Macroeconomic Effects of Secondary Market Trading*

Daniel Neuhann†

January 17, 2016

Abstract

Starting around 1990, financial intermediaries in the United States increasingly began to sell, rather than hold to maturity, many of the loans that they provided to households and firms. This paper presents a theory in which the endogenous growth of such *secondary market trading* generates a macroeconomic credit cycle. Growing secondary markets initially boost credit volumes but gradually lead credit to flow to excessively risky investments. Aggregate risk exposure builds as asset quality falls. Ultimately, a negative shock leads to a simultaneous collapse of secondary markets and credit volumes – as in the financial crisis of 2008. Booms are triggered by periods of 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.

*JOB MARKET PAPER. Click here for the latest version. I am grateful for many helpful conversations with my advisors Harold Cole, Itay Goldstein, Dirk Krueger, and Guillermo Ordoñez. 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.

†University of Pennsylvania; neuhann@sas.upenn.edu

1 Introduction

Starting around 1990, financial intermediaries in the United States increasingly began to sell, rather than hold to maturity, many of the loans that they provided to households and firms. The rise of such *secondary market trading* of financial assets was accompanied by a credit boom that ended in the financial crisis of 2008. In the aftermath of the crisis, policymakers and academics alike have argued that growing secondary markets were a crucial driver of both the credit boom and eventual bust.[1] Yet, the underlying mechanisms are not fully understood. This paper offers a theory in which the *endogenous growth* of secondary markets generates a macroeconomic credit cycle. I use the theory to understand why secondary market credit booms arise, why they eventually lead to financial crises, and how policy affects their macroeconomic consequences.

In this theory, secondary markets allow financial intermediaries to sell off risk exposure to other intermediaries. This has two conflicting effects: first, a more efficient allocation of risk can increase the borrowing capacity of intermediaries and allow for the expansion of credit volumes. Second, asset sales reduce intermediaries' incentives to screen or monitor investment opportunities ex-ante, hampering the efficiency of investment. The adverse incentive effect arises only if secondary market volumes are high and intermediaries sell off a sufficiently large fraction of their investments. Credit cycles arise because the two effects are linked over time. The transfer of risk leads secondary market volumes to grow during macroeconomic expansions because the wealth of those intermediaries who buy risky assets grows when this risk pays off. Growing secondary market volumes in turn lead to deteriorating lending incentives. Financial fragility grows during upturns because capital increasingly flows to low-quality investments. Ultimately, a negative shock leads to a simultaneous collapse of secondary markets and credit volumes. Booms are triggered by low interest rates – due to, for example, expansionary monetary policy or saving gluts – because cheap funding increases the value of increased borrowing capacity to intermediaries.

I study a segmented-markets economy in which risk-neutral financial intermediaries make risky investments on behalf of risk-averse outside investors subject to moral hazard. Krishnamurthy and Vissing-Jorgensen (2012) provide evidence that investors pay a *safety premium* for risk-free financial assets, while Gorton and Pennacchi (1990) and Krishnamurthy and Vissing-Jorgensen (2015) argue that the production of safe assets is a key function of the financial sector. Intermediaries thus borrow by issuing risk-free debt.[2] As a result, their funding ability is constrained by their net worth and risk exposure, and secondary market sales serve to reduce risk exposure in order to increase borrowing. But, who buys risk exposure? When investments are

[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. 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 boom and bust.

[2] My results generalize to any setting in which financial intermediaries are constrained by their risk exposure. This may be the case even when all outside investors are, in principle, willing to hold risk exposure. In Hebert (2015), debt is the optimal security in settings that includes flexible moral hazard, i.e. an effort choice that affects average returns *and* volatility. In the Diamond (1984) model of delegated monitoring, the efficiency of financial intermediation improves when intermediaries can offload aggregate risk exposure.

subject to aggregate risk, there are no gains from trade among symmetric intermediaries – when one intermediary's risk decreases, another's increases. Yet Coval, Jurek, and Stafford (2009) show that the financial assets traded on secondary markets typically carry strong exposure to aggregate risk. I therefore study an economy with two types of intermediaries: *bankers*, who have the requisite skill to access investment opportunities in the real economy, and *financiers*, who cannot access these opportunities directly and instead purchase assets on secondary markets. Bankers represent commercial banks or mortgage originators who directly provide loans to households and firms. Financiers represent investors with an appetite for risk, such as hedge funds, broker dealers, and asset managers.

Financiers are willing to take on aggregate risk exposure precisely because they do not make investments directly and thus do not face the same funding constraints as bankers. Rather, financiers earn intermediation rents because their risk-taking behavior allows bankers to expand borrowing and lending. This is socially valuable: bankers are less likely to engage in moral hazard when they are less exposed to risk. Indeed, when total intermediary net worth is scarce, a financial system with both financiers and bankers allows for more borrowing and lending than one of equal size featuring only bankers. Secondary markets thus boost credit volumes through the transfer of risk away from bankers. Greenlaw, Hatzius, Kashyap, and Shin (2008) estimate that financial institutions who purchased mortgage-backed securities on secondary markets were more exposed to mortgage default risk than commercial banks during the 2008 financial crisis. Mian and Sufi (2009) and Ivashina and Sun (2011) provide evidence that growing secondary markets were associated with increased credit to households and firms.

The balance sheets of financial institutions are hard to monitor in real time. Moreover, bankers typically trade with many financiers at the same time, and they are more informed about the quality of the assets they produce than potential buyers. That is, secondary markets are *non-exclusive* and hampered by *asymmetric information*. When trade is non-exclusive, Attar, Mariotti, and Salanié (2011) and Kurlat (forthcoming) argue that buyers cannot screen sellers by restricting the quantity of assets that is sold. As in Bigio (2015) and Kurlat (2013), secondary market assets thus trade at a marginal price that is independent of (i) how many assets the originating banker sells and (ii) the quality of the underlying asset. This creates a pernicious motive for secondary market trading. Rather than selling assets to alleviate funding constraints, bankers may opt to produce low-quality, high-risk assets just to sell them. In equilibrium, bankers find it optimal to "shirk and sell" when the secondary market price is sufficiently high and financiers purchase a large number of assets.[3] Strong demand for secondary market assets therefore affords bankers the opportunity to sell off low-quality assets under the guise of borrowing capacity-enhancing risk transfer. Investment efficiency falls.

[3] One concern is why bankers *ever* sell high-quality assets on secondary markets, given that all assets trade at a pooling price. I circumvent this problem in reduced form by assuming that the banker must produce either only high-quality assets or only low-quality assets. This assumption is without loss of generality if financiers are always guaranteed to receive at least the *average* quality of all assets produced by a banker when purchasing claims on secondary markets. Under this restriction, bankers optimally monitor either all of their investments or none of their investments. In practice, secondary markets are structured to eliminate excessive "cream-skimming" by bankers. Sellers typically offer a whole portfolio of loans for sale, and buyers select the subset of loans they want to purchase. Buyers can guarantee themselves at least the average portfolio quality by using a random selection rule. The assumption can also be rationalized by fixed costs in the monitoring or screening of borrowers.

The secondary market price is determined endogenously by the net worth of financiers and bankers. When financiers have small net worth, the secondary market price is low and bankers sell assets only to increase their borrowing capacity. When instead financiers have large net worth, the secondary market price is sufficiently high that some bankers begin originating low-quality assets, even as financiers earn positive returns on average. Investment efficiency falls and financial fragility grows in the aggregate. The root cause of this inefficiency is a pecuniary externality. Individual financiers do not internalize that they worsen the pool of all assets by buying more assets on secondary markets.The welfare consequences may be severe. A partial destruction of financier wealth can lead to a *Pareto*-superior allocation. Policy that limits the accumulation of financier net worth or hampers financiers' ability to purchase excess amounts of loan-backed assets may therefore 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 model's key dynamic is the evolution of the intermediary net worth distribution. Because financiers buy aggregate risk exposure on secondary markets, their net worth typically grows faster than that of bankers during macroeconomic expansions. Credit volumes initially increase as financier net worth grows because bankers are able to sell off more risk exposure. Over time, however, rising secondary market prices induce a growing fraction of bankers to produce low-quality assets, leading to excess risk exposure in the financial system. Ultimately, a negative aggregate shocks is enough to trigger sharp collapses in secondary market trading and credit. Financier net worth falls because financiers end up holding a large fraction of low-quality assets. Credit volumes fall because bankers can no longer manage risk on secondary markets. Secondary markets recover slowly because financiers need time to rebuild their net worth. As a result, bankers grow vulnerable to negative shocks, and prolonged crises also harm bank balance sheets. Longer crises thus lead to slower recoveries. Because credit quality deteriorates gradually over the course of the boom, longer booms similarly lead to sharper crises. Gorton and Metrick (2012) and Krishnamurthy, Nagel, and Orlov (2014) provide evidence 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. Adrian and Shin (2010b) 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. Bigio (2014) provides evidence of the slow recovery of bank equity and interbank markets in the aftermath of the 2008 crisis. Schularick and Taylor (2012), Mendoza and Terrones (2012), and Reinhart and Rogoff (2009) provide evidence that longer credit booms predict sharper crises.

