Working Paper Series

No 25 / September 2016

Macroeconomic effects of secondary market trading

by
Daniel Neuhann

ESRB
European Systemic Risk Board
European System of Financial Supervision

Abstract

This paper develops a theory of the secondary market trading of financial securitities in which endogenous asset market dynamics generate periods of growing aggregate credit volumes and falling credit standards even in the absence of "financial shocks." Falling credit standards in turn lead to excess risk exposure in the aggregate, precipitating future crises. The credit cycle is triggered by low interest rates, and longer booms lead to sharper crises. Saving gluts and expansionary monetary policy thus lead to financial fragility over time. Pro-cyclical regulation of secondary market traders, such as asset managers or hedge funds, can improve welfare even when such traders are not levered.

JEL classification: G01, E32, E44

Keywords: secondary markets, securitization, credit cycles, financial crisis, financial fragility, credit booms, saving gluts, risk-taking channel of monetary policy.

Credit booms are pervasive phenomena that frequently precipitate financial crises. Prominent examples, such as the Great Depression, the Financial Crisis of 2008, and the 1997 Asian Financial Crisis, were also preceded by sharp increases in the securitization and syndication of financial assets, and a deterioration in the quality of loans backing these securities.[1] Recent literature frequently relies on exogenous "financial shocks" that relax borrowing constraints or lower credit standards to account for these facts.[2] This paper instead develops a theory of securitization in which these "shocks" arise endogenously. The theory generates credit cycles in which growing credit volumes gradually lead to falling lending standards, and sheds light on how policy and macroeconomic conditions shape the emergence and evolution of credit booms.

I study securitization as a means of reallocating systematic risk exposure in an economy where financial intermediaries ("banks") must (i) provide collateral in order to borrow, and (ii) exert costly effort to originate high-quality loans ("moral hazard"). Risky assets are poor collateral because they tempt banks into originating excessively risky assets at expense of bondholders ("savers"). The motive for securitization is that selling risk exposure alleviates this concern and allows banks to borrow and lend more (*risk transfer channel*). The cost is that securitization may harm the quality of credit because asset sales harm banks' incentives to produce high-quality assets in the first place (*shirking channel*). The strength of these channels depends on macroeconomic conditions, and their dynamic interplay generates credit booms and busts.

The risk transfer channel boosts aggregate lending volumes only if banks are able to sell systematic risk exposure to buyers better able to bear it. Because banks are subject to moral hazard, I identify investors who lack the know-how to directly originate loans in primary markets ("financiers") as natural buyers. Precisely because these investors are distant from origination, they do not affect the quality of investment. As a result, they may be able to borrow more from savers per unit of collateral than bankers, boosting the flow of funds into the financial system. Securitization thus allows for the efficient use of scarce collateral by separating the *production* of risky loans from the *holding* of the resulting risk. This prediction is consistent with the growing credit market participation of non-bank financial institutions, such as hedge funds or mutual funds, prior to the 2008 financial crisis. Greenlaw, Hatzius, Kashyap, and Shin (2008) show that non-bank intermediaries were more exposed to downside risk than loan-originating banks, while Coval, Jurek, and Stafford (2009) document that securitized assets were primarily exposed to systematic risk.

The shirking channel operates only if (i) financiers cannot perfectly screen the quality of assets, (ii) banks weakly prefer to shirk and sell low-quality assets, and (iii) financiers are willing to buy securities even when they know that (some) banks are selling low-quality assets. I impose the first requirement as a technological constraint, and analyze the conditions under which the latter two are satisfied. My first main result is that, depending on the distribution of wealth across savers, banks, and financiers, asset prices may simultaneously be (i) high enough that it is prof-

[1] Typically, secondary market trading occurs through the *securitization* of financial assets. While financial intermediaries issued less than \$100 billion in securitized assets in 1900, they issued more than \$3.5 trillion in 2006. Gorton and Metrick (2012) survey the development of secondary markets and securitization in the United States prior to the 2008 crisis. Mian and Sufi (2009) and Ivashina and Sun (2011) provide evidence of a credit boom for households and firms. Brunnermeier (2009), Shin (2009), and the Report of the U.S. Financial Crisis Inquiry Commission (2011) review the role of secondary markets and securitization in the 2000-2008 boom and bust. White (2009) and Kaminsky (2008) provide evidence for the Great Depression and Asian Financial Crisis, respectively.

[2] See Justiniano, Primiceri, and Tambalotti (2015) for one example.

itable for a fraction of banks to produce low-quality assets just to sell them, and (ii) low enough that buying assets remains profitable on average for financiers. 'Financial shocks' are thus a manifestation of the wealth distribution, and credit cycles a consequence of its endogenous evolution.

The argument relies on the presence of intermediation rents. When bank capital is scarce, savers must pay a premium in order to access investment opportunities. Financiers capture part of these rents because banks' shadow cost of risk depresses prices below par. An immediate corollary is this: financiers are willing to tolerate losses on a fraction of low-quality assets because they can earn sufficient profits on the rest. Intermediation rents thus provide the scope for equilibrium shirking.

Do banks want to shirk? It depends on the asset price. No bank produces low-quality assets in order to retain them. Hence, shirking is worthwhile only if the asset price p is above a threshold $\bar{p}$. Crucially, $\bar{p}$ is strictly below the expected value of high-quality assets $\mathbb{E}[R]$ because shirking is privately costly. As a result, there exist prices at which banks are willing to shirk and financiers earn rents as long as they receive a sufficiently large share of good assets. Griffin, Lowery, and Saretto (2014) show that, prior to 2008, issuers of complex securities produced and sold excessively risky products that underperform during downturns.

The equilibrium asset price p^* is pinned down by the relative scarcity of bank and financier capital. It is large when financiers are relatively wealthy, and attains the threshold $\bar{p}$ when financier wealth is large enough. Naturally, there are no equilibria in which $p^* > \bar{p}$ since all banks would then begin to shirk. Hence, the asset price is equal to $\bar{p}$ when financier wealth is above a threshold, and banks are indifferent between shirking and exerting effort. Still, some banks must shirk in equilibrium. Because the price is bounded by $\bar{p} < \mathbb{E}[R]$, excess demand at $p^* = \bar{p}$ must be absorbed by increasing the quantity of assets sold at $\bar{p}$. The way to do so is to increase the share of bankers who shirk, because those who produce low-quality assets sell more at $\bar{p}$ than those who exert effort. Since $\bar{p} < \mathbb{E}[R]$, financiers continue buying assets as long as the fraction of shirking banks is not too large. Yet private incentives are not aligned with social welfare: while individual financiers find it optimal to buy assets even when a fraction of banks shirk, a social planner would prevent them from doing so. The source of this inefficiency is a pecuniary externality. Financiers do not internalize that their asset purchases increase excess demand and harm bank incentives. The welfare effects can be severe: a partial destruction of financier wealth may generate strict a Pareto improvement.

The adverse effects of securitization thus stem from two imbalances: an excess of savings that boosts intermediation rents, and an excess of financier wealth that raises asset prices. The theory's dynamic implications, in turn, stem from the endogenous evolution of the wealth distribution. The key force is the risk-transfer channel, which allocates aggregate risk exposure to financiers. The result is that financier wealth grows disproportionately during macroeconomic upturns in which this risk-taking pays off. Initially, this strengthens the risk transfer channel and boosts lending volumes without harming asset quality. Over time, however, financiers may grow wealthy enough to trigger the shirking channel, leading to deteriorating lending standards and falling asset quality. Low-quality assets expose the financial system to excess downside risk, providing the scope for an eventual crisis. Macroeconomic upturns can thus produce precisely the wealth dynamics that generate credit booms with falling asset quality.

Credit cycles are triggered by low risk-free rates that can be due to either monetary policy or saving gluts. Financiers lever more when borrowing is cheap, and thus grow faster during the upturn. Simultaneously, the risk-transfer and bad incentive channel combine to disproportionately expose financiers to both aggregate risk and low-quality assets. As a result, they suffer disproportionately during downturns. Recoveries are slow because the risk transfer channel is weak when financiers are impaired. Longer booms precede sharper crises because lending standards fall over time. Adrian and Shin (2010) estimate that the combined balance sheet size of hedge funds and broker-dealers was smaller than that of bank holding companies before 1990 but almost twice as large by 2007. Keys, Mukherjee, Seru, and Vig (2010) and Piskorski, Seru, and Witkin (2015) provide empirical evidence of falling credit standards and growing moral hazard over the course of the 2000-2007 U.S. credit boom. Schularick and Taylor (2012), Mendoza and Terrones (2012), and Reinhart and Rogoff (2009) document that longer credit booms predict sharper crises. Gorton and Metrick (2012) and Krishnamurthy, Nagel, and Orlov (2014) show that the fragility of leveraged secondary market traders was at the heart of the 2008 financial crisis, and that the migration of risk back onto bank balance sheets was an important determinant of the larger credit crunch to follow.

I derive three main policy implications. The first is that regulation which limits the accumulation of financier wealth or hampers financiers' ability to purchase excess amounts of loan-backed assets can be welfare-enhancing. Notably, this motive for regulation is independent of the financial structure of financiers. Indeed, it applies to zero-leverage financial institutions, such as asset managers, who have traditionally been outside the scope of financial regulation precisely because their lack of leverage was thought to eliminate financial fragility and agency frictions. The second is that low risk-free rates, due to for example monetary policy or saving gluts, can trigger disproportionate financier growth. Falling funding costs boost intermediation rents and increase the value of risk-transfer. This sets in motion the growth of financiers and may lead to lower credit standards in the future. Caballero and Krishnamurthy (2009) argue that monetary policy was indeed expansionary during the early stages of the pre-2008 U.S. credit boom. A unique feature of this particular risk-taking channel of monetary policy is that it is *persistent* and *asymmetric*: one-time shocks to the interest rate may be sufficient to generate a credit cycle going forward, but the cycle cannot be reined in by resetting the policy rate to the previous levels. Third, leverage restrictions on banks may prematurely trigger the bad incentive channel because banks respond by selling fewer assets, raising asset prices.

Related Literature. The seminal study of collateral constraints in general equilibrium is Geanakoplos (1997). Fostel and Geanakoplos (2015) characterize the resulting leverage and collateral constraints in binomial economies, and show that it results in a VaR=0-borrowing constraint. I take such a collateral-based borrowing constraint as given and study how the reallocation of risk exposure stretches scarce collateral and affects the efficiency of investment. Fostel and Geanakoplos (2008) study how collateral and leverage affect asset prices and generate spillovers across asset classes, while Garleanu and Pedersen (2011) do so using collateral-based margin constraints. Gromb and Vayanos (2002) study liquidity provision in segmented markets by collateral-constrained arbitrageurs. They find that arbitrageurs may overexposed or underexposed to risky assets from a welfare perspective due to a pecuniary externality. The inefficiency in my model also stems from a pecuniary externality, but it affects real investment quality rather than liquidity.

Mendoza (2010) quantitatively studies collateral constraints and leverage over the business cycle, and shows that these constraints amplify the response to negative macroeconomic shocks. Lorenzoni (2008), Bianchi (2011), and Bianchi and Mendoza (2012) show that pecuniary externalities can trigger fire sales that amplify credit crunches, while Bigio (2014) and Kurlat (2013) study market shutdowns during downturns. I study how the reallocation of risk during upturns harms credit quality and generates excessive risk-taking. Di Tella (2014) shows that financial intermediaries want to insure themselves against aggregate shocks that tighten equity-based borrowing constraints. I study the allocation of risk across heteregeneous intermediaries when a $VaR = 0$ rule prevents such risk-sharing amongst all agents. Gorton and Ordoñez (2014) propose a dynamic model of credit booms and busts based on the desire of agents to trade informationally-insensitive assets. Booms and busts occur due to the evolution of beliefs. I emphasize the evolution of the wealth distribution and the deterioration of investment efficiency over the credit cycle. Gennaioli, Shleifer, and Vishny (2013) argue that securitization allows for improved sharing of idiosyncratic risk, and is efficient unless agents neglect aggregate risk. I study the re-allocation of aggregate risk, and show that excessive secondary market trading can have deleterious effects even in a fully rational framework. Moreover, I explicitly model the dynamics of secondary markets and argue why booms can endogenously lead to financial fragility. Parlour and Plantin (2008) and Vanasco (2014) have studied the effects of secondary market liquidity on moral hazard and information acquisition in primary markets in static partial equilibrium settings. I differ in that I study the macroeconomic dynamics of secondary markets and emphasize the endogenous evolution of intermediary wealth. Chari, Shourideh, and Zetlin-Jones (2014) show how secondary markets may collapse suddenly in the presence of adverse selection. I study how growing secondary markets can lead to falling asset quality. This concern is shared by Bolton, Santos, and Scheinkman (2016), who study how origination incentives vary with the demand for assets by informed and uninformed buyers.

1 A Single Period Model

I begin my analysis using a single-period model in which the wealth of all agents is fixed. I use this setting to derive the risk-transfer channel and the shirking channel. I embed the model in an overlapping generation setting in Section 2. I use the dynamic model to study how the risk-transfer channel and the shirking channel shape the endogenous evolution of the wealth distribution.

1.1 Setting

There are three types of agents in the economy, each forming a continuum of unit mass: *savers*, *bankers*, and *financiers*. Types are indexed by subscripts S, B, and F, respectively. Financiers and banks are risk neutral. I create a role for collateral by assuming that savers are infinitely risk-averse as in Gennaioli, Shleifer, and Vishny (2013) or Caballero and Farhi (2014). As a result, all lending by savers to financiers and bankers must be in the form of risk-free debt. The assumption thus yields the the VaR=0-collateral constraint derived by Fostel and Geanakoplos (2015) and assumed in reduced form by e.g. Gromb and Vayanos (2002).

There is a single good that can be used for consumption and investment. Agents are born with endowment w_S, w_B and w_F of this good, respectively. The only source of risk is an aggregate state $z \in \{l, h\}$ whose realizations I refer to as the *low state* and the *high state*. The probability of state z is $\pi_z \in (0, 1)$. There are two investment technologies. The *storage technology* is accessible to all agents. It generates a certain rate of return $R \in \{\underline{R}, \bar{R}\}$ per unit of capital invested. I assume that $R = \underline{R} \leq \bar{R}$ for savers, while $R = \bar{R} = 1$ for bankers and financiers. I use $\underline{R}$ to parametrize savers' willingness to pay for financial services. The majority of my analysis relies on the benchmark case $\bar{R} = \underline{R}$.

Only banks can invest in the risky technology. The technology operates efficiently only if monitored by banks at private cost m per unit of investment. It generates a rate of return Y_z in state z if the banker monitors, and y_z otherwise.

Assumption 1 (Payoffs)**.**
The payoffs of the risky technology satisfy $\mathbb{E}_z Y_z > \mathbb{E}_z y_z + m$*,* $y_l < Y_l$*, and* $\mathbb{E}_z Y_z > \bar{R}$*.*

The assumption states that (i) shirking is inefficient and induces additional downside risk, and (ii) the monitored risky technology delivers a higher return than safe the technology. Going forward, I use the shorthand $\hat{Y} = \mathbb{E}_z Y_z$ and $\hat{y} = \mathbb{E}_z y_z$.

Asset Markets. Markets are incomplete. There are two financial assets in zero net supply: a risk-free zero-coupon bond with face value one, and a risky asset that represents a claim on the return of the risky technology. Financiers and banks borrow from savers by issuing bonds at price p. I assume that banks do not issue bonds to financiers, and later show that this assumption is immaterial to my results.

Banks use the risky asset to offload risk exposure to financiers. I assume that it represents infinitely divisible claims on the returns of the risky technology. I say that a risky asset is *good* if it represents a claim on monitored risky investment, and *bad* otherwise. The goal of this paper is to characterize the macroeconomic conditions under which securitization hampers origination incentives. To do so, I assume that information frictions prevent financiers from screening the quality of assets sold by banks. In the absence of such an assumption, no bad assets would ever be traded, and there would be no feedback to origination incentives, contradicting the empirical evidence in Piskorski, Seru, and Witkin (2015) and Griffin, Lowery, and Saretto (2014). There are two aspects to this assumption. First, financiers cannot tell *individual* good assets apart from bad assets. Second, financiers cannot draw perfect inferences about asset quality by observing bank balance sheets. In practice, bank balance sheets are opaque and particularly difficult to observe in real time. Moreover, financial institutions typically trade *non-exclusively* with many other institutions simultaneously. Attar, Mariotti, and Salanié (2011) show that non-exclusivity renders it impossible to screen by quantity under asymmetric information. I therefore assume that banks' asset sales are only partially observable: banks can commit to selling *at least* $\underline{a}$ risky assets, but can deviate to selling *more than* $\underline{a}_B$ should they want to. This structure effectively assumes that banks can provide verifiable documentation for a certain number of loan sales while retaining the ability to engage in hidden trades or offloading risk using other financial instruments. Note that banks would never want to *buy back* risky loans in equilibrium because they would have to pay at least the expected value of the loan once they are owned by financiers. An alternative interpretation is that there are be multiple stages of bond trading, agents observe loan sales at every stage, but banks cannot commit to not selling assets again in the future. The partial commitment provided by $\underline{a}$ captures these concerns while maintaining tractability.

As I show below, the choice of $\underline{a}$ determines the degree to which banks will rely on asset sales to provide collateral to savers. Since this choice may impact banks' monitoring incentives, financiers will use $\underline{a}$ as well as aggregate trade volumes to form inferences about the average quality of assets traded by $\underline{a}$-banks. I therefore assume that there are sub-markets indexed by $\underline{a}$ in which the risky asset trades competitively at marginal price $p(\underline{a})$. I denote the fraction of low-quality assets trading on submarket $\underline{a}$ by $\phi(\underline{a})$, and the fraction of $\underline{a}$-banks who shirk by $\Phi(\underline{a})$. All financiers who purchase assets in a sub-market are allocated an equal share of low-quality assets. As a result, risky assets purchased on submarket $\underline{a}$ generate the return $x_z(\underline{a}) = \phi(\underline{a})y_z + (1 - \phi(\underline{a}))\,Y_z$ in state z. I denote the expected return by $\hat{x}(\underline{a})$. Submarket $\underline{a}$ is *active* if the demand and supply of risky assets issued by $\underline{a}$-bankers is strictly positive. Not all submarkets will be active in equilibrium. Nevertheless, it will be important to specify what prices and demand would look like if banks were to move to an inactive submarket. Naturally, this will depend on off-equilibrium beliefs and the assumed distribution of bargaining power in inactive submarkets. I provide further details during the course of the analysis. Throughout, I restrict attention to equilibria in which the pricing function $p : \mathbb{R}_+ \to \mathbb{R}_+$ is differentiable. As will become clear, this is a natural assumption in my setting.

