eta)^2 + \Lambda + r - \rho + o(1).$$

The Kolmogorov forward equation (see (C7)) implies that asymptotically the stationary density of η_t takes the form

$$d(\eta) = C_d \eta^{\beta_d}, \quad \text{where} \quad \beta_d = 2\left(\frac{\mu_t^\eta}{(\sigma_t^\eta)^2} - 1\right) = 2\sigma^2 \frac{\Lambda + r - \rho}{\Lambda^2}.$$

Thus, unlike in the risk-neutral case, the stationary density is nonsingular if $2\sigma^2(\Lambda + r - \rho) + \Lambda^2 > 0$ and has a hump at 0 if $\rho > r + \Lambda$.

PROOF OF PROPOSITION IV.1:

From the proof of Lemma C.1,

$$\psi_t = \eta_t \frac{\Lambda}{\sigma^2} + o(\eta_t)$$

under logarithmic utility. Under risk neutrality,

$$\psi_t = C_\psi \eta_t + o(\eta_t), \quad \text{where} \quad C_\psi = \frac{\Lambda}{\beta\sigma^2} + 1.$$

The variable β is determined by equation (C6), which implies that $\beta = 1 + O(\sigma^2)$ when σ is small. Thus,

$$\psi_t = \eta_t\left(\frac{\Lambda}{\sigma^2} + O(1)\right) + o(\eta_t).$$

In both cases,

$$\sigma_t^\eta = \frac{\psi_t - \eta_t}{\eta_t}(\sigma + \sigma_t^q),$$

and $\sigma_t^q \to 0$ as $\eta \to 0$. Thus,

$$\sigma_t^\eta = \frac{\Lambda}{\sigma} + O(\sigma)$$

as $\eta \to 0$.