Credit booms driven by growing secondary markets can emerge even when bankers and financiers receive the same equilibrium return on equity. Indeed, financiers earn rents precisely *because* they take on aggregate risk exposure. As a result, they grow faster during booms even when earning the same average return. Moreover, there are asymmetries in how financiers and bankers achieve these (same) returns. When total intermediary net worth is scarce and funding interest rates are low, financiers employ more leverage than bankers, and earn disproportionately high returns when this risk-taking behavior pays off. Indeed, because bankers highly value increased borrowing capacity when interest rates are low, financiers earn large rents by taking on risk-exposure when funding is cheap. As a result, secondary market booms are triggered by strong demand for financial assets and low interest rates. Bernanke (2005), Caballero and Krishnamurthy (2009), and Caballero, Farhi, and Gourinchas (2008) argue that the early 2000s were characterized by a "global saving glut" that led to a large inflow of global savings in search of safe assets produced by the U.S. financial system. In my model, such inflows generate gradually falling asset quality because they trigger growing imbalances between financiers and bankers. To the extent that expansionary monetary policy leads to falling funding costs for intermediaries, the theory also generates a novel risk-taking channel of monetary policy that operates through the dynamics of financial intermediary net worth. Caballero and Krishnamurthy (2009) argue that monetary policy was indeed expansionary during the early stages of the U.S. credit boom.

Finally, I find that tight leverage constraints on bankers may lead origination incentives to deteriorate sooner than in their absence. Bankers use secondary markets to increase their leverage. When leverage is limited, secondary market supply falls and prices increase. Increasing prices in turn tempt bankers into shirking, with adverse aggregate consequences. The effects of policy must therefore be studied in the context of the aggregate financial system.

Related Literature. Beginning with Bernanke and Gertler (1989) and Kiyotaki and Moore (1997), a rich literature in macroeconomics has emphasized the role of borrower net worth and credit constraints in the amplification and persistence of macroeconomic fluctuations. Recent contributions include Brunnermeier and Sannikov (2014), Gertler and Kiyotaki (2010), Gertler and Karadi (2011), He and Krishnamurthy (2011) and Di Tella (2014). A common theme is that borrower or financial intermediary net worth serves to alleviate financial frictions and facilitates more efficient financial intermediation. I emphasize the *distribution* of net worth, and show how endogenous imbalances in this distribution can harm the efficiency of investment even when net worth increases in the aggregate. Adrian and Shin (2010b) and Adrian and Shin (2014) argue that intermediary *leverage*, rather than net worth alone, is a key determinant credit conditions. My paper is complementary to theirs in that I show how intermediary leverage is determined in the aggregate of the financial system. Lorenzoni (2008), Bianchi (2011), and Bianchi and Mendoza (2012) study pecuniary externalities during credit booms. Excessive leverage leads to inefficient fire sales during the ensuing bust. I show how pecuniary externalities can generate falling investment efficiency during the boom phase. Bigio (2014), Bigio (2015), and Kurlat (2013) study interbank market shutdowns during macroeconomic downturns, while Rampini and Viswanathan (2010) study risk management among heterogeneous agents. I study how excessive trade among intermediaries during upturns leads to falling asset quality.

A growing literature in macroeconomics and finance emphasizes that there is strong demand for safe assets and that safe assets are a key output of the financial system. The seminal paper in this literature is Gorton and Pennacchi (1990). Krishnamurthy and Vissing-Jorgensen (2012), Krishnamurthy and Vissing-Jorgensen (2015), and Gorton, Lewellen, and Metrick (2012) provide evidence of a safety premium and the role of the financial system in producing such assets. Gorton and Ordoñez (2013) provide a theoretical analysis of safe asset production. Caballero and Farhi (2014) study how safe asset shortages can lead to stagnation, while Caballero and Krishnamurthy (2009) link the demand for safe assets to financial intermediary leverage.

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, with busts being triggered by shocks that induce information acquisition. I emphasize the evolution of net worth and the deterioration of investment efficiency over the credit cycle. Gennaioli, Shleifer, and Vishny (2013) study role of securitization within the shadow banking sector in driving aggregate outcomes. Securitization allows for improved sharing of *idiosyncratic risk*, and is efficient unless agents neglect aggregate risk. I study the re-allocation of *aggregate risk* via securitization, 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 thus give a reason why booms endogenously lead to financial fragility.

A rich literature in financial economics emphasizes the role of risk in shaping intermediation incentives. Early examples are the risk shifting model of Jensen and Meckling (1976) and the model of delegated monitoring in Diamond (1984). I build on these micro-foundations by explicitly studying the process by which intermediaries diversify risk. The seminal study of loan sales by bankers is Gorton and Pennacchi (1995). In their model, banks must retain a fraction of any loan to ensure monitoring incentives, and do so in equilibrium. I differ in that I allow for shirking on the equilibrium path and focus the aggregate consequences of loan sales. More recently, 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 net worth. 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.

Adrian and Shin (2010a), Adrian and Shin (2009), and Stein (2012) study the role of monetary policy in shaping financial stability. In Adrian and Shin (2010a) and Adrian and Shin (2009), the emphasis is on the role of the short-term interest rate in driving the risk appetite and leverage of financial intermediaries and, thus, credit conditions and risk-taking. In Stein (2012), the main role of policy is to restrict the issuance of private money that relies excessively on short-term debt. I focus instead on how short-term interest rates shape the dynamics of intermediary net worth. By emphasizing the effects of asymmetrically regulating different classes of intermediary, my paper is also related to Plantin (2015), who discusses the role of differential regulation between a core banking system and a lightly regulated shadow banking sector and shows how relaxing core leverage requirements may make the financial system as a whole safer.

Layout. Section 2 presents a static model of financial intermediation in which the distribution of net worth is fixed. I use the static model to establish the key channels through which secondary market trading affects credit volumes and investment efficiency. In Section 3, I embed the static model into an overlapping generations framework to study the endogenous evolution of net worth. Section 4 studies policy. Section 5 concludes. All proofs are in Appendix A.

2 A Static Model of Financial Intermediation with Secondary Markets

I begin my analysis by studying a static model of financial intermediation with secondary markets. The distinguishing feature of this static model is that the net worth of all agents is fixed. I use this setting to characterize the role of secondary market trading for financial intermediation and to study comparative statics with respect to the net worth distribution. The key friction in the model is that the funding ability of intermediaries is limited by their risk exposure. In Section 3 I then embed the model into a dynamic framework to study the endogenous evolution of net worth.

2.1 Environment

There is a single period, comprising of multiple stages. The economy is populated by three types of agents, each of unit mass: *depositors* indexed by d, *bankers* indexed by b and *financiers* indexed by f. Depositors are outside investors that lend money to intermediaries to invest on their behalf. The key friction is that depositors have a strong preference for safe assets. As a result, all aggregate risk exposure must be held within the financial system.[4] Krishnamurthy and Vissing-Jorgensen (2012) and Krishnamurthy and Vissing-Jorgensen (2015) provide empirical evidence of this safety premium. I use the name "depositors" to indicate the risk-aversion of outside investors. Mapped into the real world, they may represent both individual depositors and financial institutions with a strong preference for safe assets, such as money market funds or pension funds. Bankers are unique in that only they can lend money to households and firms directly. Financiers purchase financial securities produced by bankers on secondary markets. Bankers and financiers partially finance their investments by borrowing from depositors, and are protected by limited liability. Because depositors are infinitely risk-averse, bankers and financiers borrow by issuing risk-free bonds subject to an endogenous risk-weighted borrowing constraints. Because financiers do not lend money to households and firms directly, they face a different borrowing constraint than bankers. These asymmetric borrowing constraints constitute the basic motive for trade on secondary markets.

2.2 Technology

There is a single good that can be used for consumption and investment. An agent of type $j \in \{d, b, f\}$ receives an endowment w_j at the beginning of the period. At the end of the period, an aggregate state of the world $z \in \{l, h\}$ is realized. The probability of state z is π_z. All agents derive

[4] All my results generalize to any setting in which financial intermediaries are constrained by their risk exposure. This may be the case even when depositors are, in principle, willing to hold risk exposure. In Hebert (2015), debt is the optimal security in settings that includes flexible moral hazard, i.e. an effort choice *and* risk shifting. In the Diamond (1984) model of delegated monitoring, the efficiency of financial intermediation improves when intermediaries can offload aggregate risk exposure.

utility from consumption at the end of the period. The consumption of agent i of type j in state z is $c_j^i(z)$. To capture depositors' preference for safe assets as simply as possible, I follow Gennaioli, Shleifer, and Vishny (2013) and Caballero and Farhi (2014) and assume that depositors are infinitely risk-averse and evaluate consumption streams according to $U_d^i(c_d^i) = \min_z c_d^i(z)$. Bankers and financiers are risk-neutral and evaluate consumption streams according to $U_b^i(c_b^i) = \mathbb{E}_z c_b^i(z)$ and $U_f^i(c_f^i) = \mathbb{E}_z c_f^i(z)$, respectively. This allows me to isolate how the risk exposure of intermediaries affects the efficiency of investment even when intermediaries are, in principle, indifferent towards holding risk.