Figure 1 summarizes the timing of events within the period. In stage 1, all agents receive their endowments. In stage 2, banks post a commitment to sell at least $\underline{a}$ units of the risky asset in stage 4, and risk-free bonds are traded. As will become clear, this partial commitment allows banks to rely on future asset sales as collateral *today*, while maintaining the scope for a deviation to excessive asset sales ex-post. This timing convention also allows me to straightforwardly embed the static model into a dynamic context. In stage 3, banks make investments in the risky technology using their own wealth and the proceeds from bond issuances in the funding market, and all agents make their investments in the safe technology. Moreover, banks make their monitoring decision. In stage 4, risky assets are traded subject to the constraint that banks must sell at least $\underline{a}$. In stage 5, the productivity shock z is realized, returns on investment accrue, accounts are settled, and all agents consume.

A notable missing market is an equity market in which banks can raise funds by selling shares to financiers. This segmentation is consistent with the data. Ivashina and Sun (2011) provide evidence that tranches of loans sold in secondary markets had lower yields than those held via direct claims on banks. Nevertheless, Section 2 shows that credit cycles with falling asset quality can arise even when banks are able to freely issue *inside* equity to financiers at no cost.

1.2 The Supply and Demand for Financial Assets

I now characterize the supply and demand for bonds and risky assets by analyzing agents' decision problems. Given a bond price q, the risk-free rate is $r_b(q) = \frac{1}{q}$. Bonds are demanded by savers as a means of lending funds to bankers and financiers. Risky assets are traded by bankers and

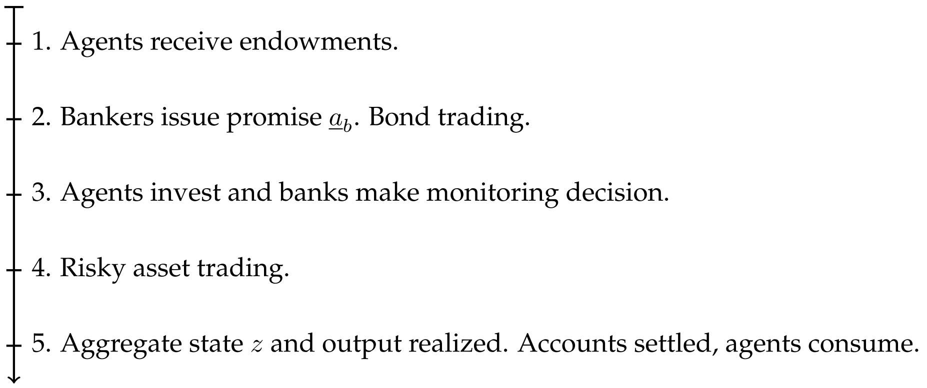

Figure 1: Timing of Events

financiers as a means of transferring risk exposure. All agents are price takers.

Savers' Problem. Since savers are infinitely risk averse, their portfolio consists only of risk-free bonds and the storage technology. Savers thus buy as many bonds as they can when $r_b(q) > \underline{R}$, and use the storage technology otherwise:

$$b_S(w_S, q) = \begin{cases} \frac{w_S}{q} & \text{if } r_b(q) > \underline{R} \\ [0, \frac{w_S}{q}] & \text{if } r_b(q) = \underline{R} \\ 0 & \text{if } r_b(q) < \underline{R}, \end{cases}$$

where b_S denotes savers' bond purchases. The bond price is thus bounded above by $\frac{1}{\underline{R}}$, and bond demand is increasing in w_S.

Financiers' Problem. Financiers choose the amount of capital to invest in storage s_F, risky asset purchases a_F, and bond issuances b_F. Generically, exactly one sub-market will offers the highest return on risky assets. I denote this submarket by $\underline{a}^*$ and solve a relaxed decision problem in which the financier purchases a non-negative quantity of assets in $\underline{a}^*$ only. Formally, this decision problem is:

$$\max_{s_F, a_F(\underline{a}^*), b_F} \quad \mathbb{E}_z \left[s_F + x_z(\underline{a}) a_F(\underline{a}^*) - b_F \right]$$

$$\text{s.t.} \quad s_F + p(\underline{a}^*) a_F(\underline{a}^*) \leq w_F \tag{1}$$

$$b_F \leq s_F + x_z(\underline{a}^*) a_F(\underline{a}^*) \ \text{ for all } \ z. \tag{2}$$

(1) is the budget constraint restricting asset purchases and storage to be weakly smaller than financier wealth. (2) is a collateral constraint that ensures that financiers are always able to fully honor their debts. The binding collateral constraint is the one corresponding to the low state ($z = l$). The two pertinent choices are (i) whether to buy risky assets, and (ii) whether to issue bonds to do so. A financier who does not issue any bonds can buy at most $a_F = \frac{w_F}{p(\underline{a}^*)}$ risky assets, and earns an expected rate of return on equity of

$$r(\underline{a}^*) = \frac{\hat{x}(\underline{a}^*)}{p(\underline{a}^*)}.$$

Financiers increase their asset purchases by issuing bonds subject to (2). Conditional on not investing in storage ($s_F = 0$), the collateral can be rearranged to yield a maximum bond issuance of

$$\bar{b}_F(\underline{a}^*) = \frac{x_b(\underline{a}^*)w_F}{p(\underline{a}^*) - qx_l(\underline{a}^*)}$$

Issuing $\bar{b}_F(\underline{a}^*)$ bonds allows the financier to purchase $a_F = \frac{w_F}{p(\underline{a}^*)-qx_b(\underline{a}^*)}$ risky assets, generating a expected rate of return on equity of

$$\tilde{r}(\underline{a}^*) = \frac{\hat{x}(\underline{a}^*) - x_l(\underline{a}^*)}{p(\underline{a}^*) - qx_l(\underline{a}^*)}.$$

Financiers are willing to invest in risky assets only if $r^*(\underline{a}^*) = \max\{r(\underline{a}^*), \tilde{r}(\underline{a}^*)\} \geq 1$, and strictly prefer to do so if $r^*(\underline{a}^*) > 1$. Moreover, it is straightforward to verify that financiers strictly prefer to invest in risky assets if $r^*(\underline{a}^*) = 1$ but $b_F > \bar{b}_F(\underline{a}^*)$ because buying the risky asset shifts risk exposure onto bondholders. The collateral constraint (2) is thus equivalent to $b_F \leq \bar{b}_F(\underline{a}^*)$ no matter the financier's portfolio.

The leveraged return is higher than the unleveraged return if $q\hat{x}(\underline{a}^*) > p(\underline{a}^*)$. Financiers thus issue $\bar{b}_F(\underline{a}^*)$ bonds if $q\hat{x}(\underline{a}^*) > p(\underline{a}^*)$, do not issue any bonds if $q\hat{x}(\underline{a}^*) < p(\underline{a}^*)$, and are indifferent when $q\hat{x}(\underline{a}^*) = p(\underline{a}^*)$. I summarize financier leverage by rewriting (2) as an equality constraint of the form $b_F = \mu \cdot \bar{b}_F(\underline{a}^*)$, and use $\mu \in [0,1]$ to summarize the degree to which financiers exhaust their borrowing capacity. Accordingly, the optimal portfolio is

$$a_F = \frac{w_F}{p(\underline{a}^*) - \mu(p,q)qx_l(\underline{a}^*)}, \quad b_F = \frac{\mu(p,q)w_F}{p(\underline{a}^*) - \mu(p,q)qx_l(\underline{a}^*)}, \quad \text{where } \mu(p,q) = \begin{cases} 1 & \text{if } q\hat{x}(\underline{a}^*) > p(\underline{a}^*) \\ [0,1] & \text{if } q\hat{x}(\underline{a}^*) = p(\underline{a}^*) \\ 0 & \text{otherwise.} \end{cases}$$

The demand for risky asset is (i) strictly increasing in w_F, (ii) strictly decreasing in $p(\underline{a}^*)$, and (iii) weakly increasing in q. Going forward, I summarize financier leverage by

$$\lambda_F(p,q) = \frac{1}{p - \mu(p,q)qx_l}.$$

Banks' problem. I study the banks' problem under the presumption that banks earn intermediation rents $\left(r_b(q) < \hat{Y}\right)$ and thus want to issue as many bonds as possible. Banks may not exert monitoring effort in equilibrium. I therefore let $e \in \{0,1\}$ denote the banks' monitoring action, with $e = 1$ if the bank exerts effort. The private benefit associated with e is $m^*(e) = (1-e)m$, and the associated return of the risky technology is $Y_z^*(e) = eY_z + (1-e)y_z$. Given e, the bank's consumption in state z is

$$c_B(z, \underline{a}, a_B, k, s_B, b_B, e) \equiv \max\{s_B + Y_z^*(e)(k - a_B) - b_B + p(\underline{a})a_B, 0\},$$

where s_B, k, a_B and b_B denote the bank's storage, risky investment, asset sales and bond issuance, respectively. Consumption is non-negative because all borrowing is collateralized only by the banks' assets. Bank utility in state z is

$$u_B(z, \underline{a}, a_B, k, s_B, b_B, e) \equiv c_B(z, \underline{a}, a_B, k, s_B, b_B, e) + m^*(e)k.$$

The bank's optimal monitoring choice conditional on (a_B, k, s_B, b_B) is:

$$e^*(a_B, k, s_B, b_B) = \arg\max_{e' \in \{0,1\}} \mathbb{E}_z u_B(z, \underline{a}, a_B, k, s_B, b_B, e') \tag{3}$$

Asset sales, in turn, must be ex-post optimal given $(\underline{a}, k, S_B, b_B)$. Moreover, the bank takes into account that the optimal monitoring decision depends on a_B.Formally, optimal asset sales are given by

$$a_b^*(\underline{a}, k, s_B, b_B) = \arg\max_{k \geq a_B \geq \underline{a}} \mathbb{E}_z \left[\max\left\{s_B + Y_z^* R(e^*)\left(k - a_B\right) - b_B + p(\underline{a})a_B, 0\right\}\right] + m^*(e^*)k. \tag{4}$$

Given these preliminaries, the bank's problem is

$$\max_{s_B, k, b_B, \underline{a}} \quad \mathbb{E}_z \left[\max\left\{s_B + Y_z^*(e^*)\left(k - a_B^*\right) - b_B + p(\underline{a})a_B, 0\right\}\right] + m^*(e^*)k$$

$$\text{s.t. } s_B + k \leq w_B + q b_B, \tag{5}$$

$$b_B \leq s_B + Y_z^*(e^*)\left(k - a_B^*\right) + p(\underline{a})a_B \text{ for all } z, \tag{6}$$

where a_B^* and e^* denote optimal choices in accordance with (3) and (4), and (5) and (6) are the bank's budget and solvency constraint, respectively. I analyze this problem in steps. I first derive the decision rule that determines whether the bank's ex-post optimal asset sales are consistent with monitoring effort. Second, I derive the optimal asset sale promise $\underline{a}$, investment k and bond issuances b_B, and show how these are shaped by an endogenous collateral constraint. The analysis is simplified by the observation that banks will never invest in storage because the risky technology offers a strictly higher return ($\hat{Y} > \bar{R}$). Hence, I take as given that $s_B = 0$ throughout.

Ex-post optimal asset sales. I now characterize the optimal asset-sale policy. Recall that the bank chooses $e = e^*(a_B, \underline{a}_B, k, b_B)$ when he sells a_B risky claims. Monitoring effort is thus conditional on asset sales, and we must worry about "double deviations" in which the bank sells more assets than promised and shirks as a result. Two observations simplify the analysis. First, the objective function is linear because banks are risk neutral and (6) imposes solvency in every state of the world. As a result, the solution is bang-bang, and the bank either sells everything ($a_B = k$) or just as much as initially promised ($a_B = \underline{a}$). Second, the bank will certainly shirk when he sells his entire portfolio ($e^*(k, \underline{a}_b, k, b_B) = 0$).

For there to be monitoring in equilibrium, it must therefore be the case that banks monitor when they sell just as much as they had promised. I assume that this is the case and study the appropriate restrictions below. Bankers then either sell $\underline{a}$ and monitor or sell k and shirk. The payoffs to *shirking and selling* is $p(\underline{a})k - b_B + m$. The payoff to *exerting effort and holding* is $\hat{Y}(k - \underline{a}) - b_B + p(\underline{a})\underline{a}$. Selling assets thus substitutes the market return $p(\underline{a})$ for the technological return $\hat{Y}$ for $k - \underline{a}$ assets. Comparing payoffs yields the following decision rule.

Proposition 1 (Ex-Post Optimal Asset Sales and Monitoring)**.**

Assume that monitoring is optimal at $\underline{a}$. Then the bank sells $\underline{a}$ assets and monitors only if

$$p \leq \bar{p}(k, \underline{a}) \equiv \hat{Y} - m\left(\frac{k}{k - \underline{a}}\right) \tag{7}$$

The proposition states that banks sell everything and shirk when the market return is too high. The market thus fails to prevent the "double deviation" to selling and shirking when asset prices are excessive. Notably, $\bar{p}(k, \underline{a}) < \hat{Y}$, which implies that the upper bound is low enough that it is profitable to buy high-quality assets at $\bar{p}(k, \underline{a})$. A sufficiently rich financier sector may therefore bid up *equilibrium* asset prices until the upper bound is reached.

Collateral constraints. The next step is to characterize the bank's optimal choice of bonds b_B and investment k, taking as given the asset-sale promises $\underline{a}$ and the constraint that banks must monitor. The previous section argued that banks monitor only if $a_B = \underline{a}$. I take this as given and then verify whether it is the case in equilibrium. There are two constraints that limit banks' ability to issue bonds. The first is the solvency constraint (6) stating that the bank must be able to repay his debts in full in every state of the world. The second is the incentive constraint that ensures banks prefer monitoring over shirking,

$$\sum_z \pi_z \left[Y_z \left(k - \underline{a}\right) - b_B + p(\underline{a})\underline{a}\right] \geq \sum_z \pi_z \left[\max\left\{y_z \left(k - \underline{a}\right) - b_B + p(\underline{a})\underline{a}, 0\right\}\right] + mk. \tag{8}$$

This incentive constraint binds before the solvency constraint because shirking generates more downside risk than monitoring ($y_l < Y_l$). Let $\omega_z \equiv y_z \left(k - \underline{a}\right) + p(\underline{a})\underline{a}$ denote the banker's *cash-on-hand* in state z conditional on shirking. Accordingly, the banker can repay his debt in full conditional on shirking only if $\omega_z \geq b_B$. Whether this is the case depends on how many assets he sells. If $\underline{a}$ is such that $\omega_z < b_B$, then the incentive constraint is equivalent to a borrowing constraint of the form

$$b_B \leq \bar{b}_B(k, \underline{a}) = \left[\frac{\pi_h}{\pi_l}\left(Y_h - y_h\right) + Y_l - \frac{m}{\pi_l}\right]k + \left[p(\underline{a}) - Y_l - \frac{\pi_h}{\pi_l}\left(Y_h - y_h\right)\right]\underline{a}$$

Banks can relax this constraint by selling assets only if shirking represents a risk-shifting problem (Jensen and Meckling (1976)).

Lemma 1 (Risk-shifting Problem)**.**
Asset sales ($a_B > 0$) can increase bank borrowing capacity if and only if

$$y_h > \hat{Y}.$$

The result follows from observing that the asset price is bounded above by $\hat{Y}$. This implies that there exists a $p_a(\underline{a})$ such that the coefficient on $\underline{a}$ in the borrowing constraint is positive if and only if $y_h > \hat{Y}$. Lemma 1 thus holds that the losses from shirking must be sufficiently concentrated in the low state. The intuition is that asset sales serve as a form of insurance – the bank has more capital in the low state but less in the high state. This insurance is valuable if the bank is constrained by lack of capital in the low state. This is the case when the returns of the risky technology are particularly poor in the low state conditional on shirking. In order to simplify the exposition by writing the borrowing constraint in terms of low-state payoffs only, I use the following special case.

Assumption 2.
The returns of the risky technology in the high state are the same under shirking and monitoring:

$$Y_h = y_h$$

In order to obtain easily interpretable closed-form solutions for equilibrium prices and trading behavior, I also sometimes assume that the shirking technology pays off nothing in the low state, i.e. $y_l = 0$. Under this assumption, one can summarize the severity of the moral hazard problem by

$$\tilde{m} \equiv 1 - \frac{m}{\pi_l Y_l} \in (0, 1).$$

This statistic is close to one when the moral hazard problem is not severe (m is close to zero) and close to zero when the moral hazard problem is severe (m is close to the output loss from shirking $\pi_l Y_l$). High values of $\tilde{m}$ therefore indicate a loose bank moral hazard problem. Under Assumption 2, the bank's borrowing constraint can then be stated as a collateral constraint

$$\bar{b}_B(k, \underline{a}) = \tilde{m} Y_l k + (p(\underline{a}) - Y_l)\, \underline{a}. \tag{9}$$

The first term is the collateral capacity of the risky asset itself. It is composed of the worst-case return scaled by the moral-hazard discount factor $\tilde{m}$. The second term is the collateral capacity provided by asset sales. It exceeds that of the risky asset to the extent that $p(\underline{a})$ is larger than the worst-case return. The resulting increase in borrowing capacity translates to an increase in the investment opportunity set of banks. Exploiting the budget constraint shows that total investment is

$$k \leq \bar{k}(\underline{a}) = \lambda_B \left[w_B + (p(\underline{a}) - Y_l) \underline{a} \right]$$

where

$$\lambda_B(q) = \frac{1}{1 - q \tilde{m} Y_l} \tag{10}$$

is bank leverage.

Asset sales thus substitute for bank equity by replacing risky investment with safe asset market returns. But only to a point. If a bank sells so many assets that its cash-on-hand after shirking exceeds its debts ($\omega_z > b_B$), then the incentive constraint can be restated as a skin-in-the-game constraint

$$\underline{a} \leq \tilde{m} k. \tag{11}$$

That is, selling assets beyond the point where it serves to increase pledgeable income in the worst state of the world induces shirking. Banks will thus never issue a promise $\underline{a} > \tilde{m} k$.

In sum, the impact of asset sales on borrowing capacity is as follows. If banks have too little cash on hand after shirking, asset sales alleviate the borrowing constraint by improving the bank's collateral position. If instead banks have sufficient cash on hand after shirking, then asset sales no longer boost borrowing capacity and may instead harm origination incentives.

Optimal promises. Do banks find it optimal to issue a promise to sell assets in order to alleviate borrowing constraints? It depends on the asset price. By issuing bonds, banks earn the leveraged intermediation premium

$$\rho(q) = \lambda_B(q)(\hat{Y}q - 1) \tag{12}$$

According to (9), the collateral capacity of risky assets itself is $Y_l\tilde{m}$. In the absence of asset sales, this allows the bank to earn a return of $\lambda_B(\hat{Y} - \tilde{m}Y_l)w_B$. The benefit of selling assets at p is that banks generate $p - Y_l$ units of additional collateral that can be levered to earn $\rho(q)$. The cost is that risky assets trade at the discount $\hat{Y} - p$. The utility impact of asset sales can therefore be summarized using the indirect utility function conditional on $\underline{a}$,

$$u_B(q, p, \underline{a}) = \lambda_B(q)(\hat{Y} - \tilde{m}Y_l)w_B + \underline{a}\left[\rho(q)\left(p(\underline{a}) - Y_l\right) - \left(\hat{Y} - p(\underline{a})\right)\right].$$

The first-order condition with respect to $\underline{a}$ is

$$u_B'(q, p, \underline{a}) = \left[\rho(q)\left(p(\underline{a}) - Y_l\right) - \left(\hat{Y} - p(\underline{a})\right)\right] + p'(\underline{a})\underline{a}\left[\lambda(\hat{Y}q - 1) + 1\right]. \tag{13}$$

The second term reflects the price impact of changes in the promise $\underline{a}$, because promises may be a signal of asset quality. An instructive special case is when the price is constant on the relevant range $[0, \tilde{m}k]$. In this case, banks are willing to sell assets ($u' \geq 0$) if

$$p(\underline{a}) \geq \underline{p}(q) = \frac{\hat{Y} + \rho(q)Y_l}{1 + \rho(q)}. \tag{14}$$

That is, trade occurs only if prices exceed a lower bound $\underline{p}(q)$. This lower bound is strictly decreasing in q since intermediation rents $\rho(q)$ are strictly increasing in q, and $\underline{p}\left(\frac{1}{\hat{Y}}\right) = \hat{Y}$ because $\rho\left(\frac{1}{\hat{Y}}\right) = 0$. Banks are thus willing to sell at deep discounts if intermediation rents are large, but are willing to sell at no less than par if there are no rents. The magnitude of intermediation rents thus generates the scope for asset trade and shapes the pass-through of rents to financiers. The same lower bound obtains when evaluating (13) at $\underline{a} = 0$. Given q and an equilibrium promise $\underline{a}$, a necessary condition for trade to occur on sub-market $\underline{a}$ is that the asset price lies in the interval $\mathcal{P}_{\underline{a}}(q) = [\underline{p}(q), \bar{p}(k(q), \underline{a})]$. In the presence of wealth constraints, the asset price moves within this interval as a function of the relative wealth of financiers and bankers. The bond price in turn lives on the interval $\left[\frac{1}{\underline{R}}, \frac{1}{\hat{Y}}\right]$, which q growing large when savers are wealthy relative to intermediaries. I therefore classify equilibrium outcomes using the following definitions.

Definition 1 (Asset Market Tightness)**.**
Given q and $\underline{a}$, the asset market is ***slack*** *if $p(\underline{a}) = \underline{p}(q)$, and* ***tight*** *if $\underline{p}(q) \in \left(\underline{p}(q), \bar{p}(k(q), \underline{a})\right)$. There is* ***excessive demand*** *if $p(\underline{a}) = \bar{p}(k(q), \underline{a})$.*

Definition 2 (Intermediaries' Borrowing Capacity)**.**
Financial intermediaries are ***highly constrained*** *if $q = \frac{1}{\underline{R}}$,* ***constrained*** *if $q \in \left(\frac{1}{\underline{R}}, \frac{1}{\hat{Y}}\right)$, and* ***unconstrained*** *if $q = \frac{1}{\hat{Y}}$.*

In a slack asset market, all trading rents accrue to financiers, while banks retain some rents when asset markets are tight. Similarly, all bond market rents accrue to intermediaries when they are highly constrained, but some returns are passed on to savers when they are merely constrained. There may be excessive demand even if the asset market is slack: the former is a statement about aggregate quantities related to market clearing in the presence of price bounds, the latter slackness pertains to the pass-through of rents among intermediaries.

1.3 Equilibrium

I now turn to characterizing the competitive equilibrium. Whether the asset price is below its upper bound cannot be verified ex-ante. I therefore use a guess-and-verify approach and first construct an *efficient monitoring equilibrium* in which the constraint $p \leq \bar{p}$ is presumed to hold. As a result, all banks exert effort. I then verify whether this is indeed the case. If it is not, I construct an *excessive trading equilibrium* in which a fraction of banks shirks. Finally, I show why the competitive equilibrium must involve shirking if financier wealth is above a threshold.

Efficient monitoring equilibrium. Asset prices in efficient monitoring equilibrium must be determined by financier wealth ("cash in the market"). Else, financiers would bid up prices until $p = \hat{Y}$, violating the asset price's upper bound (7). Given that all firms monitor and pricing is cash-in-the-market, I therefore restrict attention to price schedules in which the asset price is constant for all promises at which banks have an incentive to monitor, $p(\underline{a}) = p$ for all $\underline{a} \in [0, \tilde{m}k]$. Since bankers who issue a promise greater than $\tilde{m}k$ shirk for sure, let $p = 0$ for all $\underline{a} > \tilde{m}k$. Conditional on the asset price, the banks' decision problem is simple. If the asset market is tight ($p > \underline{p}(q)$), then banks earn strictly positive rents by increasing borrowing capacity through asset sales. As a result, they sell as many assets as possible, $\underline{a}^* = \tilde{m}k^*$.[3] The bank's portfolio thus is

$$k^* = \frac{w_B}{1 - q\tilde{m}p}, \quad b_B^* = \frac{\tilde{m}pw_B}{1 - q\tilde{m}p} \quad \text{and} \quad \underline{a}_B^* = \frac{\tilde{m}w_B}{1 - q\tilde{m}p}.$$

Comparing the expression for k to the definition of bank leverage in (10), asset sales thus boost bank's effective leverage by substituting the safe cash flow p for the worst-case return Y_l. Naturally, banks invest more when the asset price is higher. The degree to which the bank can rely on outside collateral is determined by $\tilde{m}$. Banks require little skin-in-the-game when $\tilde{m}$ is large, and can thus fund their investments largely using financier funds when this is the case. As a result, they sell a large fraction of their investments off to financiers. All agent's policy functions are linear in wealth, permitting straightforward aggregation. As a result, bond and asset market clearing conditions are

$$\frac{\tilde{m}pw_B}{1 - q\tilde{m}p} + \frac{\mu(q,p)Y_l w_F}{p - \mu(q,p)qY_l} = \frac{w_S}{q} \qquad \text{and} \qquad \frac{\tilde{m}w_B}{1 - q\tilde{m}p} = \frac{w_F}{p - qY_l},$$

respectively, where I have used that $x_z = Y_z$ because all bankers are presumed to exert monitoring effort. Conditional on q, the asset price thus is a function of relative wealth $\omega = \frac{w_F}{w_B}$ only,

$$p = \frac{1}{\tilde{m}} \left(\frac{\tilde{m}qY_l + \omega}{1 + q\omega} \right).$$

[3] Note that this quantity of asset sales simultaneously ensures that the banks cash-on-hand upon shirking is equal to b_B, the face value of its debt, in the low state. We have already seen that this is the point beyond which asset sales no longer boost borrowing capacity.

If instead asset markets are slack ($p = \underline{p}(q)$), banks are indifferent toward selling assets. Asset sales are thus determined by financier demand within the interval $[0, \tilde{m}k]$, $\underline{a}^* = a_F$. Accordingly, the bank portfolio is

$$k^* = \frac{w_B + q\left(\underline{p}(q) - Y_l\right) a_F}{1 - q\tilde{m}Y_l} \quad \text{and} \quad b_B^* = \frac{Y_l \tilde{m} w_B + (\underline{p}(q) - Y_l)a_F}{1 - q\tilde{m}Y_l},$$

and the bond market clearing condition is

$$\frac{Y_l \tilde{m} w_B + (\underline{p}(q) - Y_l)a_F}{1 - q\tilde{m}Y_l} + \frac{\mu(q, \underline{p}(q))Y_l w_F}{\underline{p}(q) - \mu(q, \underline{p}(q))qY_l} = \frac{w_S}{q}, \qquad \text{where} \qquad a_F = \frac{w_F}{\underline{p}(q) - \mu(q, \underline{p}(q))qY_l}.$$

Whether the market is slack or tight depends on the relative wealth of financiers and bankers. Specifically, the market is slack only if $a_F \leq \tilde{m}k$ conditional on $p = \underline{p}(q)$. This is the case if finanicers are sufficiently poor relative to bankers,

$$\omega \leq \frac{\lambda_B}{\lambda_F}\left(\frac{1}{1 - \lambda_B q(\underline{p}(q) - Y_l)}\right)$$

Accordingly, intermediation rents are fully passed through to financiers when w_F is small.

The key comparative statics of prices with respect to the wealth distribution are straightforward: q is increasing in w_S and decreasing in w_B, while p is increasing in w_F. These observations have immediate implications for the existence of efficient monitoring equilibrium with asset market trading.

Proposition 2.
1. For any w_S, there exists a cutoff $\bar{w}_B(w_S) \geq 0$ such that no efficient monitoring equilibrium with asset trade exists if $w_B > \bar{w}_B(w_S)$.
2. For any w_S and $w_B \leq \bar{w}_B(w_S)$, there exists a cutoff $\bar{w}_F(w_S, w_B) > 0$ such that no efficient monitoring equilibrium exists if $w_F > \bar{w}_F(w_S, w_B)$.

The first result follows because q is decreasing in w_B and $\underline{p}(q)$ is decreasing in q. That is, increases in bank wealth drive down intermediation rents and push up the minimum price at which banks are willing to sell assets. Hence, there exists a w_B large enough such that the lower bound $\underline{p}(q^*)$ violates the upper bound $\bar{p}(k^*, 0)$ even when no assets are traded. Banks are thus willing to sell only at prices which, if financiers paid them, would lead banks to shirk. As a result, asset markets break down. Securitization thus requires a sufficiently constrained banking sector.

The second result follows because p is increasing in financier wealth. As a result, the asset price thus breaches the upper bound $\bar{p}$ for sufficiently large w_F. As the next section will show, an excessively large financier sector may therefore lead to equilibrium shirking. The presence of moral hazard and hidden trading thus limits the ability of the price to adjust to clear markets, and excess demand triggers the production of low-quality assets. An immediate implication is that cash-in-the-market pricing is a necessary condition for efficient monitoring. This result contrasts with previous literature that identified cash-in-the-market pricing as a source of instability in models where the set of assets is fixed.

When w_F is *not* excessively large, increases in financier wealth typically boost lending in the aggregate. Indeed, increases in w_F may lead to sharper aggregate credit growth than increases in bank wealth w_B.

Proposition 3.
Assume that $w_B < \bar{w}_B(w_S)$ and $w_F < \bar{w}_F(w_S, w_B)$. Investment k is strictly increasing in w_F if

(i) asset markets are tight, or

(ii) asset markets are slack and intermediaries are highly constrained.

Moreover, $\frac{\partial k^*}{\partial w_F} > \frac{\partial k^*}{\partial w_B}$ if funding is scarce, intermediaries are highly constrained and $\underline{R} < 1$.

The intuition behind the first part of the proposition is straightforward. As financier wealth increases, so does the demand for risky assets. Rising asset prices mean that banks receive more collateral per risky asset sold. Borrowing and investment increase. Since all banks monitor, expected aggregate output also increases. A caveat applies when funding is scarce. In this case, financiers receive all asset market rents, and increases in financier wealth may increase the total supply of bonds more than bank's borrowing capacity. This may lead to a drop in bond prices that crowds out bank borrowing at intermediate levels of w_F.

The third part of the proposition shows that the social benefits of increased financier wealth can be large. Specifically, increases in financier funding spur investment disproportionately when the financial sector is highly constrained ($q = \frac{1}{\underline{R}}$) and savers pay a premium for intermediation services ($\underline{R} < 1$). The reason is that banks are able to lever more than banks when bond prices are high and asset prices are low because they are not subject to banks' moral hazard concern. That is, aggregate lending volumes are boosted by the separation of asset origination and the holding of the resulting risk. The mechanism is limited by the fact that the asset price eventually reaches its upper bound. The next section characterizes equilibrium outcomes when this is the case.

Excessive trading equilibrium. I now turn to characterizing the competitive equilibrium financiers are too wealthy to sustain monitoring by all bankers, $w_F > \bar{w}_F(w_S, w_B)$. The fundamental problem is that the asset price cannot increase beyond $\bar{p}(k, a)$, since all banks would shirk otherwise. The price mechanism thus fails to clear the market. This section's basic argument is that an increase in equilibrium shirking *can* clear the market at fixed prices because (i) banks are indifferent between shirking and effort at $\bar{p}$, (ii) banks who shirk do so because they sell more, raising asset supply, (iii) financiers are willing to tolerate some shirking because they buy at $\bar{p} < \hat{Y}$. The only subtlety is that $\bar{p}(\bar{k}(\underline{a}), \underline{a})$ is a function of banks' endogenous promises $\underline{a}$. We must therefore check whether banks have an incentive to use $\underline{a}$ to signal the quality of their assets.

The first part of the argument holds by construction: given $\underline{a}$, $\bar{p}$ is defined to be the price at which banks are indifferent between (sell $\underline{a}$, effort) and (sell $\bar{k}(\underline{a})$, shirk). The second part follows because those shirk necessarily sell more assets than those who exert effort. Defining $\Phi \in [0, 1]$ to be the fraction of shirking banks (or, alternatively, the probability with which each bank chooses to shirk and sell) excess asset market demand is

$$\eta(\underline{a}, \Phi) = \frac{w_F}{\bar{p}(\bar{k}(\underline{a}), \underline{a}) - qx_b(\underline{a})} - \left[\Phi\bar{k}(\underline{a}) + (1 - \Phi)\underline{a}\right],$$

and is decreasing in Φ. The third part requires that financiers weakly prefer to buy risky assets rather than invest in storage even when Φ banks shirk. Given $\underline{a}$, the fraction of low-quality assets is

$$\phi = \frac{\Phi \bar{k}(\underline{a})}{\Phi \bar{k}(\underline{a}) + (1-\Phi)\underline{a}} \geq \Phi,$$

and the risky asset's expected rate of return is $\frac{\hat{x}}{\bar{p}(\bar{k}(\underline{a}),\underline{a})} = \frac{\phi\hat{y}+(1-\phi)\hat{Y}}{\bar{p}(\bar{k}(\underline{a}),\underline{a})}$. The risky asset is preferable to storage if

$$\phi \leq \bar{\phi}(\underline{a}) = \frac{\hat{Y} - \bar{p}(\bar{k}(\underline{a}),\underline{a})}{\hat{Y} - \hat{y}},$$

and $\bar{\phi}(\underline{a}) > 0$ for all $\underline{a}$ since because $\bar{p}(\bar{k}(\underline{a}),\underline{a}) < \hat{Y}$ for all $\underline{a}$. Financiers are thus willing to tolerate a degree of shirking because they earn rents on average.

If the upper bound $\bar{p}(k(\underline{a},\underline{a})) = \hat{Y} - m\frac{\bar{k}(\underline{a})}{\bar{k}(\underline{a})-\underline{a}}$ binds, prices need no longer be constant in $\underline{a}$ on the relevant interval $[0, \tilde{m}k]$. If this is so, banks may want to shade their promises in order to maximize revenue. Their incentives to do so depend on the price schedule $p(\cdot)$.

I work from the *benchmark excessive trading equilibrium* in which $p(\underline{a})$ *is* constant for $\underline{a} \in [0, \tilde{m}k]$. In this case, banks have no incentive to shade their promises to boost prices, and choose the same optimal portfolio $(b_B^*, k^*, \underline{a}^*)$ as in the efficient monitoring equilibrium. The only difference is that Φ^* banks deviate to (sell $\bar{k}(\underline{a})$, shirk) ex post, with Φ^* determined such that excess asset market demand is zero, $\eta(\underline{a}^*, \Phi^*) = 0$.

This benchmark is natural when the asset market is slack and financiers need not offer a price above $\underline{p}(q)$ at any $\underline{a}$. It is perhaps less convincing when the asset market is tight and there is excess demand at $\bar{p}$. In the presence of wealth constraints and a feedback from prices to incentives, fully specifying the price schedule would require a theory of counterfactual off-the-equilibrium-path market tightnesses for all $\underline{a} \in [0, \tilde{m}k]$. My approach is to circumvent this requirement by asking whether there are profitable local deviations from the benchmark portfolio $(b_B^*, k^*, \underline{a}^*)$ under the best-case scenario that banks always receive the maximum feasible price $\bar{p}(\bar{k}(\underline{a}),\underline{a})$ after a deviation. The following result provides a sufficient condition that ensures banks do not want to shade their promises in order to boost revenues.