Endowments can be invested into two constant-returns-to-scale investment opportunities: a *risky* technology indexed by R and a *safe* technology indexed by S. There are no capacity constraints – an infinite amount of capital can be invested either technology. The safe technology represents investment opportunities that do not *require* intermediation. For example, all agents in the economy can purchase treasury bills and widely traded AAA-rated corporate bonds. However, I assume that intermediaries may receive a higher return on the safe technology than depositors. Specifically, the safe technology yields a return of $\bar{y}_S$ per unit of investment in every state of the world when bankers or financiers invest, and a return of $\underline{y}_S \leq \bar{y}_S$ when depositors invest. Here $\bar{y}_S - \underline{y}_S \geq 0$ can be viewed as a cost advantage accruing to specialized financial intermediaries when investing in the safe technology. This could be due to economies of scale or informational costs. In the model, I will use $\bar{y}_S - \underline{y}_S$ to parametrize the *intermediation premium* that depositors are willing to pay for financial services. For simplicity, I normalize the safe technology's return to intermediaries to one: $\bar{y}_S = 1$. Agent i of type τ invests $k_{S,\tau}^i$ in the safe technology

The risky technology represents investment opportunities in the real economy, such as lending to households and firms. Only bankers can invest in this technology. The assumption here is that bankers have the requisite expertise to appropriately evaluate prospective borrowers and the technology to interact directly with households and firms. Banker i invests $k_{R,b}^i$ in the risky technology. The risky technology requires costly effort at the time of investment to operate efficiently. This assumption is motivated by the notion that bankers may have to engage in costly screening and monitoring to make sure that borrowers are likely to repay their loans and behave so as to maximize the expected returns on investment as in Holmstrom and Tirole (1997). For simplicity, I refer to all costly actions undertaken by the banker as *monitoring*, and to the absence of monitoring as *shirking*.

Monitoring has a utility cost of m per unit of investment. If monitored, the risky technology yields a return $y_R(z)$ per unit of investment in state z. If it is not monitored, it yields $y_R'(z)$ in state z. To simplify notation, I write $\hat{y}_R = \mathbb{E}_z y_R(z)$ and $\hat{y}_R' = \mathbb{E}_z y_R'(z)$ Monitoring is efficient, and shirking increases the downside risk: $\hat{y}_R > \hat{y}_R' + m$ and $y_R'(l) < y_R(l) < y_R'(h)$. When it is monitored, the risky technology yields a higher expected return but a lower worst-case return than the safe technology: $\hat{y}_R > \bar{y}_S$ but $y_R(l) < \underline{y}_S$. I let $e \in \{0, 1\}$ denote the monitoring effort exerted by the bank, with $e = 1$ if the bank monitors. The bank's monitoring decision is private information. Financiers and depositors thus do not know whether the claims on investment produced by bankers are of high-quality (monitored) or low-quality (unmonitored). Hence, there is moral hazard – monitoring occurs only if it is in the private interest of bankers to do so. Moreover, the monitoring decision applies to the banker's entire investment. That is, the banker produces either high-quality claims *or* low-quality claims but not both. This assumption is without loss of generality if financiers are always guaranteed to receive at least the *average* quality of all assets produced by a banker when purchasing claims on secondary markets. Under this restriction, bankers optimally monitor either all assets or on none. In the real world, secondary markets are structured to eliminate excessive "cream-skimming" by bankers. Sellers typically offer a whole portfolio of loans for sale, and buyers select the subset of loans they want to purchase. Buyers can guarantee themselves at least the average portfolio quality by using a random selection rule. The assumption can also be rationalized by fixed costs in the monitoring or screening of borrowers.

2.3 Asset Markets and Investment

Agents trade two financial assets: a risk-free bond and a risky claim. The risky claim is a direct claim on the output of the risky technology: it pays out $y_R(z)$ in state z if monitoring occurs, and $y'_R(z)$ otherwise. The banker's investment splits into a continuum of identical risky claims that can be traded individually. The risk-free bond is zero coupon bond with face value one. Bankers and financiers use the bond market to borrow funds from depositors. Bankers use risky claims trade risk exposure to financiers. Both financial assets are in zero net supply. I refer to bond market as the *funding market*, and the market for risky claims as the *secondary market*. The two markets open sequentially: the funding market closes before the secondary market opens. Investment occurs after the funding market has closed, but before the secondary market opens. All investment choices are not contractible: each agent makes individually rationally investment choices conditional on the bond holdings determined in the funding market.

The role of secondary markets is to allow to bankers to sell off risk exposure in order to increase borrowing. To simplify the timing of the model, I assume that bankers can issue a commitment to sell *at least* $\underline{a}_b$ claims when secondary markets open. In this manner, the banker can expand borrowing through secondary market sales even though markets open sequentially. Yet, I also allow bankers to sell *more* than $\underline{a}_b$ claims should they find it optimal to do so ex-post. This captures the idea that bankers can credibly promise to sell to sell a given amount of loans – for example, by offloading credit risk from a previous origination round – while always being able to return to secondary markets at a later date.[5]

Given these assumptions, I now detail the market structure in each market. I summarize the timing of events in Figure 1 below.

2.3.1 Funding Market Structure

In the funding market, financiers and bankers issue risk-free zero-coupon bond with face value one to depositors in order to fund investment and risky claim purchases. They do so subject to

[5] Bankers may find it optimal ex-post to shirk and sell a large fraction of his assets to financiers. If bankers do so, however, depositor payoffs are not adversely affected. As a result, there is no incentive for *depositors* to require bankers to commit to not selling more than $\underline{a}_b$, even if doing so were possible. Such ex-post shirking will affect financier's payoffs in secondary markets, however. I discuss this issue in detail below.

a solvency constraint – to be specified below – that ensures that all bonds are indeed risk-free. Before bond trading commences, each banker posts a commitment to sell at least $\underline{a}_b$ risky claims when the secondary market opens. As I will show below, these asset sale commitments will affect the tightness of the banker's solvency constraint. Given that the solvency constraint must hold for every banker and every financier, all bonds are identical. I therefore model the bond market as perfectly competitive, with price Q_b and return $R_b = \frac{1}{Q_b}$. Depositor i purchases b^i_d units of the bond, and intermediary i of type τ issues b^i_τ units of the bond subject to the solvency constraint.

The key simplifying assumptions of this funding market structure are that (i) financiers do not fund bankers by buying bonds and (ii) bankers do not issue equity to financiers. I show below that these assumptions are immaterial to the main results of the paper. Specifically, credit booms can arise even when financiers achieve weakly higher returns on equity than bankers – so that bankers would not want to issue equity, even if doing so were costless – and the return on secondary market assets strictly dominates the return on bonds. Moreover, market 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 bankers.

2.3.2 Secondary Market Structure

The secondary market opens after the funding market closes and is organized in two stages: bidding and trading. Banker i enters the bidding stage having issued b^i_b bonds and a promise to sell at least $\underline{a}^i_b$ risky claims. Financiers observe the pair $\mu \equiv (\underline{a}_b, b_b)$ associated with every risky claim that is sold. That is, each financier knows the bond position and asset-sale promises made by the banker issuing the risky claim. Because bankers can sell assets to many financiers at the same time, trade is non-exclusive. Financiers thus cannot directly observe either the quality of the claims or the total quantity of risky claims sold given banker. The motivation for this assumption is as follows. First, bankers have better information about their own actions. Second, secondary markets are typically large and opaque. Indeed, many financial securities are traded in over-the-counter markets that are hard to monitor in real time. Attar, Mariotti, and Salanié (2011) and Kurlat (forthcoming) show that buyers cannot screen sellers by restricting the quantity of assets that is sold when trade is non-exclusive. That is, bankers cannot signal that they engaged in costly monitoring by promising to retain a fraction of their assets.[6] I thus restrict the contract space in secondary markets to menus consisting of a *per-unit* price $Q_a(\mu)$ and a quantity $a_f(\mu)$ that the financier is willing to purchase at $Q_a(\mu)$. These bids are conditional on μ because financiers can use μ to make inferences about the quality of risky claims. Bankers then sell risky claims to the highest bidder.

This market structure allows me to tackle two concerns that would arise in a standard competitive market. The first is that financiers are able to form inferences about asset quality as a function of bankers funding market choices. As a result, multiple asset qualities can trade simul-

[6] One way to overcome such "anonymity" in financial markets is to allow for reputations. Yet reputations are typically fragile – see Ordoñez (2013) – and may even serve to *sustain* pooling equilibria in which both low-quality and high-quality assets are sold in dynamic settings (Chari, Shourideh, and Zetlin-Jones (2014)). As a result, they may attenuate, but don't eliminate, the potential for harmful hidden trading in secondary markets. I thus abstract from reputational concerns for simplicity.

taneously. The second is that I can accommodate secondary market shutdowns. That is, there exist equilibria in which no claims are traded on secondary markets. This feature will turn out to be useful to guarantee equilibrium existence more generally. The structure nevertheless preserves an appealing feature of competitive markets. Specifically, financiers act as *price takers* for any given μ when assets are traded on secondary markets. The result is that the intermediation rents on secondary markets are split according to secondary market prices, with market prices in turn being determined by the relative wealth of bankers and financiers.