Proposition 4.
There is no profitable local deviation to $\underline{a} < \underline{a}^* = \tilde{m}k^*$ if $\frac{\tilde{m}}{1-\tilde{m}} > \frac{1}{q[\pi_l(\pi_l Y_l)+\pi_h m]}$. This condition holds for all q if $\frac{\tilde{m}}{1-\tilde{m}} > \frac{\hat{Y}}{[\pi_l(\pi_l Y_l)+\pi_h m]}$. There exists m sufficiently close to zero such that this condition is satisfied for all π_l and Y_l.

That is, banks do not want to decrease their reliance on outside funding when intermediation rents are large (q is high) and when they can easily collateralize outside funding in order to borrow more ($\tilde{m}$ is large). As a result, the benchmark excessive trading equilibrium obtains even when the asset market is tight, and increases in financier wealth are deleterious.

Proposition 5.
The share of shirking banks Φ is strictly increasing in w_F. If (i) financiers do not borrow or (ii) intermediaries are highly constrained, then a partial destruction of financier wealth from w_F to some $w_F > \bar{w}_F(w_S, w_B)$ is Pareto-improving.

When the price is at its upper bound, markets clear through quantities, and some banks must shirk. More banks shirk the larger is excess demand at $\bar{p}$. A pecuniary externality drives the inefficiency: individual financiers do not internalize that they contribute to a worsening asset pool when buying assets. Under certain conditions, this negative externality is severe enough that financiers can be made better off by a uniform destruction of their wealth. This externality also contributes to excess risk exposure in the aggregate, as low-quality assets carry more downside risk than high-quality assets.

Figure 2 provides a numerical illustration of the equilibrium effects of w_F. Growing financier wealth initially increases investment and expected output, but gradually induces banks to shirk. As a result, investment efficiency falls. Blue corresponds to efficient monitoring equilibrium, while

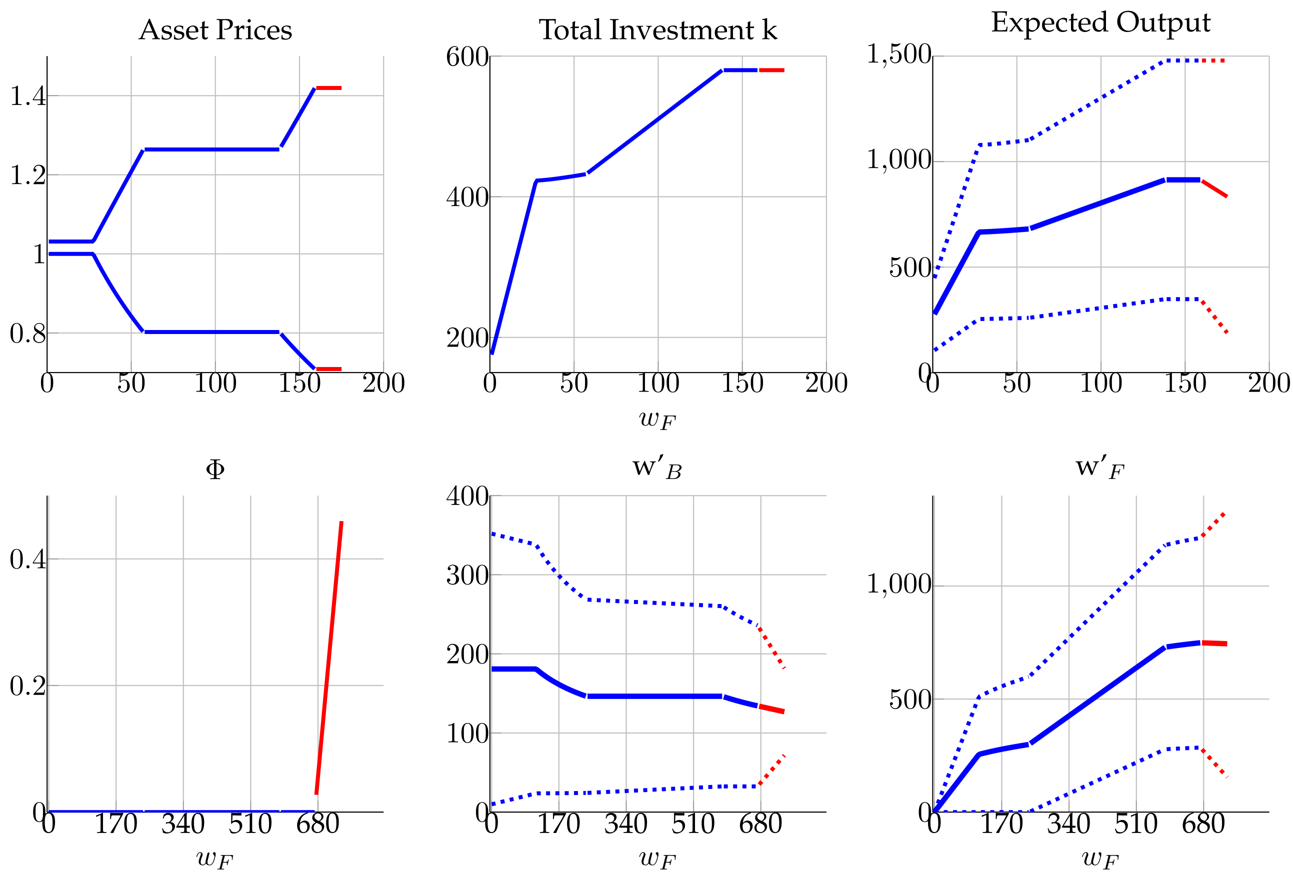

Figure 2: Equilibrium Outcomes as a Function of Financier Net Worth w_F.

red depicts the excess trading equilibrium. The top left panel depicts asset prices, with the upper line representing p and the lower line showing q. Initially, increases in w_F boost investment and expected output. Eventually, the asset price reaches its upper bound, investment no longer increases, and expected output declines because a growing fration of banks shirk (bottom left panel). The top right panel shows the increase in aggregate risk, with the dotted lines depicting aggregate output after a good and a bad shock. Given that low-quality assets are more exposed to downside risk, increases in financier wealth lead to poorer worst-case outcomes. The two last figures in the bottom row show the risk exposure of both classes of financial intermediary. The solid line depicts the *expected* wealth of an intermediary at the end of the period, with the dotted lines corresponding to a high and low aggregate shock, respectively. Financiers take on a greater fraction of total risk exposure as w_F increases. The next section shows that this risk-transfer channel may generate credit volumes with growing w_F in which lending standards gradually fall.

2 Dynamics

I now incorporate the static model into a simple overlapping generations setting to study the endogenous evolution of the wealth distribution. The key question is whether the dynamics of asset market trading are such that wealth distribution may endogenously move towards regions of the state space in which only excessive trading equilibria can be sustained even when starting out in efficient monitoring equilibrium.

Time is discrete and runs from 0 to T. A generic period is indexed by t. The model economy is populated by overlapping generations of financiers and banks, each of whom lives for two periods, and savers who live for one period. I refer to intermediaries in the first period of their life as the *young*, and to those in the second period as the *old*. There are two goods: a consumption good and an intermediary wealth good. Only the consumption good can be consumed. Intermediary wealth is special in that banks and financiers must use wealth in order to intermediate and invest. Every generation of agents is born with an endowment of the consumption good. Only the initial generation of intermediaries are born with an endowment of wealth. When young, intermediaries and depositors play the intermediation game described in the static model. Before consuming their end-of-period wealth, they have the opportunity to sell it to the young. In this intergenerational wealth market, young intermediaries pool their endowment and instruct a market maker to purchase the old intermediaries' equity capital. The market maker then approaches each old intermediary individually and bargains over the equity capital. If a trade is agreed, the old consume the proceeds from the sale and the market maker distributes the equity capital evenly to all young intermediaries. If no trade is agreed, the old consume their equity capital and the young do not receive any equity capital from the old.[4]

I assume throughout that the young have all the bargaining power. The objective function of a young intermediary then is to maximize expected end-of-life wealth. As a result, the dynamic model is equivalent to repeating the static model period-by-period, with the evolution of the wealth distribution linking periods. This setting parsimoniously illustrates how the risk-transfer and shirking channel can generate credit booms with falling asset quality. In Appendix B I consider the case where the old have all the bargaining power, and show that doing so admits the same qualitative dynamics.[5]

All bargaining power to the young. Given that the young have all the bargaining power, the evolution of wealth follows directly from the statically optimal asset portfolios computed above. w_F and w_B increase or decrease jointly because the only source of risk is aggregate. Key to the evolution of credit volumes and asset quality will therefore be the relative wealth of financiers $\omega = \frac{w_F}{w_B}$. Naturally, asset prices and credit volumes can increase only when intermediary wealth grows. I therefore focus on the evolution of ω after good aggregate shocks. The next proposition

[4] The role of the market maker is solely to make sure that each intermediary in every generation starts out with the same wealth. As a result, I do not have to keep track of a wealth distribution for the same type of agents. All results go through without this assumption.

[5] I assume that if a generation of intermediaries has zero wealth at the end of their life, then the new generation receives start-up funds of ϵ_0. This ensures that both types of intermediaries are always active.

characterizes the evolution of relative wealth and the asset price when w_F is already enough for asset markets to be tight.

Proposition 6.
Assume that the wealth distribution is such that asset markets are tight in period t. Then the law of motion for relative wealth is $\omega_{t+1} = \frac{\tilde{m}}{1-\tilde{m}} \frac{Y_h - Y_l}{Y_h}$ *if financiers' borrowing constraints bind in period t, and* $\omega_{t+1} = \frac{\tilde{m}}{1-\tilde{m}}$ *if financiers do not borrow in period t. The asset price increases after a good shock if*

$$\frac{\tilde{m}}{1-\tilde{m}} > (\tilde{m} Y_l)^{\frac{1}{2}} \left(\frac{Y_h - Y_l}{Y_h} \right).$$

For all π_l, Y_h *and* Y_l *there exists* $\tilde{m}$ *sufficiently close to zero such that this condition is satisfied.*

When asset markets are tight, the evolution of relative wealth is therefore shaped by $\tilde{m}$. Perhaps counterintuitively, financiers grow to be large when banks' moral hazard problem is not too severe ($\tilde{m}$ is large). The reason is that large $\tilde{m}$ allow banks to easily use revenues from asset sales as collateral, strengthening the risk-transfer channel. This mechanism is attenuated somewhat when financiers' borrowing constraint is binding and they rely on leverage to purchase assets, since $\frac{Y_h - Y_l}{Y_h} < 1$. Nevertheless, financiers' *absolute* wealth is large enough that good shocks lead to growing asset prices when $\tilde{m}$ is above a treshold. As a result, good aggregate shocks may gradually trigger the shirking channel.

The previous proposition assumed that w_F was large enough for asset markets to be tight. Can financiers' relative wealth grow even when their initial wealth is such that assets markets are initially slack?

Proposition 7.
Define the return on equity earned by banks and financiers in state z when the asset market is slack as $ROE_B(z) = \lambda_B(q)(Y_z - \tilde{m} Y_l)$ *and* $ROE_F(z) = \lambda_F(\underline{p}(q), q)(Y_z - Y_l)$*, respectively. If asset markets are slack, relative wealth* ω *strictly increases after a good shock if*

$$ROE_B(h) - ROE_F(h) < (1-\tilde{m})\lambda_F(\underline{p}(q), q) Y_l \left(\frac{Y_h - \hat{Y}_R}{\hat{Y} - Y_l \tilde{m}} \right) .\omega \tag{15}$$

This inequality is satisfied for any $\omega > 0$ *if* $q \geq 1$.

The right-hand side of (15) is equal to the degree of risk transfer from banks to financiers. Given that the asset market is slack, risk transfer is demand-determined and thus decreasing in current relative wealth. Financier wealth grows even if it is vanishingly small to begin with as long as $\text{ROE}_F(h) > \text{ROE}_B(h)$. Since financiers are not subject to moral hazard, they can leverage disproportionately at low interest rates. This advantage allows them to earn higher returns on equity when $q \geq 1$. Since q is strictly increasing in w_D, the model therefore predicts that financiers begin growing when there is sufficient demand for financial services even when they are very small to begin with. Caballero, Farhi, and Gourinchas (2008) and Krishnamurthy and Vissing-Jorgensen (2015) argue that this demand existed in the run-up to the 2008 financial crisis.

More generally, low interest rates substitute for financier wealth, and financiers grow when they can borrow cheaply or when they are rich to begin with. Figures 3 and 4 depict the importance of initial conditions graphically. I plot the evolution of financier and bank wealth after a sequence of positive aggregate shocks. In both figures, the left panel depicts a baseline scenario in which financier wealth is smaller than bank wealth initially, but grows to be larger over time. The right panel depicts deviations from this baseline. Figure 3 shows the effect of a reduction in initial financier wealth. This reduction leads to less risk being transferred to financiers. As a result, financier wealth no longer catches up with bank wealth. Figure 3 shows the effect of a reduction in

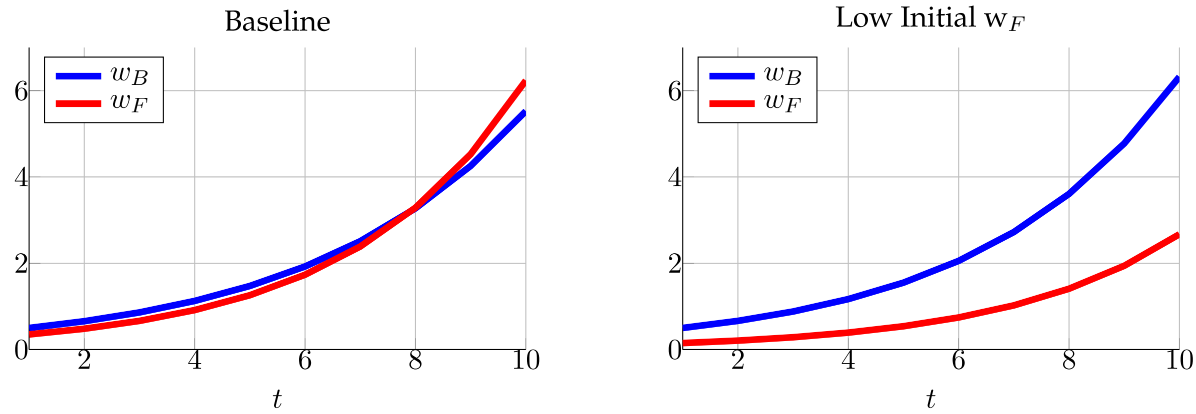

Figure 3: The effects of initial conditions – reduction in initial financier wealth w_F^0. Baseline parameter values: $\pi_h = 0.8, Y_l = 0.5, Y_h = 1.2, \tilde{m} = 0.82$. Initial wealth distribution: $(w_S, w_B^0, w_F^0) = (25, 0.5, 0.35)$. Comparative static: w_F^0 from 0.3 to 0.15.

depositor wealth. Lower depositor wealth causes a fall in the equilibrium bond price. The resulting decrease in financier leverage induces a disproportionate fall in financiers' return on equity and total purchases of risky assets. Given suitable initial conditions, the model can therefore give

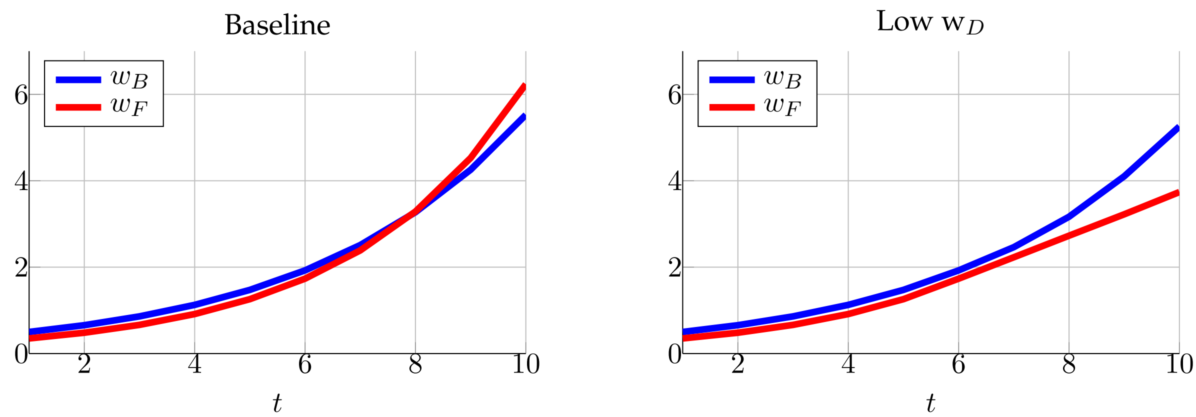

Figure 4: The effects of initial conditions – reduction in depositor wealth w_S. Baseline parameter values: $\pi_h = 0.8, Y_l = 0.5, Y_h = 1.2, \tilde{m} = 0.82$. Initial wealth distribution: $(w_S, w_B^0, w_F^0) = (25, 0.5, 0.35)$. Comparative static: w_S from 25 to 5.

rise to *destabilizing credit booms* in which growing credit volumes and the increased transfer of risk to financiers gradually give rise to the incentive channel and falling credit quality.

Characteristics of credit booms. In this section, I now characterize the properties credit booms. I do so by computing equilibrium outcomes as a function of a time path for the exogenous shock z

and the initial wealth distribution $\mathbf{w}^0 = (w_S^0, w_B^0, w_F^0)$. I simulate the economy for T periods. The initial T_{boom} shocks are good shocks. The next T_{crisis} shocks are negative. The remaining shocks are good.

Figure 5 depicts a typical destabilizing credit boom. I simulate the economy for 11 periods. There is an initial period, 8 positive shocks, a single negative shock, and then another positive shock. Financiers and banks each start out with 0.5 units of wealth. Initial conditions are such that the economy starts out in a efficient monitoring equilibrium. The left panel plots the evolution of wealth over time. There is a rapid build-up of wealth in the aggregate, with financiers growing faster than banks. When a negative shock occurs, financier wealth collapses sharply because financiers are disproportionately exposed to risk. Bank wealth drops only moderately because financiers provide partial insurance to banks. The middle panel plots the evolution of investment over time. The blue line depicts total investment and the red line depicts the fraction of investment that goes to low-quality projects because banks begin to shirk. Initially all banks exert effort and there is no investment in low-quality projects. Over time, however, continued financier growth pushes the economy into an excessive trading equilibrium. As a result, the fraction of investment that flows to low-quality projects increases steadily during the boom. In the aftermath of the crisis, investment falls. The right panel plots the evolution of output. The solid line depicts actual output in the model economy. The dashed line depicts output in a fictitious economy in which capital accumulation is unaltered but all banks are forced to exert effort. During the boom phase, output increases steadily. During the crisis, output collapses sharply. As the comparison between the solid and dashed lines shows, however, almost one third of the drop is accounted for by falling credit quality over the course of the boom. Excessive asset trading can therefore generate credit booms that end in sharp crises.

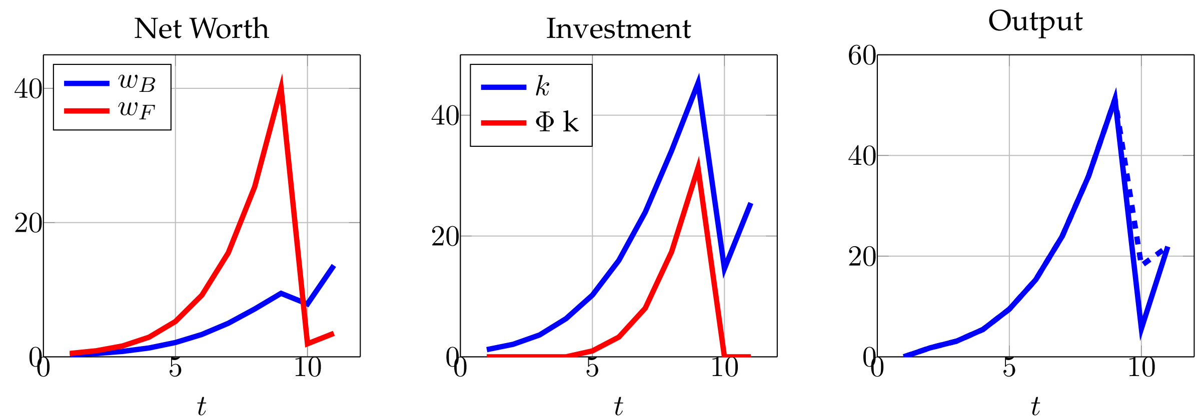

Figure 5: Credit Boom - low $\tilde{m}$. Parameter values: $\pi_h = 0.65, Y_l = 0.4, Y_h = 1.5, \tilde{m} = 0.82$. Initial wealth distribution: $(w_B^0, w_F^0) = (0.5, 0.5)$. Depositor wealth: $w_S = 450$.

Figure 6 shows the importance of the moral hazard parameter $\tilde{m}$ for the dynamics of secondary market booms. While I set $\tilde{m} = 0.82$ in Figure 5, I now set $\tilde{m} = 0.85$. Recall that larger values of $\tilde{m}$ mean that the bank's moral hazard problem is less severe and banks can leverage each unit of wealth more. The time path of aggregate shocks and initial conditions are the same for both simulations. Three observations stand out. First, financier wealth grows *faster* when $\tilde{m}$

is large. This is perhaps counterintuitive given that increases in $\tilde{m}$ allow *banks* to lever more. In equilibrium, however, the portfolios of banks and financiers are intertwined. When $\tilde{m}$ is high, the shadow value of collateral is high for banks. When bank wealth is scarce, banks increase their collateral position by selling claims on asset markets. Increases in $\tilde{m}$ thus boost supply and reduce asset prices. This allows financiers to lever more and purchase more risk exposure. In equilibrium, increased *potential* bank leverage may disproportionately boost *realized* financier leverage. Conditional on a good shock, financiers thus grow faster than banks.

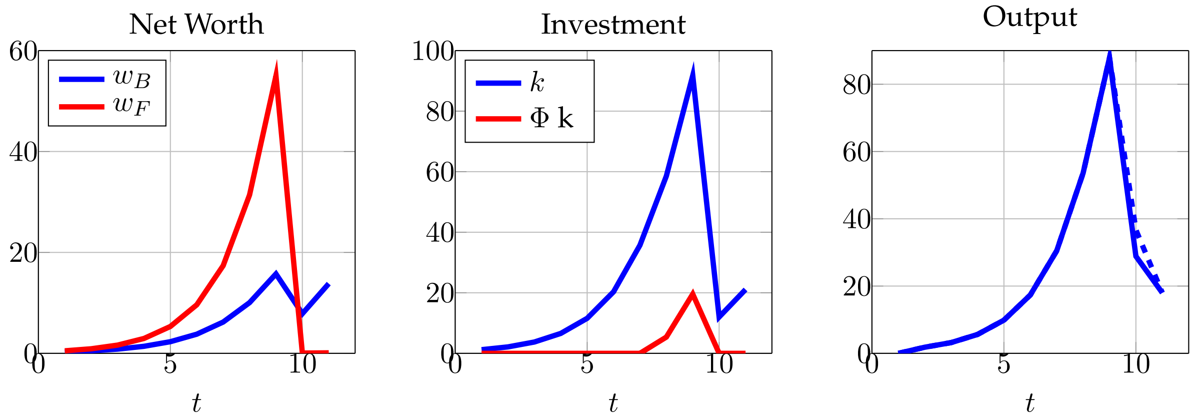

Figure 6: Credit Boom - high $\tilde{m}$. Parameter values: $\pi_h = 0.65, Y_l = 0.4, Y_h = 1.5, \tilde{m} = 0.85$. Initial wealth distribution: $(w_S^0, w_B^0, w_F^0) = (450, 0.5, 0.5)$.

Second, aggregate investment also increases faster because banks can lever each unit of collateral acquired on secondary markets by more. Third, increased supply of secondary market assets means that the fraction of low-quality loans is lower and so investment efficiency is higher. Moreover, financiers borrow *more* when $\tilde{m}$ is high because asset prices are relatively low. Because banks sell off more risk exposure when $\tilde{m}$ is high, they suffer less in a crisis, and financier wealth declines disproportionately. Nevertheless, looser bank constraints allow output and investment efficiency to increase throughout even as volatility grows.

Next, I turn to the effects of boom *duration*. Figure 7 plots two simulated time paths for identical parameters and initial conditions. The only difference is the timing of the negative shock. Solid lines depict the case where the negative shock hits in period 9, while dashed lines depict the case where the negative shock hits in period 8. The left panel shows the evolution of intermediary wealth. The middle panel plots total investment and low-quality investment. The right panel plots output. Two observations stand out. First the fraction of low-quality investment is increasing in the duration of the boom, as is the relative wealth of financiers. Second, the decline in output is increasing in duration – the peak is higher and the trough is lower. Longer booms generate deeper recessions because of increased origination of low-quality credit.

Finally, I study how the number of bad shocks hitting the economy shape the evolution of wealth and the recovery from a crisis. Specifically, Figure 8 plots two simulations that differ only in the number of negative aggregate shocks that hit the economy. The solid line depicts a simulation in which a single negative shock hits in period 8. The dashed line depicts a simulation in which there are negative shocks in period 8 and 9. The left panel depicts the evolution of wealth,

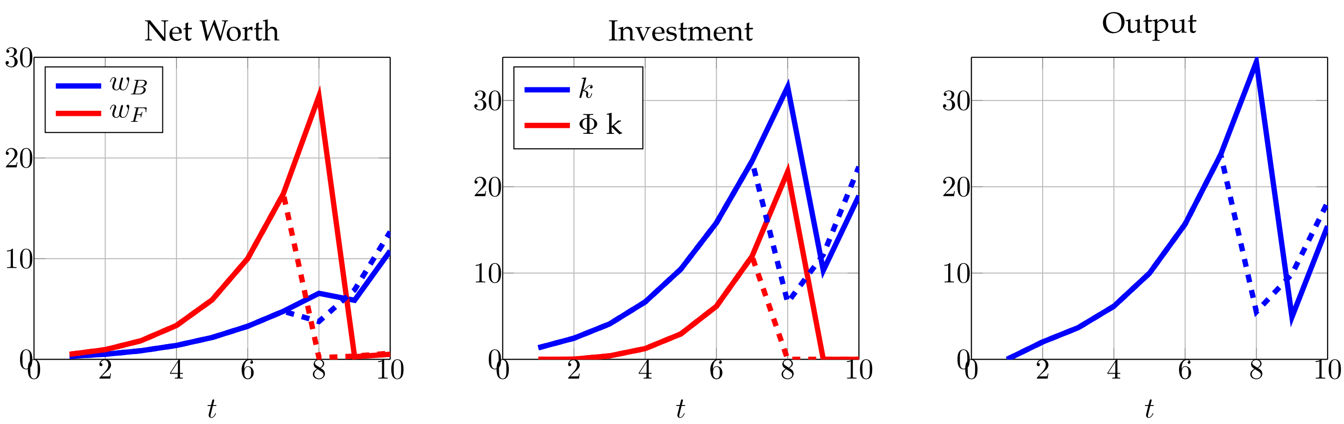

Figure 7: **Effects of increased boom length** - negative shock in period 9 (solid) vs. negative shock in period 8 (dashed).

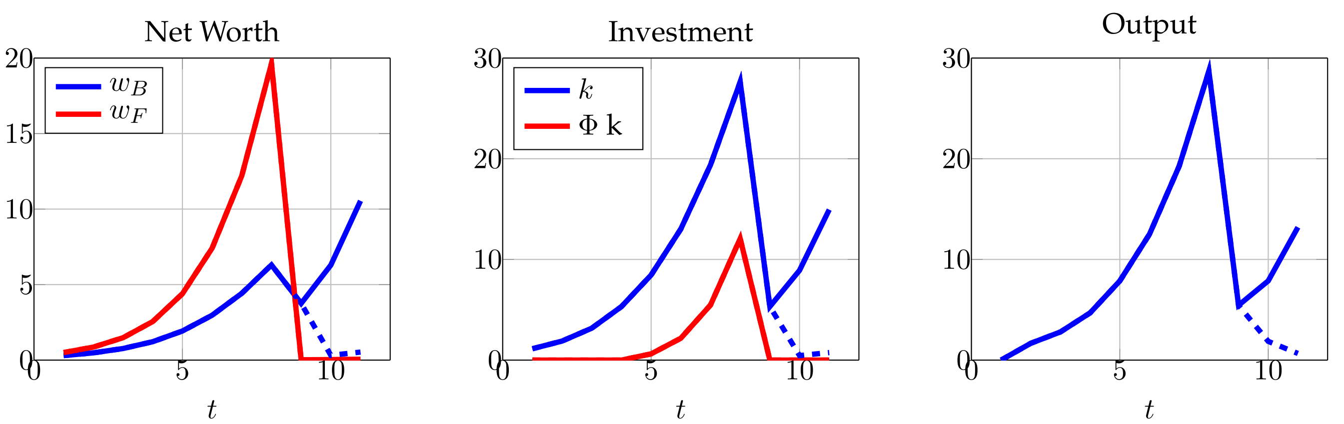

Figure 8: **Effects of increased crisis duration** - negative shock in period 8 (solid) vs. negative shock in periods 8 *and* 9 (dashed).

and shows how the model generates the migration of risk exposure back onto bank balance sheets once the initial negative shock has depleted financier wealth. In particular, the second negative shock leads to a dramatic fall in bank wealth, even as total investment falls. This is consistent with the evidence in Krishnamurthy, Nagel, and Orlov (2014) that credit conditions were poor in the aftermath of the 2008 financial crisis because banks had to carry more risk exposure on their balance sheets. The second negative shock can be thought as representing the endogenous amplification of the initial shock through the real side of the economy. This could be due to foreclosure externalities in housing markets or deteriorating labor market conditions that force increased defaults among outstanding loans.

Relative returns on equity during booms. The bad incentive channel is driven by a growing wealth imbalance between banks and financiers. Do banks and financiers have incentives to "correct" these imbalances by re-allocating equity across intermediaries? If issuing (inside) equity were costless, this would be the case whenever the equilibrium return on financier equity is below that of banks. The next proposition shows that the model can generate destabilizing secondary market credit booms even when financiers always receive higher returns on equity than banks.

Proposition 8.
There exist parameters and values for w_S such that the model generates credit booms with falling asset quality in which (i) relative financier wealth ω grows, (ii) financiers earn a higher expected return on equity, $\hat{ROE}_F > \hat{ROE}_B$ and (iii) financiers optimally choose not to lend to banks.

The intuition is that the harmful effects of secondary markets arise as a function of the imbalance between banks and financiers, while the *rents* accruing to both intermediaries are partially determined by depositor's demand for financial services. Because financiers benefit disproportionately from low interest rates, one can always find a level of saver wealth w_S such that financiers receive higher rents than banks as long as some parametric conditions are satisfied. As a result, the model's results are robust to allowing for endogenous equity issuances or letting banks issue debt to financiers.

3 Policy

I now study how policy shapes the likelihood and evolution of credit booms. I study the positive implications of three prominent policy approaches: monetary policy as a determinant of short-term interest rates, restrictions on bank leverage, and macro-prudential to eliminate pecuniary externalities in asset markets.

Monetary policy. I study monetary policy in reduced form by assuming that the short-term policy rate affects savers' return to storage. That is, $\underline{R} = M(\rho)$, where ρ denotes the monetary policy environment and $M'(\rho) > 0$. Monetary policy thus works through affecting the *required return on deposits*. This is in line the evidence in Krishnamurthy and Vissing-Jorgensen (2012) that treasuries are valued for their safety by risk-averse investors and are thus a substitute for safe assets produced by the financial system. Since $q \leq \frac{1}{\underline{R}}$ in equilibrium, the monetary policy environment places a lower bound on funding market interest rates, $r_b \geq \underline{r}_b(\rho) \equiv \frac{M(\rho)-1}{M(\rho)}$. Since $\underline{r}_b'(\rho) > 0$, I use ρ to denote the *tightness* of monetary policy. To the extent that equilibrium interest rates are at their lower bound, tight monetary policy raises interest rates.

For monetary policy to have bite, bond prices must be at their lower bound. I therefore assume that the financial system is highly constrained. To understand whether expansionary monetary policy can trigger destabilizing credit booms when asset markets are small, I assume that asset markets are slack initially. The combination of these two assumptions implies that all banks exert effort.

Proposition 9 (Monetary Policy and Secondary Market Booms)**.**
Fix a efficient monitoring equilibrium in which asset markets are slack. Then a loosening of monetary policy (a reduction in ρ) increases investment and the growth rate of relative financier wealth after a good shock.

Due to asymmetric leverage constraints, loose monetary policy biases the growth rates of intermediary wealth towards financiers, even as both banks and financiers can borrow at cheaper rates. As a result, expansionary monetary policies can contribute to the build-up of financial fragility over time by encouraging imbalances in the distribution of wealth in the financial system, leading to a *dynamic* risk-taking channel of monetary policy. The reason is that initially low interest rates set the economy on a path towards excessive growth in relative financier wealth, which ultimately manifests itself in deteriorating monitoring incentives and increased risk-taking. Altunbas, Gambacorta, and Marques-Ibanez (forthcoming) provide evidence for these precise dynamics: extended periods of loose monetary policy are associated with increased risk-taking and higher default risk among financial institutions, but with a lag.

Importantly, short-lived monetary impulses may have persistent effects. Proposition 6 showed that financier wealth is a substitute for low interest rates in generating future increases in ω: financiers grow faster when they are already large. An initial monetary policy boost to financier wealth may then mean that financiers continue to grow when the crutch of low interest rates is removed. Halting a secondary market boom by "taking away the punch bowl" is difficult if financiers have already stashed away the punch.

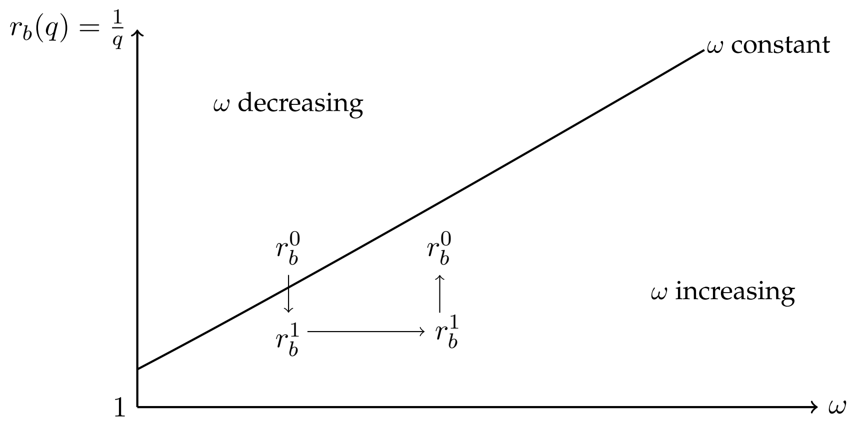

Figure 9: **Persistent effects of monetary policy** - iso-ω curve in (q,ω) space. Monetary policy shift from r_b^0 to r_b^1 to r_b^0.

Figure 9 depicts this concern graphically. The solid line depicts all combinations of (q, ω) such that ω is constant over time. A policy that lowers the risk-free rate from r_b^0 to r_b^1 may set financiers on a path towards growth that cannot be halted by reversing the policy action.