At the commencement of the bidding stage, each financier j posts a pair $\left(Q_a^j(\mu_b), \hat{a}_f^j(\mu_b)\right)$ for all possible pairs of banker asset sale promises and bond issuances μ. Note that I require financiers to post prices and quantities for *any* μ, regardless of whether any banker posts such a μ in equilibrium. Since financier offers are a function of μ, I refer to all trades with bankers who post μ as occurring on a *sub-market* μ. Naturally, banker i can only trade in sub-market μ^i. I denote the set of bankers who post μ by $\mathcal{I}_\mu^b$, and the set of financiers who trade in sub-market μ by $\mathcal{I}_\mu^f$. A sub-market is **active** if a strictly positive mass of bankers and financiers makes strictly positive bids in this sub-market. The set of active secondary sub-markets is defined as $\mathcal{M}$.

In the trading stage, banker i in sub-market μ^i offers to sell $\hat{a}_b^{i,j}$ units of the risky claim to financier j at the posted price $Q_a^j(\mu^i)$. Given bankers' offers $\left\{a_b^{i,j}\right\}_{i\in\mathcal{I}_{\mu^i}^b}$ to financier j, risky claims are allocated as follows. If $\hat{a}_b^j(\mu) \geq \int_{i\in\mathcal{I}_\mu^b}\hat{a}_b^{i,j}$, so that there is excess demand for risky claims, then each banker sells exactly $a_b^{i,j}$ risky claims to financier j at price $Q_a^j(\mu)$. If $\hat{a}_b^j(\mu) > \int_{i\in\mathcal{I}_\mu^b}\hat{a}_b^{i,j}$, so that there is excess supply of risky claims, then each fraction of risky claim supplied is sold to the financier with equal probability, with the total amount of claims sold equal to financier j's demand. Financier j therefore receives exactly $a_b^j \equiv \min\left\{\int_{i\in\mathcal{I}_\mu^b}\hat{a}_b^{i,j}, \hat{a}_f^j\right\}$ units of the risky claim at price $Q_a^j(\mu)$, while banker i sells

$$a_b^{i,j} = \min\left\{\left(\frac{\hat{a}_b^{i,j}}{\int_{i'\in\mathcal{I}_\mu^b}\hat{a}_b^{i',j}}\right)\hat{a}_f^j, \hat{a}_b^{i,j}\right\}$$

risky claims to financier j. I refer to these as *realized* quantities. Across all financiers, the banker sells $a_b^i = \int_{j\in\mathcal{I}_\mu^f} a_b^{i,j}$ risky claims and receives $\int_{j\in\mathcal{I}_{\mu^i}^f} Q_a^j(\mu^i)a_b^{i,j}$ in revenue. By the law of large numbers, these quantities are not random variables. Finally, banker i's stage-2 bidding strategy must satisfy $a_b^i \geq \underline{a}_b^i$. Whenever multiple financiers make identical bids in a given sub-market, each financier receives a representative slice of all risky claims in that market. Specifically, if both high and low-quality loans are traded in a given sub-market then the fraction of low-quality loans received is the same for every financier. When bidding, the financier must therefore form beliefs only about the *average* quality risky claims in sub-market μ. A sufficient statistic is, of course, the fraction of low-quality loans. I denote financiers' beliefs about this fraction by $\phi(\mu)$. Throughout, I require that financier beliefs satisfy Bayes' rule wherever possible. In equilibrium, furthermore, beliefs must be correct – they must coincide with the true fraction of low-quality loans. To economize on notation, I take this condition as given and use $\phi(\mu)$ to denote the equilibrium fraction of low-quality loans in sub-market μ.

When turning to the characterization of equilibrium, I will also require that financier bids satisfy a regularity condition across sub-markets. In particular, I impose that the terms of trade offered by financiers must not "decrease" in beliefs.

Definition 1 (Bid Consistency)**.**
The bidding behavior of financiers satisfies ***bid consistency*** *if for all sub-markets* $\mu \in (\underline{a}_b, b_b) \in R_+^2$ *and* $\mu' \in R_+^2$ *, if* $\phi(\mu',) \leq \phi(\mu)$ *then* $Q_a^j(\mu') \geq Q_a^j(\mu)$ *and* $\hat{a}_f^j(\mu') \geq \hat{a}_f^j(\mu)$*.*

Bid consistency requires that financier bids be conditioned on the *quality* of risky claims only: whenever the financier believes assets to be of weakly higher quality in one of two sub-markets, he cannot offer worse terms in the sub-market where he believes the quality to be higher. I impose this restriction to prevent "collusive" equilibria in which financiers coordinate to punish bankers for deviating from some μ to a μ' by offering low prices when doing so does not change the quality of claims. It is directly linked to my assumption that secondary markets are *anonymous*: financiers must make offers conditional on their beliefs regarding the quality of the assets rather than the identity or balance sheet characteristics of the issuer. Note that bid consistency is not a constraint on the bidding behavior in *active* sub-markets since it is *implied* by a *no-arbitrage condition* stating that financiers do not achieve strictly higher returns in one active sub-market than in another. This no-arbitrage condition must hold in equilibrium whenever there are multiple active sub-markets. Bid consistency thus only constrains financier bids in currently inactive sub-markets to be consistent with those in active sub-markets. It is in this manner that the restriction rules out the aforementioned collusive equilibria.

The next result simplifies the analysis by showing that each *active* sub-market behaves as if it were a competitive market.

Lemma 1 (Secondary Market Prices and Rationing)**.**
If every sub-market μ*, there exists a unique marginal price* $Q_a(\mu)$ *such that* $Q_a^j(\mu) = Q_a(\mu)$ *for all* j*. No individual banker or financier is rationed at* $Q_a(\mu)$ *in any sub-market* μ*.*

The proof is standard and follows from all financiers being infinitesimally small and holding the same beliefs. Since no agent can impact market quantities in the aggregate, no agent can acquire risky claims below the marginal price. Yet no agent must pay more to acquire as many claims as he wants. This line of reasoning also accommodates the requirement that financier bids must satisfy bid consistency. The reason is that bid-consistency is implied by a no-arbitrage condition for financier's across active sub-markets – a condition that must hold when there are multiple active sub-markets – while bids on inactive sub-markets are irrelevant for financier utility because they are never accepted on the equilibrium path. Going forward, I thus assume that financiers take the set of active sub-markets and the market price within that sub-market as given. Accordingly, I write all decision problems in terms of market prices rather than bid prices, and *realized* quantities a_b and a_f rather than *bid* quantities $\hat{a}_b$ and $\hat{a}_f$. Note that all financiers receive a representative slice of all claims traded in all sub-markets in which they trade because all financiers bid the same marginal price in that sub-market.

Figure 1 summarizes the timing of events. In stage 1, all agents receive their endowments. In stage 2, bankers post a commitment to sell at least $\underline{a}_b$ units of the risky asset in secondary markets, and risk-free bonds are traded in the funding market. Once the funding market closes, we move on to stage 3. Here, bankers make investments in the risky technology using their own net worth and the proceeds from bond issuances in the funding market, and all agents make their investments in the safe technology. Moreover, bankers make their monitoring decision. The secondary market opens in stage 4. Financiers post bids for risky claims and bankers choose which financier to sell to, subject to the constraint that they must sell at least $\underline{a}_b$ units in total. To simplify notation, I take as given that the proceeds from risky asset sales are automatically invested in the storage technology, and thus yield a sure return of $\bar{y}_S \equiv 1$. Once the secondary market closes, we move on to stage 5. In this stage, the productivity shock z is realized, returns on investment accrue, accounts are settled, and all agents consume.

1. Agents receive endowments.
2. Bankers issue promise $\underline{a}_b$. Bond trading.
3. Agents invest and bankers make monitoring decision.
4. Secondary market trading. Bankers sell $a_b \in [\underline{a}_b, k_{R,b}]$ at price Q_a.
5. Aggregate state z and output realized. Accounts settled.

Figure 1: Timing of Events

2.4 Decision Problems

I begin the equilibrium characterization by discussing the decision problem of each type of agent. Given that the decision problems are symmetric for all agents of type j, I simplify notation by dropping the superscript i. Every agent takes the bond price, the set of active secondary sub-markets $\mathcal{M}$, and the marginal prices in each secondary sub-market as given.