Caps on bank leverage. Next, consider bank leverage caps that restrict banks from investing more than a fixed multiple of their wealth:

$$k \leq \bar{\lambda}_B w_B.$$

For simplicity, I focus on regions of the state space where asset markets are tight in the absence of capital requirements. Recall that bank leverage in the absence of secondary markets and capital requirements is

$$\lambda_B^0 = \frac{k^0}{w_B} = \frac{1}{1 - q^0 Y_l \tilde{m}}.$$

while equilibrium leverage in the absence of leverage caps is

$$\lambda_B^* = \frac{k}{w_B} = \frac{1}{1 - q^* p^* \tilde{m}}.$$

For leverage requirements to influence equilibrium outcomes without shutting down secondary markets altogether, I assume that

$$\lambda_B^0 < \bar{\lambda}_B < \lambda_B^*.$$

In the presence of leverage caps, banks sell just enough assets to exactly hit the leverage constraint. A binding leverage constraint therefore acts as a negative supply shock in the asset market, boosting asset prices. High asset prices feed back into bad origination incentives. Accordingly, the next proposition shows that binding leverage constraints may directly harm asset quality.

Proposition 10.
Assume that the financial system is highly constrained whether or not capital requirements are binding. Suppose that the competitive equilibrium without leverage constraints is an efficient monitoring equilibrium with tight asset markets, and let λ_B^* denote the associated bank leverage. If $\bar{\lambda}_B < \lambda_B^*$, then a strictly positive fraction of banks must shirk under the leverage cap.

In Figure 10, I plot equilibrium outcomes as a function of the capital requirement $\bar{\lambda}_b$ in an example economy in which secondary market liquidity is high in the absence of capital requirements. The top-left panel shows that asset sales fall as the leverage cap shrinks. The next two

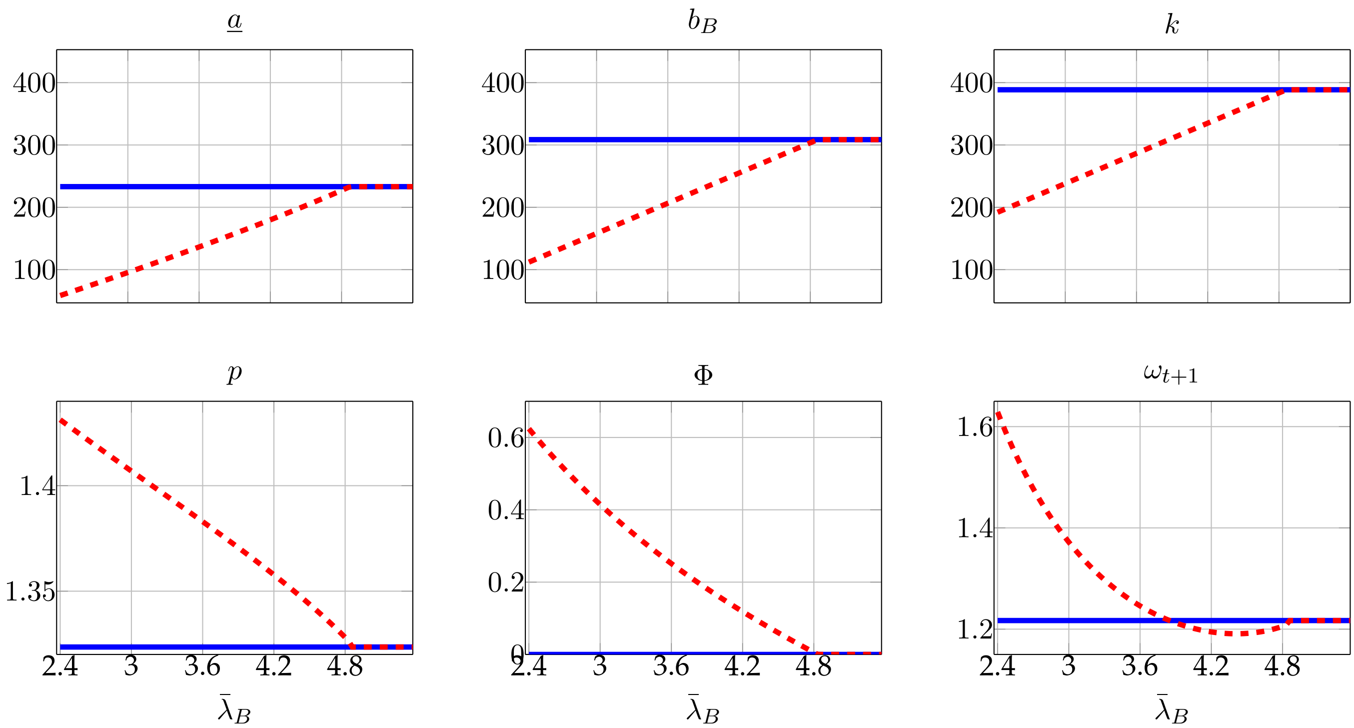

Figure 10: Equilibrium outcomes as a function of the bank capital requirement $\bar{\lambda}_B$. The unconstrained equilibrium is depicted in solid blue, the equilibrium with capital requirements in dashed red. Leverage in the unconstrained equilibrium is equal to $\lambda_B^* = 4.85$. Parameter values: $\pi_h = 0.5$, $Y_l = 0.5$, $\hat{Y} = 2.65$, $\underline{R} = 1$, $m = 0.1$. Wealth distribution: $w_S = 1000$, $w_B = 80$, $w_F = 190$.

panels show that both bond issuances and the level of investment increase as $\bar{\lambda}_b$ increases. The left two panels on the bottom row show that p and Φ both decrease in $\bar{\lambda}_b$, as per the proposition. The intuition is simple: tighter leverage caps push banks to sell fewer assets, leading to excess demand. The bottom-right panel plots the evolution of relative financier wealth ω conditional on a good aggregate shock. Relative to the unconstrained equilibrium depicted in blue, two effects jointly shape the degree of risk transferred to financiers. First, tighter capital requirements lead banks to reduce asset-sale promises and *decreases* the amount of risk transferred. Second, an increase in the fraction of shirking banks leads to an *increase* in risk transfer because shirking bank sell more assets. When leverage caps are not too tight, the first effect dominates and ω grows more slowly in the constrained equilibrium. When capital requirements are tight, the second effect dominates and relative financier wealth grows *faster* in the constrained equilibrium.

The example thus reveals a static and a dynamic channel through which leverage caps adversely impact the flow of credit: statically, lending standards deteriorate as supply shortfalls push up prices; dynamically, increased risk transfer leads financiers to grow faster than banks, inducing further falls in investment quality. It also stands to reason that the first channel is particularly strong when capital requirements are counter-cyclical – asset demand is particularly high at the peak of a boom – while the second channel is particularly strong when regulations are risk-weighted – if selling assets for cash relaxes leverage caps, then selling risky assets is particularly attractive. For this reason, the model's predictions are also consistent with the *regulatory arbitrage* view articulated in, e.g., Acharya, Schnabl, and Suarez (2013), that banks used asset markets to bypass capital requirements.

Equity injections and macro-prudential market interventions. I now briefly discuss the model's key implications for equity injections and macro-prudential regulation. Proposition 3 showed that aggregate lending volumes may increase more sharply in w_F than w_B when intermediaries are highly constrained and financiers are impaired. If this is the case after a crisis, then providing equity to financiers is more cost-effective than funding banks.

On the downside, increases in w_F may trigger the incentive channel in the future, leading to excessive risk-taking and inefficient credit standards (Proposition 5). This suggests a role for macro-prudential policy in regulating the dynamics of secondary market booms more generally. The inefficiency stems from a pecuniary exernality that leads to excess asset market demand. A simple policy is therefore to place a cap $\hat{w}_F$ on financiers' asset purchases, and choose this cap such that no bank shirks. Such a policy naturally eliminates within-period inefficiencies. Yet it may also have dynamic benefits. Indeed, it is easy to see that *aggregate* wealth $w_F + w_B$ is larger in any state of the world under this policy. The reason is that the policy eliminates shirking on the equilibrium path, and thus improves the allocation of capital. While aggregate wealth is generally not a sufficient statistic for welfare or total investment, Figure 11 presents an example in which the policy leads to strictly higher investment and output in every period.

Such a policy may well be impractical. Yet the thought experiment does suggest that constraints on the asset side of financier balance sheets are a useful macro-prudential policy tool. Three aspects of such a policy are of note. First, financier wealth is harmful only when it is large. Because financiers grow during expansions, the policy is pro-cyclical. Second, the policy is independent of financier capital structure. That is, there is a motive for regulation independent of whether financiers are levered or not. Third, the aggregate size of the financier sector, rather than the systemic relevance of individual financial institutions, is the relevant concern. The last two points contrast with regulatory discussions at the Financial Stability Board, which focused on the designating individual asset managers as systemically important because they were fearful of

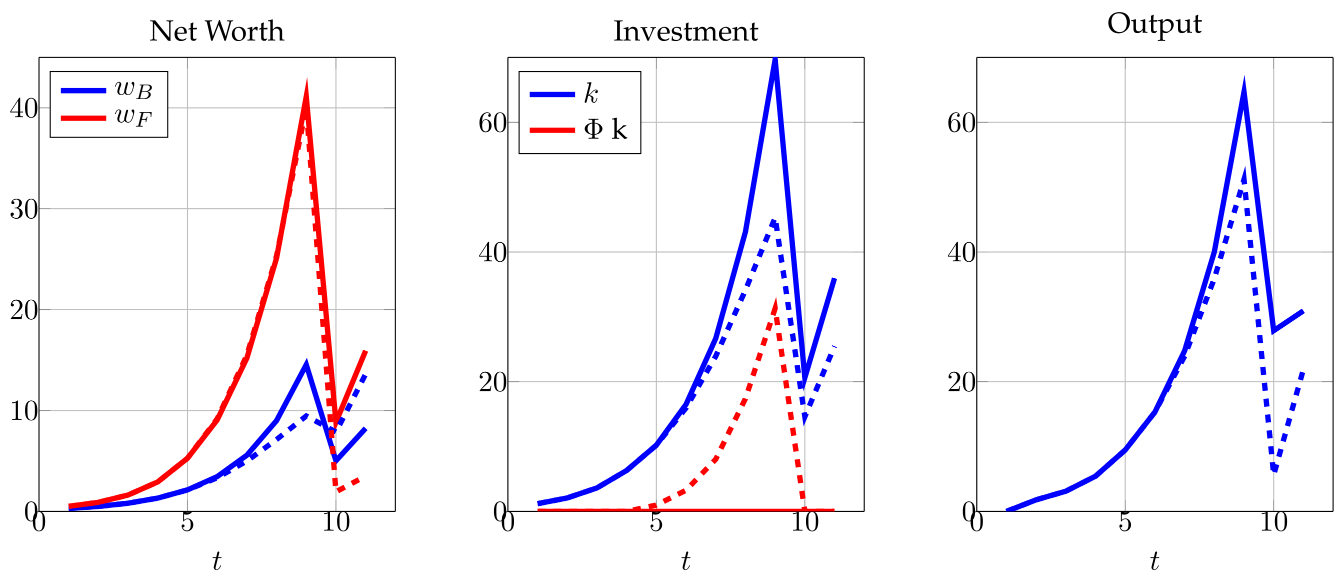

Figure 11: Equilibrium outcomes when financiers can invest no more than $\hat{w}_F$ in risky assets, and $\hat{w}_F$ chosen such that $\Phi = 0$ in whenever reductions in financier wealth are Pareto-improving as in Proposition 5. Solid lines depict the equilibrium with the policy and dashed lines the equilibrium in the absence of the policy.

sudden withdrawals from such institutions. In this sense, my results suggest a novel motive for financial regulation.

4 Conclusion

This paper offers a theory of the macroeconomic effects of secondary market trading. Secondary market trading impacts the flow of credit through the distribution of aggregate risk exposure in the cross-section of financial intermediaries. Some risk transfer away from constrained lenders relaxes a borrowing constraint and allows for the expansion of credit volumes. Excessive risk transfer destroys monitoring incentives and leads to lax credit standards and excessive aggregate risk exposure. The level of risk transfer is determined by the distribution of wealth in the financial system. I distinguish between "banks" – intermediaries that lend to firms and household directly, such as commercial banks or mortgage originators – and "financiers" – those who do trade in assets originated by other intermediaries, such as hedge funds or dealer banks. There is excessive risk transfer when financiers are too well-capitalized relative to banks. Dynamically, the risk transfer that allows credit volumes to expand when financiers are not too large causes financier wealth to grow disproportionately after a sequence of good shocks. As a result, there are credit cycles with gradually declining investment efficiency and increasing financial fragility.

Secondary market booms are triggered by periods of low interest rates. The model therefore provides a novel link from expansionary monetary policy and "saving gluts" to future financial fragility. In this manner, it sheds new light on the origins of the U.S. credit boom that eventually ended in the 2008 financial crisis. I also show that bank leverage restrictions may be harmful, and that there is a strong motive for restricting purchases of asset-backed securities over the cycle.

There are two main avenues for future research. The first is to study the optimal design of policy in the context of secondary market trading. The second is to undertake a quantitative evaluation of the mechanisms proposed in this paper.

References

ACHARYA, V. V., P. SCHNABL, AND G. SUAREZ (2013): "Securitization without Risk Transfer," *Journal of Financial Economics*, 107(3), 515–536.

ADRIAN, T., AND H. S. SHIN (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19(3), 418–437.

ALTUNBAS, Y., L. GAMBACORTA, AND D. MARQUES-IBANEZ (forthcoming): "Does Monetary Policy Affect Bank Risk?," *International Journal of Central Banking*.

ATTAR, A., T. MARIOTTI, AND F. SALANIÉ (2011): "Nonexclusive Competition in the Market for Lemons," *Econometrica*, 79(6), 1869–1918.

BIANCHI, J. (2011): "Overborrowing and Systemic Externalities in the Business Cycle," *American Economic Review*, 101(7), 3400–3426.

BIANCHI, J., AND E. G. MENDOZA (2012): "Overborrowing, Financial Crises and 'Macro-prudential' Policy," *University of Pennsylvania Working Paper*.

BIGIO, S. (2014): "Financial Risk Capacity," *Working Paper*.

BOLTON, P., T. SANTOS, AND J. A. SCHEINKMAN (2016): "Savings Gluts and Financial Fragility," *Columbia University Working Paper*.

BRUNNERMEIER, M. K. (2009): "Deciphering the Liquidity and Credit Crunch 2007-2008," *Journal of Economic Perspectives*, 23(1), 77–100.

CABALLERO, R. J., AND E. FARHI (2014): "The Safety Trap," *NBER Working Paper 19927*.

CABALLERO, R. J., E. FARHI, AND P.-O. GOURINCHAS (2008): "An Equilibrium Model of "Global Imbalances" and Low Interest Rates," *American Economic Review*, 98(1), 358–93.

CABALLERO, R. J., AND A. KRISHNAMURTHY (2009): "Global Imbalances and Financial Fragility," *American Economic Review, Papers and Proceedings*, 99(2), 584–88.

CHARI, V., A. SHOURIDEH, AND A. ZETLIN-JONES (2014): "Reputation and Persistence of Adverse Selection in Secondary Loan Markets," *American Economic Review*, 104(12), 4027–70.

COVAL, J. D., J. W. JUREK, AND E. STAFFORD (2009): "Economic Catastrophe Bonds," *American Economic Review*, 99(3), 628–66.

DI TELLA, S. (2014): "Uncertainty Shocks and Balance Sheet Recessions," *Working Paper*.

FOSTEL, A., AND J. GEANAKOPLOS (2008): "Leverage Cycles and the Anxious Economy," *American Economic Review*, 98(4), 1211–1244.

——— (2015): "Leverage and Default in Binomial Economies: A Complete Characterization," *Econometrica*, 83(6), 2191–2229.

GARLEANU, N., AND L. PEDERSEN (2011): "Margin-based Asset Pricing and the Law of One Price," *Review of Financial Studies*, 24(6), 1980–2022.

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

GENNAIOLI, N., A. SHLEIFER, AND R. W. VISHNY (2013): "A Model of Shadow Banking," *The Journal of Finance*, 68(4), 1331–1363.

GORTON, G., AND A. METRICK (2012): "Securitized Banking and the Run on Repo," *Journal of Financial Economics*, 104(3), 425–451.

GORTON, G., AND G. ORDOÑEZ (2014): "Collateral Crises," *American Economic Review*, 104(2), 343–78.

GREENLAW, D., J. HATZIUS, A. K. KASHYAP, AND H. S. S. SHIN (2008): "Leveraged Losses: Lessons from the Mortgage Market Meltdown," in *Proceedings of the U.S. Monetary Policy Forum 2008*.

GRIFFIN, J., R. LOWERY, AND A. SARETTO (2014): "Complex Securities and Undewriter Reputation: Do Reputable Underwriters," *Review of Financial Studies*, 27(10), 2872–2925.

GROMB, D., AND D. VAYANOS (2002): "Equilibrium and welfare in markets with financially constrained arbitrageurs," *Journal of Financial Economics*, 66, 361–407.

IVASHINA, V., AND Z. SUN (2011): "Institutional Demand Pressure and the Cost of Corporate Loans," *Journal of Financial Economics*, 99(3), 500–522.

JENSEN, M. C., AND W. H. MECKLING (1976): "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3(4), 305–360.

JUSTINIANO, A., G. E. PRIMICERI, AND A. TAMBALOTTI (2015): "Credit Supply and the Housing Boom," *Federal Reserve Bank of New York Staff Report No. 709*.

KAMINSKY, G. L. (2008): "Crises and Sudden Stops: Evidence from International Bond and Syndicated-Loan Markets," *NBER Working Paper 14249*.

KEYS, B. J., T. MUKHERJEE, A. SERU, AND V. VIG (2010): "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *The Quarterly Journal of Economics*, 125(1), 307–362.

KRISHNAMURTHY, A., S. NAGEL, AND D. ORLOV (2014): "Sizing up Repo," *The Journal of Finance*, 69(6), 2381–2417.

KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2012): "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 120(2), 233–267.

KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2015): "The Impact of Treasury Supply on Financial Sector Lending and Stability," *Journal of Financial Economics*, forthcoming.

KURLAT, P. (2013): "Lemons Markets and the Transmission of Aggregate Shocks," *American Economic Review*, 103(4), 1463–89.

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

MENDOZA, E. G. (2010): "Sudden Stops, Financial Crises, and Leverage," *American Economic Review*, 100, 1941–1966.

MENDOZA, E. G., AND M. E. TERRONES (2012): "An Anatomy of Credit Booms and Their Demise," *NBER Working Paper 18379*.

MIAN, A., AND A. SUFI (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *The Quarterly Journal of Economics*, 124(4), 1449–1496.

PARLOUR, C. A., AND G. PLANTIN (2008): "Loan Sales and Relationship Banking," *The Journal of Finance*, 63(3), 1291–1314.

PISKORSKI, T., A. SERU, AND J. WITKIN (2015): "Asset Quality Misrepresentation by Financial Intermediaries: Evidence from RMBS Market," *Journal of Finance*, Forthcoming.

REINHART, C., AND K. ROGOFF (2009): *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton.

SCHULARICK, M., AND A. M. TAYLOR (2012): "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," *American Economic Review*, 102(2), 1029–61.

SHIN, H. S. (2009): "Securitisation and Financial Stability," *The Economic Journal*, 119, 309–332.

VANASCO, V. (2014): "Information Acquisition vs. Liquidity in Financial Markets," *Working Paper*.

WHITE, E. N. (2009): "Lessons from the Great American Real Estate Boom and Bust of the 1920s," *NBER Working Paper 15573*.

A Proofs

Proof of Proposition 3

Let asset markets be tight. Assume first that financiers are fully levered ($\mu = 1$). The optimal bank portfolio satisfies $b_B = pa_B$, while the asset market clearing condition is $a_B = a_F$. Since the financier portfolio satisfies $b_F = Y_l a_F$, the bond market clearing condition is $pa_F + Y_l a_F = \frac{w_S}{q}$ Rearranging gives the bond price as $q(p) = \min\left(\frac{pw_S}{(p+Y_l)w_F + Y_l w_S}, \frac{1}{\underline{R}}\right)$. The secondary market clearing condition in turn gives the asset price as $p = \frac{w_F + w_B \tilde{m} Y_l q}{\tilde{m} w_B + \tilde{m} q w_F}$. Assume first that $q = \frac{1}{\underline{R}}$.

Then $p^* = \frac{w_F + w_B \tilde{m} Y_l \frac{1}{\underline{R}}}{\tilde{m} w_B + \tilde{m} \frac{1}{\underline{R}} w_F}$. Differentiating yields that p^* is increasing in w_F if and only if $w_B > \tilde{m} w_B \frac{Y_l}{\underline{R}}$, which always holds because $\tilde{m} \in (0,1)$ and $Y_l < \underline{R}$. Moreover, $k = \frac{w_B}{1-\tilde{m}qp}$ is increasing because p is increasing and q is a constant. Since all banks monitor, expected output also increases. Now assume that $q < \frac{1}{\underline{R}}$. Solving the system of two unknowns generated by the market clearing conditions gives the asset price as

$$p^* = \frac{w_F - \tilde{m} Y_l w_B + \sqrt{(w_F - \tilde{m} Y_l w_B)^2 + 4\tilde{m}(w_B + w_S) Y_l (w_F + w_S)}}{2\tilde{m}(w_B + w_S)}$$

which is clearly increasing in w_F. Next, we need to show that k is strictly increasing in w_F. Given the optimal bank portfolio, $k = \frac{1}{\tilde{m}} a_B$. By market clearing, $k = \frac{1}{\tilde{m}} a_F$. Since a_F is strictly increasing in w_F, the result follows. Moreover, expected output is increasing because all banks monitor. Next, assume that financiers do not borrow ($\mu = 0$). Then the market clearing conditions yield $q^* = \min\left(\frac{w_S}{p\tilde{m}(w_B + w_S)}, \frac{1}{\underline{R}}\right)$ and $p^* = \frac{w_F}{\tilde{m}(w_B + q w_F)}$. If $q^* = \frac{1}{\underline{R}}$, then p is clearly increasing in w_F. If $q < \frac{1}{\underline{R}}$, then $p^* = \frac{w_F}{\tilde{m}(w_S + w_B)}$ which is again increasing in w_F. Next, note that $q^* p^* = \frac{w_S}{\tilde{m}(w_B + w_S)}$. Hence $k = \frac{w_B}{1 - qp\tilde{m}}$ is non-decreasing in w_F. Next, assume that financiers are indifferent between borrowing and lending ($\mu \in (0,1)$). Then by definition, $p^* = \hat{Y} q^*$, and $b_F = \mu Y_l a_F$. The secondary market clearing condition is $\frac{\tilde{m} w_B}{1 - p^2 \frac{\tilde{m}}{\hat{Y}}} = \frac{w_F}{p(1 - \frac{\mu Y_l}{\hat{Y}})}$. Suppose for a contradiction that p is decreasing in w_F. Then $a_B = \frac{\tilde{m} w_B}{1 - p^2 \frac{\tilde{m}}{\hat{Y}}}$ is also decreasing in w_F. To maintain market clearing, a_F must be decreasing in w_F, and hence μ, $b_B = p a_B$ and $b_F = \mu Y_l a_F$ must also be decreasing. But if b_B and b_F are decreasing in w_F, then q must be *increasing* in w_F. This is a contradiction with the fact that q must decreasing because $p^* = \hat{Y} q^*$ and p was presumed to be decreasing. It then follows that $k = \frac{w_B}{1 - p^2 \frac{\tilde{m}}{\hat{Y}}}$ is increasing in w_F. Because all banks monitor, expected output is increasing in w_F also.

Now turn to the second part of the proposition and assume that asset markets are slack and intermediaries are highly constrained. Then $q^* = \frac{1}{\underline{R}}$, and $p^* = \underline{p}(q^*)$ are both constants. Hence a_F is increasing in w_F and, as a result, so is k.

Now turn to the third part of the proposition. Inspecting the optimal intermediary portfolios reveals that k is proportional to $w_B + \frac{q^*[\underline{p}(q^*) - Y_l] w_F}{\underline{p}(q^*) - q^* Y_l}$. Hence $\frac{\partial k}{\partial w_B} = \frac{\partial k}{\partial w_F}$ if $q = 1$ and $\frac{\partial k}{\partial w_B} < \frac{\partial k}{\partial w_F}$ if $q > 1$. Given that intermediaries are highly constrained, this is the case when $\underline{R} < 1$. □

Proof of Proposition 4

First-order condition (13) shows that a bank weakly prefers to increase its promise at $\underline{a}$ given $p(\underline{a}) = \bar{p}(\bar{k}(\underline{a}), \underline{a})$ if $u'_B(q, \bar{p}, \underline{a}) \geq 0 \Leftrightarrow -\bar{p}'(\underline{a}) \cdot \underline{a} \leq \bar{p}(\bar{k}(\underline{a}), \underline{a}) - \underline{p}(q)$, where $\bar{p}'(\underline{a}) = \frac{\partial \bar{p}(\bar{k}(\underline{a}), \underline{a})}{\partial \underline{a}}$. Computing this derivative in closed-form yields $p'(\underline{a}) = \frac{-m\lambda_B(q) w_B}{(\bar{k}(\underline{a}) - \underline{a})^2 - qm\lambda_B(q)\underline{a}^2}$. Evaluating this expression at $\underline{a}^* = \tilde{m}k^*$ gives $-p'(\underline{a}^*)\underline{a}^* = \frac{m\lambda_B(q)(\pi_l Y_l - m)}{\bar{\lambda}_B(q, \underline{a}^*)[(1-\tilde{m}) - \tilde{m}q\lambda_B(q)(\pi_l Y_l - m)]}$, where $\bar{\lambda}_B(q, \underline{a}^*) = \left[1 - q\tilde{m}\bar{p}(\bar{k}(\underline{a}), \underline{a})\right]^{-1}$ is the bank's effective leverage at $\underline{a}^*$. The numerator is strictly positive. The stated condition ensures that the denominator is strictly negative. The limit of the LHS as $m \to 0$ is ∞ because $\lim_{m \to 0} \tilde{m} = 1$, while the limit of the RHS is bounded. □

Proof of Proposition 5

Assume first that the asset market is tight. Since $\underline{a}^* = \tilde{m}k^*$, it directly follows that $p^* = \bar{p}(k^*, \underline{a}^*) = \hat{y}$ is a constant. If intermediaries are highly constrained, then q is fixed at $\frac{1}{R}$, and q^* is also constant. As a result, k is a constant. To clear secondary markets at fixed prices, Φ must be increasing in w_F. Given that k is constant, expected output must be declining. If intermedaries are merely constrained, then market clearing conditions are:

$$\frac{\hat{y}\tilde{m}W_B}{1-q\hat{y}\tilde{m}} + \frac{\mu(1-\phi)Y_l W_F}{\hat{y}-(1-\phi)q\mu Y_l} = \frac{W_d}{q} \qquad \text{and} \qquad \left(\frac{\Phi+(1-\Phi)\tilde{m}}{1-q\hat{y}\tilde{m}}\right)W_B = \frac{W_F}{\hat{y}-(1-\phi)q\mu Y_l}.$$

Suppose first that financiers are indifferent toward leverage ($\mu \in (0,1)$). Then $q^* = \frac{p^*}{\hat{Y}}$ is a constant. It follows immediately that Φ is increasing in w_F, while k is independent of w_F. Hence expected output must decline. Next, suppose financiers do not borrow ($\mu = 0$). In this case, the bond market clearing condition is independent of w_F, and so q is a constant. Hence k is a constant, and expected output must decline. Finally, assume that financiers are fully levered ($\mu = 1$). Imposing bond market clearing reveals that q must satisfy $q_{PM}(\Phi, w_F) = \frac{w_S}{\tilde{m}[\hat{y}(w_B+w_S)+(1-\Phi)Y_l w_F]}$, which is strictly increasing in Φ and strictly decreasing in w_F. The bond price that clears the secondary market is $q_{SM}(\Phi, w_F) = \frac{w_F - w_B\hat{y}(\Phi+(1-\Phi)\tilde{m})}{\tilde{m}[(w_B+w_F)Y_l + w_B Y_l \Phi]}$, which is strictly increasing in w_F but strictly decreasing in Φ. It is then straightforward to show that Φ must be strictly increasing in w_F. Suppose for a contradiction that is decreasing. By bond market clearing, an increase in w_F and a decrease in Φ leads to fall in q. But by secondary market clearing, an increase in w_F and a decrease in Φ leads to an *increase* in q. Hence both markets do not clear simultaneously.

Now assume that the asset market is slack. I will first show that whenever the financial system is highly constrained, or financiers weakly prefer to not borrow, then q^*, p^*, k and $\underline{a}$ are all invariant to w_F. Suppose first that intermediareis are highly constrained, so that $q^* = \frac{1}{R}$. Then $p^* = \underline{p}(q^*)$ is a constant, and thus invariant to w_F. Since $k = \frac{w_B + q^*(p^* - Y_l)\underline{a}}{1 - Y_l q^* \tilde{m}}$ in any excessive trading equilibrium, k^* and $p^* = \bar{p} = \hat{Y} - m\frac{k}{k-\underline{a}}$ in any excessive trading equilibrium, k and $\underline{a}$ are also invariant to w_F. Now suppose that financiers do not borrow. The market-clearing condition in the bond market is then given by $b_B = \frac{Y_l\tilde{m}w_B + (p-Y_l)\underline{a}}{1-Y_l q^*\tilde{m}} = \frac{w_S}{q}$. Hence $\underline{a}$ is fixed conditional on q. Since $p^* = \bar{p} = \hat{Y} - m\frac{k}{k-\underline{a}}$ and k is fixed once q is determined, the condition $p = \underline{p}(q)$ suffices to pin down q, p, k and $\underline{a}$ independently of w_F. But given that all quantities are pinned down independently of w_F, it must be the case that Φ increases in w_F to ensure secondary market clearing. Moreover, given fixed k, expected output must decline. Now assume that financiers are fully levered and that the financial system is not highly constrained. Then the market clearing conditions are:

$$\frac{Y_l\tilde{m}w_B + (\underline{p} - Y_l)\underline{a}}{1 - qY_l\tilde{m}} + \frac{Y_l w_F}{\underline{p} - (1-\phi)qY_l} = \frac{w_S}{q}$$

$$\Phi\left(\frac{w_B + q(\underline{p} - Y_l)\underline{a}}{1 - qY_l\tilde{m}}\right) + (1-\Phi)\underline{a} = \frac{w_F}{\underline{p} - (1-\phi)qY_l},$$

The asset price must satisfy $p^* = \underline{p}(q)$, and is thus fixed given q. The high-type bank's investment is given by:

$$k = \frac{w_B}{1 - qY_l\tilde{m}} + \left(\frac{\hat{Y} - Y_l}{\hat{Y} - Y_l\tilde{m}}\right)\underline{a}$$

Hence, k is a function of $\underline{a}$ and q only. Given that we are in a excessive trading equilibrium, $p = \bar{p}(k, \underline{a})$, and so p and q are fixed for a given $\underline{a}$. It is easy to verify that $\bar{p}$ is strictly decreasing in $\underline{a}$. We can then show that Φ must be strictly increasing in w_F. Suppose for a contradiction that Φ is weakly decreasing. For the bond market to clear, either q and/or $\underline{a}$ must decrease. Since $p = \underline{p}$, if q falls, then p must increase. Since $p = \bar{p}$ and $\bar{p}$ is strictly decreasing in $\underline{a}$, it follows that q and $\underline{a}$ must both decrease. Since b_B and q are decreasing, it follows that k must decrease. Since k, $\underline{a}$, and Φ are all decreasing, total secondary market supply must decrease. Yet a_F is weakly increasing. Hence secondary markets cannot clear, yielding a contradiction.

Lastly, we want to show that there are pareto-improving financier wealth reductions. Suppose first that asset markets are tight. Then $p^* = \hat{y}$, which is a constant. Furthermore, q^* must also be invariant to reductions in w_F, either because financiers do not borrow (so that w_F does not impact the bond market clearing condition, given that p^* is a constant), or because the financial system is highly constrained so that $q^* = \frac{1}{\underline{R}}$ and further reductions in w_F increase excess demand in the bond market. Hence k^* is invariant to w_F. Since q^* is constant, so is depositor utility. By construction, bank utility is $(1 - \tilde{m})k^*$, which is constant. If financiers do not borrow, financier utility is $(\Phi\hat{y} + (1 - \Phi)\tilde{m}\hat{Y})k^*$, and is strictly decreasing in w_F. If financiers do borrow, financier utility is $(\Phi\hat{y} + (1 - \Phi)\tilde{m}(\hat{Y} - Y_l)k^*$, which is again strictly decreasing in w_F.

Now suppose that asset markets are tight. I will first show that whenever the financial system is highly constrained, or financiers do not borrow, then q^*, p^*, k and $\underline{a}$ are all invariant to w_F. Suppose first that the financial system is highly constrained, so that $q^* = \frac{1}{\underline{R}}$. Then $p^* = \underline{p}(q^*)$ is a constant, and thus invariant to w_F. Since $k = \frac{w_B + q^*(p^* - Y_l)\underline{a}}{1 - Y_l q^* \tilde{m}}$ and $p^* = \bar{p} = \hat{Y} - m\frac{k}{k - \underline{a}}$ in any excessive trading equilibrium, k and $\underline{a}$ are also invariant to w_F. Now suppose that financiers do not borrow. The market-clearing condition in the bond market is then given by $b_B = \frac{Y_l\tilde{m}w_B + (p - Y_l)\underline{a}}{1 - Y_l q^* \tilde{m}} = \frac{w_S}{q}$. Hence $\underline{a}$ is fixed conditional on q. Since $p^* = \bar{p} = \hat{Y} - m\frac{k}{k - \underline{a}}$ and k is fixed once q is determined, the condition $p = \underline{p}(q)$ suffices to pin down q, p, k and $\underline{a}$ independently of w_F. We can then show that reductions in w_F are Pareto-improving. First, note that the utility of depositors is given by $v_d = \frac{w_B}{q^*}$ if $q^* > \frac{1}{\underline{R}}$ and $v_d = \underline{R}w_S$ otherwise. Since banks are indifferent towards selling assets on when asset markets are slack, their utility is unchanged by the presence of secondary markets: $v_b^H = \frac{\hat{Y} - \tilde{m}Y_l}{1 - q^* Y_l \tilde{m}}$. By construction, the utility of banks who shirk satisfies $v_b^L = v_b^H = v_b$. Since q^* is invariant to w_F, so are v_b and v_d. Next, turn to the *equilibrium* utility of financiers. When financiers borrow, it is $v_f = (\phi\hat{Y}' + (1 - \phi)(\hat{Y} - Y_l))a_F^*$. When they do not borrow, it is $v_f = (\phi\hat{Y}' + (1 - \phi)\hat{Y})a_F^*$. Market clearing requires that $\Phi a_B^{H*} + (1 - \Phi)k^{H*} = a_F = \frac{w_F}{p^* - (1 - \phi)Y_l q}$, where we have established that q^*, p^*, k^{H*} and a_B^{H*} are all constant. By the definition of ϕ, the utility of financiers then is $v_f = \Phi\hat{y}k^{H*} + (1 - \Phi)\hat{Y}a_B^{H*}$ when they do not borrow, and $v_f = \Phi\hat{y}k^{H*} + (1 - \Phi)(\hat{Y} - Y_l)a_B^{H*}$ when they do. Since k^{H*} and a_B^{H*} are constants and Φ is strictly increasing in w_F, v_f is strictly decreasing in w_F in equilibrium. □

Proof of Proposition 6

The law of motion for relative wealth follows directly from the optimal portfolios of banks and financiers and the asset market clearing condition. The law of motion states that relative wealth is a constant ratio pinned down by parameters only, with the ratio being larger when financiers do not issue bonds. Financiers do not issue bonds when q is too low. Because q must decrease after a good shock since the wealth of intermediaries increases after a good shock), ω is either constant or increasing when asset markets are tight. The asset market clearing condition states that $p = \frac{1}{\tilde{m}}\left(\frac{\tilde{m}qY_l+\omega}{1+q\omega}\right)$, which is strictly increasing in ω. To find a sufficient condition for p to increase after a good shock, we must therefore find conditions under which p increases even when ω is constant. Since q decreases after a good shock, we therefore require $\frac{\partial p}{\partial q} < 0$. This is the case when $\omega > \sqrt{\tilde{m}Y_l}$. Since $\omega \geq \frac{\tilde{m}}{1-\tilde{m}}\frac{Y_h-Y_l}{Y_h}$ when asset markets are tight, the condition follows.

Proof of Proposition 7

The law of motion for relative wealth can computed directly from the optimal portfolios of banks and financiers, taking into account that $p = \underline{p}(q)$. The second part of the proposition follows from noting that $\text{ROE}_F(z) - \text{ROE}_B(z)$ is strictly increasing in q, and that $\mathbb{E}_z\text{ROE}_F(z) - \text{ROE}_B(z) = 0$ if $q = 1$.

Proof of Proposition 8

By construction. Consider a low-liquidity equilibrium in which the financial system is highly constrained. Such an equilibrium always exists for w_S large enough and w_F small enough. In such an equilibrium, $q^* = \frac{1}{\underline{R}}$ and $p^* = \underline{p}(q^*)$. Let $\lambda_F = \frac{1}{p^*-q^*Y_l}$ and $\lambda_B = \frac{1}{1-q^*Y_l\tilde{m}}$ denotes the equilibrium leverage of financiers and banks in a efficient monitoring equilibrium, respectively. From the optimal bank portfolio and the definition of $\underline{p}(q)$, it follows that

$$k^* = \lambda_B w_B + \chi a_F^*$$

where $\chi = \left(\frac{\hat{Y}-Y_l}{\hat{Y}-Y_l\tilde{m}}\right) \in (0,1)$ and $a_F^* = \lambda_F w_F$. Hence, the upper bound on the asset price is

$$\bar{p} = \hat{Y} - \tilde{m}\left(\frac{\lambda_B w_B + \chi\lambda_F w_F}{\lambda_B w_B - (1-\chi)\lambda_F w_F}\right) = \hat{Y} - \tilde{m}\left(\frac{\lambda_B + \chi\lambda_F\omega}{\lambda_B - (1-\chi)\lambda_F\omega}\right)$$

Note that $\bar{p} = \hat{Y} - m$ if $\omega = 0$ and that $\bar{p}$ is strictly decreasing in ω. It follows that as long as $p^* < \hat{Y} - m$ there exists, for small enough w_F, an efficient monitoring equilibrium with slack asset markets in which the financial system is highly constrained. This parametric condition is equivalent to

$$\underline{R} < \frac{1}{\chi}(\hat{Y} - m) - \frac{(1-\tilde{m})Y_l\hat{Y}}{\hat{Y} - Y_l} \tag{16}$$

Next, note that $\bar{p} \geq \hat{y}$ because $a_B \leq \tilde{m}k$. For a excessive trading equilibrium to exist for sufficiently large w_F, we therefore require that $p^* \geq \hat{y}$. This parametric condition is equivalent to

$$\underline{R} > \frac{1}{\chi}\hat{y}' - \frac{(1-\tilde{m})Y_l\hat{Y}}{\hat{Y} - Y_l} \tag{17}$$

It is easy to see that there exist parameters such conditions (16) and (17) are jointly satisfied. For example, set $\hat{y} = \underline{R} - \epsilon$ for ϵ and m sufficiently small. Hence there exist parameters such that $p \in [\hat{y}, \hat{Y} - m)$. Assume a set of such parameters from now on, and choose initial financier wealth w_F^0 such that the economy is initially in a efficient monitoring equilibrium. We now want to show that the economy may transition into a excessive trading equilibrium after a sufficiently long sequence of large shocks. Note first that because intermediary wealth is bounded after any finite sequence of good aggregate shocks, there always exists a level of depositor wealth such that the financial system is highly constrained after any such sequence. Hence, we can construct a destabilizing secondary market boom under the presumption that the financial system is highly constrained throughout. As a result, prices are fixed throughout and $q^* \geq 1$ because $\underline{R} \leq \bar{y}_S = 1$. By Proposition 7, relative financier wealth ω thus grows after a good shock for any ω. By the parametric condition (17), a sufficiently long sequence of good aggregate shocks therefore triggers a excessive trading equilibrium. We then only need to show that there exist parameters such that *expected* return on equity is higher for financiers than for banks throughout. Recall that in efficient monitoring equilibrium with slack asset markets, $\hat{\text{ROE}}_F > \hat{\text{ROE}}_B$ for $q > 1$. Hence $\hat{\text{ROE}}_F > \hat{\text{ROE}}_B$ in a efficient monitoring equilibrium when the financial system is highly constrained if $\underline{R} < 1$. Next, turn to a excessive trading equilibrium. By construction, the return on equity of banks is independent of the fraction of shirking banks Φ, while the return on equity of financiers is strictly decreasing in Φ. As long as the return-on-equity for financiers is strictly higher in a efficient monitoring equilibrium, there exists a Φ^* such that the return on equity is also strictly higher in a excessive trading equilibrium in which Φ^* banks shirk. This is the case when $\underline{R} < 1$. □

Proof of Proposition 9

The first part of the statement (k increasing) follows directly from the optimal bank portfolio. The second part (ω increasing after a good shock) follows from Propostion 7

Proof of Proposition 10

Begin by discussing the bank's decision problem given leverage caps. In the absence of capital requirements, the bank's optimal portfolio is such that they sell off exactly the amount of risky claims that maximizes their borrowing capacity: $\underline{a} = \tilde{m}K = \frac{b_B}{p}$. When capital requirements bind, however, banks are no longer permitted to exhaust their entire borrowing capacity. Because risky claims trade below par, banks therefore withdraw assets from secondary markets until the capital requirement just binds. I summarize the degree to which banks exhaust their borrowing capacity by $\mu_B \in [0, 1]$. The secondary market supply of banks can then be written as $a_B = \mu_B \left(\frac{b_B}{p}\right)$ .. Since the borrowing and budget constraints continue to bind, the optimal portfolio for given prices is

$$k = \left[\frac{(1-\mu_B)p + \mu_B Y_l}{(1-\mu_B)p + \mu_B Y_l - pqY_l\tilde{m}}\right] w_B \qquad \text{and} \qquad b_b = \left[\frac{pY_l\tilde{m}}{(1-\mu_B)p + \mu_B Y_l - pqY_l\tilde{m}}\right] w_B.$$

Accordingly, bank leverage for a given μ_B is $\lambda_B(\mu) = \frac{(1-\mu_B)p+\mu_B Y_l}{(1-\mu_B)p+\mu_B Y_l - pqY_l\tilde{m}}$. Setting $\lambda_B(\mu) = \bar{\lambda}_b$ reveals that the degree to which the bank exhausts his borrowing capacity under capital requirements is $\mu_B^*(\bar{\lambda}_b) = \left(\frac{p}{p-Y_l}\right)\left[1 - \left(\frac{\bar{\lambda}_b}{\bar{\lambda}_b - 1}\right) qY_l\tilde{m}\right] \in (0,1)$. All else equal, the supply of risky claims is thus increasing in the capital requirement $\bar{\lambda}_b$, but *decreasing* in the asset price p.

We can then show that the equilibrium with leverage caps must include shirking. Suppose for a contradiction that all banks monitor. Let p^* denote the equilibrium asset price. Since all banks monitor, $p^* < \hat{Y}$. Since the financial system is highly constrained, $q^* = 1$ with and without leverage constraints.. As a result, financiers are fully levered, and the demand for risky claims is $a_F = \frac{w_F}{p^* - Y_l}$. From the optimal bank portfolios, the supply of risky assets is $a_B = \frac{w_B\left(\bar{\lambda}_B(1-Y_l\tilde{m})\right)}{p^* - Y_l}$. Since secondary markets clear in the absence of capital requirements, $a_F > a_B$ for any p^* if $\bar{\lambda}_B < \lambda_B^*$. Hence, there is *excess demand* for risky claims. To restore market clearing, financiers must be indifferent between risky claims the safe technology. But this requires $p = \hat{Y}$.

For completeness, I now fully characterize the equilibrium with capital requirements. The equilibrium asset price p^* satisfies $p^* = \hat{Y} - m \cdot \left[\frac{\bar{\lambda}_B(p^* - Y_l)}{\bar{\lambda}_B(p^* - m') - (\bar{\lambda}_B - 1)}\right]$ and is strictly decreasing in $\bar{\lambda}_B$. The equilibrium bank portfolio is $k = \bar{\lambda}_B w_B$, $\underline{a} = \frac{w_B(\bar{\lambda}_B(1-Y_l\tilde{m}))}{p^* - Y_l}$ and $b_B = w_B(\bar{\lambda}_B - 1)$. The fraction of shirking banks is Φ determined by the secondary market clearing condition $\Phi k + (1-\Phi)\underline{a} = \frac{w_F}{p^* - (1-\phi)Y_l}$, where $\phi = \frac{\Phi k}{\Phi k + (1-\Phi)\underline{a}}$ and Φ is decreasing in $\bar{\lambda}_b$. □

B Dynamic Model with Endogenous Risk Aversion

In this section, I study a variant of the dynamic model in Section 2 in which old intermediaries have full bargaining power. This assumption implies that the old appropriate the entire *value* of their end-of-life stock of wealth. Generically, this value of wealth is state-contingent, with intermediaries valuing a dollar of equity more highly in states of the world where intermediation rents are large. Intermediation rents are large when intermediaries are not well-capitalized in the aggregate. The health of intermediary balance sheets will in turn depend on the realization of aggregate risk. Forward-looking behavior thus leads to endogenous risk preferences.

The main goal of this section is to show that the forces that drove secondary market booms in the baseline dynamic model are not overturned by considerations of endogenous risk aversion. To do so, I construct examples in which financiers grow even in the presence of endogenous risk aversion. For simplicity, I focus on the special case $T = 3$. The key simplification inherent in this assumption is that intermediaries face a finite horizon. This allows me to characterize the value of equity capital in the final period in closed form. Since intermediaries appropriate the entire value of their end-of-life wealth, I can then analyze the problem as if the initial generation of intermediaries lived for three periods rather than two, and intermediates capital in the latter two periods. I denote the final-period value of w units of equity capital to an intermediary of type τ when the wealth distribution is $\mathbf{w}$ by $v_\tau(w, \mathbf{w})$. Since all intermediaries are risk-neutral, the following proposition follows immediately:

Proposition (The Value of Equity Capital)**.**
The final-period value of w units of equity capital to an intermediary of type τ when the wealth distribution is $\mathbf{w}$ is linear in w:

$$v_\tau(w, \mathbf{w}) = \alpha_\tau(\mathbf{w})w$$

Proof. Follows directly from all policy functions in the static game being linear in wealth. □

Since there are only three periods, there are only two generations of intermediaries and one intergenerational equity market. The second (and final) generation of intermediaries chooses the same portfolios as in the static model. The key stage of analysis is thus the initial generation's portfolio choice, taking into account that they maximize the market value of equity capital. I suppress time subscripts for simplicity. Since financiers and banks have endogenous risk preferences, they may value a risky claim differentially even in the absence of borrowing constraints. Specifically, a risky claim is of little value to an intermediary that highly values wealth conditional on a negative aggregate shock. This gives rise to a trading motive separate from selling assets to relax borrowing constraints.

Disregarding borrowing constraints, the bank weakly prefers to sell the asset at price p if and only if

$$\frac{\alpha_B(l)}{\alpha_B(h)} \geq \frac{\pi(Y_h - p)}{\pi_l(p - Y_l)}$$

A financier strictly prefers to purchase the risky asset at price p rather than hold the safe asset if and only

$$\frac{\alpha_F(l)}{\alpha_F(h)} < \frac{\pi(Y_h - p)}{\pi_l(p - Y_l)}.$$

I say that a given intermediary is the **natural bearer of risk** when his valuation of a risky claim is the highest among all intermediaries.

Lemma (Natural Bearer of Risk)**.**
Intermediary τ is the natural bearer of risk if and only if

$$\tau = \arg\min_{\tau'} \quad \frac{\alpha_{\tau'}(l)}{\alpha_{\tau'}(h)}$$

Going forward, I will use $\sigma_\tau \equiv \frac{\alpha_\tau(l)}{\alpha_\tau(h)}$ to summarize the risk attitude of the type-τ intermediary. As long as $\sigma_F < \sigma_B$, there exists a p such that financiers are willing to purchase the risky asset and banks are willing to sell. When instead $\sigma_F = \sigma_B$, there are no endogenous differences in risk-preference and intermediaries trade assets as in the static model. To show that the results from the baseline model are robust, I now construct an example in which the economy with endogenous risk aversion admits secondary market booms as in Section 2.

Proposition.
If, after any shock, the competitive equilibrium in the final period is a efficient monitoring equilibrium with slack asset markets and a highly constrained financial system then $\sigma_F = \sigma_B = 1$.

Proof. In an equilibrium with slack asset markets in which the financial system is highly constrained we have $q^* = \frac{1}{\underline{R}}$ and $p^* \underline{p}(q^*)$. By Proposition 7, $\hat{\text{ROE}}_f = \frac{\hat{Y}-Y_l}{\underline{p}(q^*)-q^*Y_l}$ and $\hat{\text{ROE}}_f = \frac{\hat{Y}-Y_l\tilde{m}}{1-q^*Y_l\tilde{m}}$. Given that the financial system is highly constrained after any shock, the result follows. □

Proposition 8 provides an example of destabilizing secondary market booms when the financial system is highly constrained and secondary market liquidity is low. The above proposition implies that the evolution of the economy under endogenous risk aversion is identical to that example as long as the economy is in a efficient monitoring equilibrium. What remains to be shown is that the economy also transitions into a excessive trading equilibrium after a sequence of good shocks.

Proposition.
If the competitive equilibrium in the final period is an efficient monitoring equilibrium *with slack asset markets and a highly constrained financial system after a bad shock, and a* excessive trading equilibrium *with slack asset markets and a highly constrained financial system after a good shock, then*

$$\sigma_B = 1 \qquad \textit{and} \qquad \sigma_F = \frac{\frac{\phi\hat{y}+(1-\phi)(\hat{Y}-Y_l)}{p^*-(1-\phi)q^*Y_l}}{\frac{\hat{Y}-Y_l}{p^*-q^*Y_l}} \leq 1$$

where $q^* = \frac{1}{\underline{R}}$ *and* $p^* = \underline{p}(q^*)$.

Proof. For banks, the result follows from the fact that return on equity is independent of Φ by construction. For financiers, the result follows from a straightforward computation of expected utility in the excessive trading equilibrium. □

It follows that the economy with endogenous risk aversion must also transition into a excessive trading equilibrium. To see this, suppose first that the economy with endogenous risk aversion does *not* transition into a excessive trading equilibrium after a good shock, while the economy without endogenous risk aversion does. Then $\Phi = 0$ after a good shock. But then the above proposition implies that $\sigma_F = \sigma_B = 1$, and there is no endogenous risk aversion. As a result, the economy must transition into a excessive trading equilibrium, yielding a contradiction. Note that $\underline{p}$ is the same in the presence of endogenous risk aversion as in its absence because $\sigma_B = 1$ throughout. Moreover, financiers are willing to buy risky assets when Φ is sufficiently small tomorrow because they receive strictly positive rents from doing so when $\sigma_F = 1$.

More generally of course, endogenous risk aversion contributes to a slower build-up of risk and fragility. Intermediaries' endogenous preference to preserve capital for downturns makes them less willing to hold risk exposure. Accounting for the channel is thus important in a quantitative sense. In a qualitative sense, however, the previous proposition shows that the model admits the same dynamics as without endogenous risk aversion.

Imprint

Note
This paper is the joint runner-up of the ESRB's Ieke van den Burg research prize, 2016. This annual prize recognises outstanding research conducted by young scholars on a topic related to the ESRB's mission. The paper was selected as joint runner-up for the prize by a committee of the Advisory Scientific Committee comprising Prasanna Gai, Ross Levine, Marco Pagano, Richard Portes and David Thesmar. The committee considered the paper to be a sophisticated theoretical analysis of the financial cycle arising from the interaction between the banking and the shadow-banking system. The novel model of liquidity panics, based on informational frictions in the interbank market, yields new and non-trivial policy implications.

Acknowledgements
This paper is based on Chapter 1 of my dissertation at the University of Pennsylvania. I am grateful for many helpful conversations with my advisors Harold Cole, Itay Goldstein, Dirk Krueger, and Guillermo Ordo~nez. I thank Luigi Bocola, Murat Alp Celik, Selman Erol, Vincent Glode, Nir Jaimovich, Urban Jermann, Ricardo Lagos, Michael J. Lee, Iourii Manovskii, Enrique Mendoza, Stephen Morris, Christian Opp, Francisco Silva, and participants at the Penn Macro Club and the Penn Economics/Wharton Liquidity and Financial Crisis Reading Group for their feedback. I gratefully acknowledge financial support from the Robert Summers Dissertation Fellowship in Economics at the University of Pennsylvania. This paper has been prepared by the author under the Lamfalussy Fellowship Program sponsored by the ECB. Any views expressed are only those of the author and do not necessarily represent the views of the ECB or the Eurosystem.

Daniel Neuhann
McCombs School of Business, University of Texas at Austin; email: **daniel.neuhann@mccombs.utexas.edu**

Postal address	60640 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Website	**www.esrb.europa.eu**

ISSN	2467-0677 (online)
ISBN	978-92-95081-55-0 (online)
DOI	10.2849/347961 (online)
EU catalogue No	DT-AD-16-025-EN-N (online)