2.4.1 The Depositor's Problem

The return on a risky claim purchased on secondary markets can never be higher than the direct return on investment on the risky technology. Hence, the worst-case return of a secondary market claim is always below that of the safe technology, and depositors invest only in the safe technology and/or risk-free bonds. Let $k_{S,d}$ and b_d denote the depositor's investment in the safe technology and bond purchases, respectively. Then the depositor's problem is

$$\begin{aligned}
\max_{k_{S,d},b_d} \quad & \min_z (c_d(z)) \\
\text{s.t.} \quad & c_d(z) = \underline{y}_S \cdot k_{S,d} + b_d \quad \text{for} \quad z \in \{l,h\} \\
& k_{S,d} + Q_b b_d \leq w_d.
\end{aligned}$$

The first constraint determines the depositor's consumption in state z. Since both investments are risk-free, consumption is independent of the state of the world. The second constraint is the budget constraint stating that the depositor cannot spend more than his endowment w_d on bonds and investment. This problem has a simple solution. In particular, the budget constraint binds and

$$b_d(w_d, Q_b) = \begin{cases} \frac{w_d}{Q_b} & \text{if } Q_b < \frac{1}{\underline{y}_S} \\ [0, \frac{w_d}{Q_b}] & \text{if } Q_b = \frac{1}{\underline{y}_S} \\ 0 & \text{if } Q_b > \frac{1}{\underline{y}_S} \end{cases}$$

Equilibrium bond prices are thus bounded above by $\bar{Q}_b \equiv \frac{1}{\underline{y}_S}$.

2.4.2 The Financier's Problem

To discuss the financier's problem, I first establish some additional notation. Recall that all financiers receive a representative risky claim in each sub-market, and that the fraction of low-quality claims sold in sub-market μ is $\phi(\mu)$. The state-z payoff of the representative claim in sub-market is $y_{R,\mu}(z) \equiv (1-\phi(\mu))y_R(z) + \phi(\mu)y'_R(z)$. The expected payoff is $\hat{y}_{R,\mu} \equiv (1-\phi(\mu))\hat{y}_R + \phi(\mu)\hat{y}'_R$. Financiers chooses investment in the safe technology $k_{S,f}$, a quantity of bonds to issue in funding markets b_f, and the number of risky claims to bid $a_f(\mu)$ for every $\mu \in \mathbb{R}^2_+$. Taking prices and the set of active secondary sub-markets $\mathcal{M}$ as given, the financier's problem is

$$\begin{aligned} \max_{k_{S,f}, b_f \geq 0, a_f(\mu)} \quad & \mathbb{E}_z \left[\bar{y}_S k_{S,f} + \int_{\mathcal{M}} y_{R,\mu}(z) a_f(\mu) d\mu - b_f \right] \\ \text{s.t.} \quad & k_{S,f} + \int_{\mathcal{M}} Q_a(\mu) a_f(\mu) d\mu \leq w_f + Q_b b_f \\ & \bar{y}_S k_{S,f} + \int_{\mathcal{M}} y_{R,\mu}(z) a_f(\mu) d\mu \geq b_f \ \text{ for all } \ z. \end{aligned}$$

The first constraint is the budget constraint. It states that sum of the expenditures on risky assets in all active sub-markets and the safe investment cannot exceed the sum of his net worth w_f and bond issuances b_f. The second constraint is a solvency constraint that ensures that the all debts are paid in full in every state of the world.

There are two main decisions: whether to purchase risky claims, and, if so, whether to issue bonds to do so. These decisions depend on the expected return of risky claims and their collateral capacity. In particular, financiers can issue more bonds more when the sub-market they are buying in has a higher proportion of high-quality claims. The reason is that the worst-case payoff of a highly-quality claim is strictly higher than that of a low-quality claim. As a result, it is of better use as collateral. Depending on prices and asset quality, financiers may issue less bonds than the solvency constraint allows them to. For example, the bond price may be so low that it is not profitable for the financier to issue bonds to invest in risky claims. I summarize the financier's optimal borrowing decision by γ in the pseudo-solvency constraint

$$b_f = \gamma \left[\bar{y}_S k_{S,f} + \int_{\mathcal{M}} y_{R,\mu}(z) a_f(\mu) d\mu \right].. \tag{1}$$

Here, $\gamma \in [0,1]$ is a decision variable determining the degree to which the financier exhausts his borrowing capacity. When $\gamma = 0$, the financier does not issue any bonds. When $\gamma = 1$, the financier issues as many bonds as he can. I also make use of the following definition.

Definition 2 (Return on Investment in Secondary Markets)**.**
The unlevered expected return on investment *in sub-market* μ *is* $\hat{R}^{unlev}(\mu) \equiv \frac{\hat{y}_{R,\mu}}{Q_a(\mu)}$. *The* fully levered expected return on investment *in sub-market* μ *is* $\hat{R}^{lev}(\mu) = \frac{\hat{y}_{R,\mu} - y_{R,\mu}(l)}{Q_a(\mu) - Q_b y_{R,\mu}(l)}$. *The* maximal expected return *in sub-market* μ *is* $\hat{R}^{\max}(\mu) \equiv \max\left\{\hat{R}^{unlev}(\mu), \hat{R}^{lev}(\mu)\right\}$.

The unlevered return is achieved by purchasing the risky claim using own net worth only. The fully levered return is achieved by purchasing claims using own net worth and the full amount of bonds that can be issued. The following corollary states the condition under which leverage is beneficial to the financier.

Corollary 1.
In sub-market μ*, the maximum expected return is equal to the fully levered expected return if and only if* $Q_b \hat{y}_{R,\mu} \geq Q_a(\mu)$.

Proof. Follows directly from comparing the rates of return. □

I now turn to the financier's optimal portfolio. If there are no active secondary sub-markets, the solution is trivial. Specifically, the financier invests all his wealth in the safe technology. He issues bonds to do so only if $Q_b \geq \bar{y}_S = 1$. To the extent that this condition holds, it is easy to verify that the solvency constraint is never binding. As a result, every financier can issue an infinite amount of bonds. Hence $Q_b \leq \bar{y}_S$ when secondary markets are inactive, and financiers issue bonds only when there is excess demand at $Q_b = \bar{y}_S$.

Next, turn to financier portfolios when there *are* active secondary markets. For simplicity, take as given that only one sub-market, μ^* say, is active – this will be the case in equilibrium. For secondary markets to be active, financiers must be willing to purchase risky claims. Hence the return on risky claims must not be lower than that of the safe technology. That is, $\hat{R}^{\text{unlev}}(\mu^*) \geq \bar{y}_S$. The first question is whether financiers will invest in the safe technology.

Lemma 2 (Financier Safe Investment with Active Secondary Markets.)**.**
Assume that $\hat{R}^{unlev}(\mu^*) \geq \bar{y}_S$*. Then financiers are indifferent between the safe and the risky technology if* $\hat{R}^{unlev}(\mu^*) = \bar{y}_S$ and $b_f \leq \frac{y_R(l) w_f}{Q_a(\mu^*) - Q_b y_R(l)}$*, and strictly prefer to invest in the risky technology otherwise.*

Proof. See appendix. □

A corollary of this result is that the financier's solvency constraint can be written as a borrowing constraint that is independent of whether the financier invests in risky claims or the safe technology.

Corollary 2 (Financier Borrowing Constraint)**.**
The financier solvency constraint is equivalent to the borrowing constraint

$$b_f \leq \frac{y_R(l)w_f}{Q_a(\mu^*) - Q_b y_R(l)}.$$

Proof. Suppose that the financier invests in risky claims only. Then the result follows directly from re-arranging the solvency constraint. Suppose instead that the financier invests in the safe technology. By Lemma 2, the stated condition must hold. □

When secondary markets are active, financiers thus face a borrowing constraint even when investing in the safe technology. The reason is that secondary markets allow the financier to engage in *risk-shifting*. Given that risky claims always offer a weakly higher return than the safe technology, and that the financier's borrowing capacity is independent of his investment strategy, I proceed under the presumption that the banker invests in the risky technology only. I later verify this presumption. The pseudo-solvency constraint allows me to write the financier's bond issuances as

$$b_f = \frac{\gamma y_R(l)w_f}{Q_a(\mu^*) - \gamma Q_b y_R(l)}.$$

for some $\gamma \in [0, 1]$.

The optimal degree of borrowing then follows directly from Corollary 1: the financier levers fully when the levered return is strictly higher than the unlevered return. Specifically, the optimal γ is given by

$$\gamma^* = \begin{cases} 1 & \text{if } \frac{\hat{y}_{R,\mu^*}}{Q_a} > \frac{1}{Q_b} \\ [0,1] & \text{if } \frac{\hat{y}_{R,\mu^*}}{Q_a} = \frac{1}{Q_b} \\ 0 & \text{if } \frac{\hat{y}_{R,\mu^*}}{Q_a} < \frac{1}{Q_b}. \end{cases}$$

Under the presumption that secondary markets are active and financiers strictly prefer risky claims to the safe technology, the financier optimally chooses the following asset allocation:

$$a_f(\mu^*) = \frac{w_f}{Q_a(\mu^*) - \gamma Q_b y_{R,\mu^*}(l)} \quad \text{and} \quad b_f^* = \frac{\gamma y_{R,\mu^*}(l)w_f}{Q_a(\mu^*) - \gamma Q_b y_{R,\mu^*}(l)}.$$

Accounting for bid consistency then only requires that the financier makes weakly better bids in all inactive sub-markets in which beliefs are weakly higher than in the active sub-market, i.e. $\hat{a}_f(\mu) \geq a_f(\mu^*)$ for all μ such that $\phi(\mu) \geq \phi(\mu^*)$.

2.4.3 The Banker's Problem

I now turn to the banker's problem. I assume throughout that bankers receive strictly positive intermediation rents from investing depositors' money on their behalf. As a result, bankers want to issue as many bonds as possible. I will show that it may not always be feasible to sustain monitoring in equilibrium for all bankers. I therefore begin by characterizing the decision problem for a given action $e \in \{0, 1\}$. I denote the realized private benefit associated with e by $m^*(e) = (1-e)m$, and the associated return on the risky technology by $y_R^*(z, e) = e y_R(z) + (1-e)y_R'(z)$. The banker's state-$z$ consumption $c_b(z)$ is the sum of payoffs from investments in the safe technology, payoffs from the risky technology net of asset sales, proceeds from asset sales, and bond repayments. By limited liability, consumption is bounded below by zero. That is,

$$c_b(z, \underline{a}_b, a_b, k_{R,b}, k_{S,b}, b_b, e) \equiv \max \left\{ \bar{y}_S k_{S,b} + y_R^*(e,z)\left(k_{R,b} - a_b\right) - b_b + Q_a(\mu) a_b, 0 \right\},$$

where $\mu = (\underline{a}_b, b_b)$. Since the banker is risk-neutral, the banker's utility in state z is

$$u_b(z, \underline{a}_b, a_b, k_{R,b}, k_{S,b}, b_b, e) \equiv c_b(z, \underline{a}_b, a_b, k_{R,b}, k_{S,b}, b_b, e) + m^*(e) k_{R,b}.$$

The banker's optimal monitoring choice *conditional* on $(a_b, k_{R,b}, k_{S,b}, b_b)$ is:

$$e^*(\underline{a}_b, a_b, k_{R,b}, k_{S,b}, b_b) = \arg \max_{e' \in \{0,1\}} \mathbb{E}_z u_b(z, \underline{a}_b, a_b, k_{R,b}, k_{S,b}, b_b, e')$$

Secondary markets open after the bond market closes and investment has taken place. The banker's asset sales must therefore be ex-post optimal given $(\underline{a}_b, k_{R,b}, k_{S,b}, b_b)$. Note that the banker must sell at least $\underline{a}_b$ claims but can sell no more than $k_{R,b}$. When deciding on how many assets to sell, the banker takes into account that he will adjust his monitoring decision optimally. For example, a banker that sells a large fraction of his portfolio may decide to stop monitoring. As a result, asset sales are ex-post optimal if and only if

$$a_b^*(\underline{a}_b, k_{R,b}, k_{S,b}, b_b) = \arg \max_{k_{R,b} \geq a' \geq \underline{a}} \mathbb{E}_z \left[\max \left\{ \bar{y}_S k_{S,b} + y_R^*(e', z)\left(k_{R,b} - a_b'\right) - b_b + Q_a(\mu) a_b', 0 \right\} \right] + m^*(e') k_{R,b}$$
$$\text{where} \quad e' = e^*(\underline{a}_b, a_b', k_{R,b}, k_{S,b}, b_b).$$

Taking prices and action e as given, the banker thus solves the problem:

$$\max_{k_{S,b}, k_{R,b}, b_b, \underline{a}} \quad \mathbb{E}_z \left[\max \left\{ y_S k_{S,b} + y_R^*(e,z)\left(k_{R,b} - a_b\right) - b_b + Q_a(\mu) a_b, 0 \right\} \right] + m^*(e) k_{R,b} \qquad (\mathrm{P_B}(e))$$
$$\text{s.t.}$$
$$k_{S,b} + k_{R,b} \leq w_b + Q_b b_b,$$
$$b_b \leq \bar{y}_S k_{S,b} + y_R^*(e,z)\left(k_{R,b} - a_b\right) + Q_a(\mu) a_b \ \text{ for all } \ z,$$
$$e = e^*(a_b, \underline{a}_b, k_{R,b}, k_{S,b}, b_b),$$
$$a_b = a_b^*(\underline{a}_b, k_{R,b}, k_{S,b}, b_b).$$

The first constraint is the budget constraint, stating that total investment in the safe and the risky technology cannot exceed net worth and the proceeds from bond issuances. The second constraint is the solvency constraint that guarantees that all debts are repaid in full in every state of the world. The third constraint is the incentive compatibility constraint that ensures that action e is privately optimal. The fourth constraint ensures that asset sales are ex-post optimal. A helpful result is that bankers will never invest in the safe technology in equilibrium. I impose this result going forward.

Lemma 3 (No Safe Investment by Bankers)**.**
Bankers never invest in the safe technology: $k_{S,b}^* = 0$ in any equilibrium.

Proof. See appendix. □

I characterize the solution to the banker's problem in two steps. I first discuss ex-post optimal asset sales and the monitoring decision *conditional* on bond issuances, investment, and asset-sale promises. I then discuss optimal bond issuances, investment, and asset-sale promises, given that asset-sales and the monitoring decision are chosen optimally ex-post.

Ex-Post Optimal Asset Sales. The secondary market opens once the funding market has closed. As a result, bond issuances b_b, investment in the risky technology $k_{R,b}$, and the asset-sale promise $\underline{a}_b$ are all sunk. The banker makes his asset sale decision with the associated optimal monitoring decision in mind. Specifically, the banker chooses $e = e^*(a_b, \underline{a}_b, k_{R,b}, b_b)$ when he sells a_b risky claims. Two observations simplify the analysis. First, the banker's objective function is linear because the banker is risk-neutral and he is required to be solvent in all states of the world. As a result, the solution is bang-bang. That is, the banker either sells everything or just as much as he initially promised, $a_b^* \in \{\underline{a}_b, k_{R,b}\}$. Second, the banker will certainly shirk when he sells his entire portfolio because asset quality is irrelevant to the banker's utility when $a_b = k_{R,b}$. That is, $e^*(k_{R,b}, \underline{a}_b, k_{R,b}, b_b) = 0$.

For there to be monitoring in equilibrium, it must therefore be the case that bankers monitor when they sell just as much as they had promised. Assume for now that this is the case. Bankers then either sell $\underline{a}_b$ and monitor or sell $k_{R,b}$ and shirk. The payoffs of these two action profiles are as follows.

$$
\begin{aligned}
\text{Shirk and Sell:} \quad & Q_a k_{R,b} - b_b + m k_{R,b} \\
\text{Effort and Hold:} \quad & \hat{y}_R(k_{R,b} - \underline{a}_b) - b_b + Q_a \underline{a}_b
\end{aligned}
$$

Comparing payoffs yields a simple decision rule in the secondary market price.

Proposition 1 (Ex-Post Optimal Asset Sales and Monitoring)**.**
Assume that monitoring is optimal at $\underline{a}_b$. Then the banker sells $\underline{a}_b$ assets and monitors only if

$$Q_a \leq \bar{Q}_a(k_{R,b}, \underline{a}_b) \equiv \hat{y}_R - m\left(\frac{k_{R,b}}{k_{R,b} - \underline{a}_b}\right) \tag{IMP}$$

That is, bankers will choose to sell everything and shirk if the asset price is too high. Because $\bar{Q}_a(k_{R,b}, \underline{a}_b) \leq \hat{y}_R - m < \hat{y}_R$, this may be the case even as financiers continue to make receive rents on secondary market assets. It is for this reason that there is scope for ex-post shirking when financiers are well-capitalized and bid up prices. Going forward, I will refer to this upper bound on the secondary market price as the *implementation constraint*. Monitoring occurs in equilibrium only if this constraint is satisfied for a some bankers.

If instead bankers shirk even when they sell just as many claims as they had promised (that is, $e^*(\underline{a}_b, \underline{a}_b, k_{R,b}, b_b) = 0$), then bankers always shirk. In this case, the optimal asset sales follow an even simpler decision rule: sell as many claims as promised if the secondary market price Q_a is below the expected return on a low-quality claim $\hat{y}_R$, and sell all claims otherwise.

Borrowing Constraints. The next step is to characterize the banker's optimal choice of bonds b_b, investment $k_{R,b}$, and asset-sale promises $\underline{a}_b$. I do so under the presumption that monitoring is optimal. The previous section showed that monitoring can only be sustained if it is ex-post optimal to sell just as many claims as promised. I therefore presume that $a_b^* = \underline{a}_b$. I then derive the borrowing constraint that ensures that bankers monitor when they do indeed sell as many claims as promised. When turning to competitive equilibrium, I compute the competitive equilibrium under this presumption and then verify whether $a_b^* = \underline{a}_b$ in equilibrium.

There are two constraints that limit bankers' ability to issue bonds. The first is the solvency constraint that states the banker must be able to repay his debts in full in every state of the world. The second is the incentive compatibility constraint that ensures bankers prefer to monitor. This takes the form

$$\underbrace{\mathbb{E}_z\left[y_R(z)\left(k_{R,b}-\underline{a}_b\right)-b_b+Q_a\underline{a}_b\right]}_{\text{Payoff Conditional on Monitoring}} \geq \underbrace{\mathbb{E}_z\left[\max\left\{y_R'(z)\left(k_{R,b}-\underline{a}_b\right)-b_b+Q_a\underline{a}_b,0\right\}\right]+mk_{R,b}}_{\text{Payoff Conditional on Shirking}}.$$

It is straightforward to see that the incentive constraint binds before the solvency constraint. The reason is that the banker is less sensitive to downside risk when he shirks than when monitors because the limited-liability constraints binds earlier under shirking. I refer to bankers as *collateral-constrained* if the limited-liability constraint binds in the low state conditional on shirking. The banker is collateral-constrained at $(k_{R,b}, b_b, Q_a)$ if and only if

$$\underline{a}_b \leq \bar{a}_b(k_{R,b}, b_b, Q_a) \equiv \underbrace{\frac{b_b - y_R'(l)k_{R,b}}{Q_a - y_R'(l)}}_{\text{Collateral Shortfall}}.$$

When the banker is collateral-constrained, the incentive-compatibility constraint can be rewritten as a borrowing constraint of the form:

$$b_b \leq \underbrace{\left[\frac{\pi_h}{\pi_l}\left(y_R(h)-y_R'(h)\right)+y_R(l)-\frac{m}{\pi_l}\right]k_{R,b}+\left[Q_a-y_R(l)-\frac{\pi_h}{\pi_l}\left(y_R(h)-y_R'(h)\right)\right]\underline{a}_b}_{\equiv\text{Banker borrowing capacity }\ \bar{b}_b(k_{R,b},Q_a,\underline{a}_b)}$$

The next result shows that bankers can relax this borrowing constraint by selling risky claims on secondary markets if the moral hazard problem is a sufficiently severe *risk-shifting problem* (Jensen and Meckling (1976)).

Lemma 4 (Risk-shifting Problem)**.**
There is scope for secondary market sales $(a_b > 0)$ *to increase banker borrowing capacity if and only if*

$$y_R'(h) > \mathbb{E}_z y_R(z).$$

Proof. Since the secondary market price cannot be higher than the return on the risky technology (that is, $Q_a \leq \hat{y}_R$) there exists a Q_a such that the coefficient on a_b in the borrowing constraint is positive if and only if $y_R'(h) > \mathbb{E}_z y_R(z)$. □

Lemma 4 states that the losses from shirking must be sufficiently concentrated in the low state. The intuition is that secondary market sales serve as a form of insurance – the banker has more capital in the low state but less in the high state. For this insurance to be valuable, the banker must be constrained by lack of capital in the low state. This is the case when the returns of the risky technology are poor in the low state of the world, and particularly so when shirking. I will impose this condition throughout. In order to simplify the exposition, I use the following special case.

Assumption 1.
The returns of the risky technology in the high state are the same under shirking and monitoring:

$$y'_R(h) = y_R(h)$$

This assumption allows me to write the borrowing constraint purely in terms of low-state payoffs. As will become clear, the assumption is innocuous in terms of the main results of the paper.[7] In order to obtain easily interpretable closed-form solutions for equilibrium prices and trading behavior, I also sometimes specialize the shirking technology as follows.

Assumption 2.
The risky technology yields zero payoff in the low state conditional on shirking: $y'_R(l) = 0$.

I summarize the severity of the moral hazard problem by

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

This reduced-form 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 $\pi_l y_R(l)$, the output loss from shirking). High values of $\tilde{m}$ therefore indicate a loose banker moral hazard problem. Under Assumption 1, the banker's borrowing constraint can then be written as

$$b_b \leq \underbrace{y_R(l)\tilde{m}k_{R,b}}_{\text{Collateral Capacity of Risky Investment}} + \underbrace{\left(Q_a - y_R(l)\right)\underline{a}_b}_{\text{Secondary Market Insurance}}.$$

The banker can back his bonds with the worst-case payoff of the risky technology – appropriately discounted by $\tilde{m}$ to account for moral hazard – or with proceeds from secondary market sales. Exploiting the budget constraint $k_{R,b} = w_b + Q_b b_b$ reveals a constraint on investment that can be relaxed by net worth and risky claim sales:

$$k_{R,b} \leq \frac{w_b + Q_b\left(Q_a - y_R(l)\right)\underline{a}_b}{1 - Q_b y_R(l)\tilde{m}} \tag{2}$$

What happens when the banker is *not* collateral-constrained – that is, when asset sales $\underline{a}_b$ exceed the collateral shortfall $\bar{a}_b(k_{R,b}, b_b, Q_a)$? In this case, the limited-liability constraint does not

[7] A caveat applies if I were to allow for *tranching* on secondary markets. If tranching were allowed, then my results go through as long as the high-state payoffs are different $y_R(h) \neq y'_R(h)$.

bind conditional on shirking. As a result, the incentive-compatibility constraint becomes a *skin-in-the-game* constraint:

$$\underline{a}_b \leq \tilde{m} k_{R,b} \tag{3}$$

Secondary market sales no longer boost borrowing capacity. Indeed, selling too many assets now induces shirking. As a result, a banker will never issue a promise to sell more than a fraction $\tilde{m}$ of his portfolio. As the previous section has shown, of course, the fact that he does not *promise* more does not mean he will not sell more ex-post.

The impact of secondary market sales on borrowing capacity is therefore as follows. If bankers are collateral-constrained (that is, $\underline{a}_b < \bar{a}_b(k_{R,b}, b_b, Q_a)$), asset sales alleviate the borrowing constraint by improving the bank's collateral position. When instead bankers are not collateral-constrained (that is, $\underline{a}_b > \bar{a}_b(k_{R,b}, b_b, Q_a)$), then asset sales reduce the stake of the banker outcome of his risky investment and do not boost borrowing capacity.

But do bankers find it optimal to sell assets to alleviate borrowing constraints? It depends on the secondary market price. By selling a risky claim, the bank is able to issue $Q_a - y_R(l)$ additional bonds. Upon investing this cash, each unit of investment can be used to back another $y_R(l)\tilde{m}$ in bonds. By issuing $\frac{(Q_a - y_R(l))}{1 - Q_b y_R(l)\tilde{m}}$ in new bonds, the banker can thus increase investment by $\frac{Q_b(Q_a - y_R(l))}{1 - Q_b y_R(l)\tilde{m}}$. The cost is that the banker receives a return of Q_a rather than the expected value $\hat{y}_R$. Bankers thus sells risky claims only if

$$\left[\underbrace{\frac{Q_b \hat{y}_R - 1}{1 - Q_b y_R(l)\tilde{m}}}_{\text{Levered Return - Bond Repayment}}\right] \cdot \left[\underbrace{Q_a - y_R(l)}_{\text{Debt Capacity of Risky Claim}}\right] \geq \underbrace{[\hat{y}_R - Q_a]}_{\text{Secondary Market Discount}}$$

This condition implies a lower bound on the price of risky assets for trade to occur in secondary markets:

$$Q_a \geq \underline{Q}_a(Q_b) \equiv \frac{\hat{y}_R - y_R(l) + y_R(l)(1 - \tilde{m})\hat{y}_R Q_b}{Q_b \left[\hat{y}_R - y_R(l)\tilde{m}\right]}. \tag{4}$$

Note that $\underline{Q}_a(Q_b)$ is strictly *decreasing* in Q_b and $Q_a(\frac{1}{\hat{y}_R}) = \hat{y}_R$. That is, when Q_b is high, borrowing capacity is valuable and banks sell claims at a discount; when Q_b is at its lowest, bankers are willing to sell claims only at par. Going forward, it will be useful to distinguish two degrees of *secondary market liquidity*.

Definition 3 (Secondary Market Liquidity)**.**
Secondary market liquidity is ***high*** *if* $Q_a^* > \underline{Q}_a(Q_b^*)$ *and* ***low*** *if* $Q_a^* = \underline{Q}_a(Q_b^*)$.

That is, secondary market liquidity is low if prices are such that bankers are exactly indifferent toward selling assets to increase borrowing. In this case, financiers receive all intermediation rents from secondary market trading. If instead secondary market liquidity is high, bankers strictly prefer to sell assets to increase borrowing capacity, and receive intermediation rents from doing so. As a result, bankers sell exactly $\underline{a}_b^* = \bar{a}_b(k_{R,b}, b_b, Q_a)$ claims. I fully characterize the optimal banker portfolio in the next section, where I study competitive equilibria.

2.5 Competitive Equilibria in the Static Model

I now turn to characterizing competitive equilibria in the static model. Because monitoring is socially efficient, I look for equilibria in which as many bankers as possible monitor. A complication is that the implementability constraint (IMP) cannot be verified ex-ante because $\bar{Q}_a$ is a function of the optimal banker portfolio. I therefore use a guess-and-verify approach to computing equilibria. Specifically, I first conjecture that the equilibrium secondary market price Q_a^* does not exceed the upper bound $\bar{Q}_a$. Given this conjecture, all bankers monitor. I then compute the resulting equilibrium allocations, and verify whether Q_a^* does indeed satisfy the implementability constraint (IMP). If the constraint is violated, I construct equilibria in which some bankers shirk.

2.5.1 Benchmark Without Secondary Markets

To understand the role of secondary markets, I begin by establishing a benchmark without secondary market trading. It is straightforward to show that bankers must always monitor in the absence of secondary markets. If bankers were to shirk on the equilibrium path, the solvency constraint would guarantee that the banker is exposed to all downside risk. Since shirking is inefficient, the banker elects to monitor. The key upshot is that secondary market trading is a *necessary* condition for shirking: investment efficiency falls only if bankers have an opportunity to sell off assets ex-post. Given that no banker shirks, the optimal portfolios of bankers and depositors are

$$k_{R,b}^0 = \frac{w_b}{1 - Q_b^0 y_R(l)\tilde{m}}, \qquad b_b^0 = \frac{y_R(l)\tilde{m}' w_b}{1 - Q_b^0 y_R(l)\tilde{m}}, \qquad b_d^0 = \frac{w_d}{Q_b^0}.$$

Imposing the market clearing condition $b_b = b_d$ yields the equilibrium price

$$Q_b^0 = \min\left\{\frac{w_d}{(w_d + w_b) y_R(l)\tilde{m}}, \frac{1}{\underline{y}_S}\right\}.$$

Here, the $\min$ operator stems from a boundary constraint on the equilibrium price. In particular, depositors are indifferent between bonds and the safe technology when $Q_b = \frac{1}{\underline{y}_S}$. Aggregate investment is

$$k_{R,b}^0 = \min\left\{W_d + W_b, \frac{\underline{y}_S W_b}{\underline{y}_S - y_R(l)\tilde{m}}\right\}.$$

An equilibrium without secondary markets always exists. Bankers do post asset-sale promises if they do not expect financiers to buy assets; financiers do not bid if bankers do not post promises.[8] Active secondary markets thus require some degree of coordination between bankers and financiers.

Proposition 2 (Existence of Equilibrium without Secondary Markets)**.**
There always exists an equilibrium without trade on secondary markets.

[8] Bankers that did not post a promise only sell (high-quality) assets at the expected value $\hat{y}_R$. As a result, financiers are indifferent between buying risky claims and investing in the safe technology. Moreover, bankers find it optimal to shirk and sell at this price because $\hat{y}_R > \bar{Q}_a$.

Proof. See appendix. □

I will show below that an equilibrium *with* active secondary markets may fail to exist. The above proposition above thus guarantees the existence of competitive equilibrium more generally. To focus on the role of secondary markets in financial intermediation, I assume that the equilibrium with active secondary markets is selected whenever it exists.

2.5.2 Active Secondary Markets

I now study competitive equilibria with active secondary markets. For ease of exposition, I focus on pure-strategy equilibria but allow different groups of bankers to pursue different strategies. There are, potentially, two groups of bankers: those who monitor and those who shirk. I refer to bankers who monitor as the *high type* and bankers who shirk as the *low type*. I denote the fraction of shirking bankers by $\Phi \in [0, 1]$. The definition of competitive equilibrium is as follows.

Definition 4 (Competitive Equilibrium With Active Secondary Markets)**.**
A pure-strategy competitive equilibrium with active secondary markets is a bond price Q_b, a set of quantities for depositors $\{b_d, k_{S,b}\}$, a set of quantities and bidding strategies for financiers $\left\{k_{S,f}, b_f, \{Q_a(\mu), a_f(\mu)\}_{\mu\in\mathbb{R}^2}\right\}$, a set of quantities and a monitor decision for bankers $\{\underline{a}_b, b_b, k_{S,b}, k_{R,b}, e, a_b\}$, a non-empty set of active secondary sub-markets $\mathcal{M}$, beliefs $\hat{\phi}(\mu)$ for each sub-market, and a fraction of shirking bankers Φ such that:

(i) All agents optimize given prices, the set of active sub-markets, and the bidding behavior of all other agents.

(ii) Financier bids satisfy bid consistency *in accordance with Definition 1.*

(iii) Prices are such that the bond market and all active secondary sub-markets clear.

(iv) The monitoring decision of bankers is individually optimal.

(v) Beliefs are correct.

(vi) The competitive equilibrium is a full-monitoring equilibrium if $\Phi = 0$ and a shirking equilibrium if $\Phi \in (0, 1]$. A pooling equilibrium is a shirking equilibrium in which high-type bankers and low-type bankers pool in the funding market. That is, they issue the same quantity of bonds b_b and asset-sale promises $\underline{a}_b$, and make the same investment $k_{R,b}$.

(vii) In a shirking equilibrium, bankers who monitor obtain the same expected utility as bankers who shirk.

In Appendix B I show that there do not exist separating equilibria in which high-type bankers sell assets on secondary markets. The intuition for the result is as follows. If bankers who monitor were to trade on secondary markets, then by no-arbitrage financiers must offer a higher price in the high-type's sub-market. By bid consistency, financiers offer the same high price to any banker posting a μ consistent with monitoring incentives. Low-type bankers can then profitably deviate to a slightly lower asset-sale promise and sell assets at the higher price. For this reason, I focus on pooling equilibria going forward.

I also distinguish equilibria by the degree to which the financial system is borrowing constrained.

Definition 5 (Highly Constrained Financial System)**.**
The financial system is ***highly constrained*** *if the equilibrium bond price* Q_b^* *satisfies* $Q_b^* = \frac{1}{\underline{y}_S}$ *and* $b_f^*(Q_b^*, Q_a^*) + b_b(Q_a^*, Q_b^*) < \frac{w_d}{Q_b^*}$*.*

The financial system is thus highly constrained when (i) depositors are exactly indifferent between investing in risk-free bonds produced by intermediaries and directly investing in the safe technology, and (ii) the financial system nevertheless cannot absorb the entire wealth of depositors. This is the case when wealth of depositors is large relative to that of financiers and bankers.

Given these preliminaries, the goal is to characterize how secondary markets influence the volume and efficiency of investment in competitive equilibrium. I proceed in steps. I begin by characterizing optimal banker portfolios in both *full-monitoring equilibrium* and *shirking equilibrium*. I then study how changes in financier net worth affect equilibrium outcomes within in each class of equilibrium. Finally, I show that only shirking equilibria exist when financier net worth is above an endogenously determined threshold.

Consider the full-monitoring equilibrium first. The first step is to characterize the optimal bank portfolio. Because all bankers are symmetric and are presumed to monitor, is without loss of generality to focus on symmetric equilibrium strategies. As a result, there is a unique active submarket and a unique marginal secondary market price Q_a. I denote equilibrium strategies and outcomes under full monitoring by the superscript $*$. To derive particularly simple expressions, I impose Assumption 2 from now on and let $y_R'(l) = 0$. The main upshot is that the collateral shortfall now takes the form $\bar{a}_b(k_{R,b}, b_b, Q_a) = \frac{b_b}{Q_a}$. That is, to maximize borrowing capacity, the bankers sells assets until his secondary market revenue is exactly equal to his debt burden.

Proposition 3 (Banker Portfolio in the Full-monitoring Equilibrium)**.**
If secondary market liquidity is high, the optimal banker portfolio is

$$k_{R,b} = \frac{w_b}{1 - Q_b Q_a \tilde{m}}, \qquad b_b = Q_a \tilde{m} k_{R,b} \qquad \underline{a}_b = \tilde{m} k_{R,b}^*.$$

If secondary market liquidity is low, the optimal banker portfolio satisfies

$$k_{R,b} = \frac{w_b + Q_b(\underline{Q}_a - y_R(l))a_b}{1 - Q_b y_R(l)\tilde{m}}, \qquad b_b = \frac{y_R(l)\tilde{m} w_b + (\underline{Q}_a - y_R(l))a_b}{1 - Q_b y_R(l)\tilde{m}}, \quad a_b = \underline{a}_b \in [0, \bar{a}_b(k_{R,b}, b_b, Q_a)].$$

Proof. When secondary market liquidity is high, the banker promises to sell $\underline{a}_b = \bar{a}_b(k_{R,b}, b_b, Q_a)$ assets so as to maximize borrowing capacity. Moreover, the borrowing constraint binds: $b_b = \bar{b}_b(k_{R,b}, Q_a, \bar{a}_b)$. When instead secondary market liquidity is low, the banker is indifferent between issuing claims on secondary markets and retaining his entire portfolio. Hence, any asset sale between zero and the collateral shortfall $\bar{b}_b(k_{S,b}, k_{R,b}, Q_a, \bar{a}_b)$ is consistent with banker optimality. The result then follows from imposing the budget constraint. □

The degree of secondary market liquidity is a function of the relative net worth of bankers and financiers. In particular, secondary market liquidity is high if

$$a_f(Q_b^{**}) > \frac{\tilde{m}w_b}{1 - Q_b^* \underline{Q}_a(Q_b^{**})\tilde{m}},$$

where Q_b^{**} is the bond price that clears the funding market given $Q_a = \underline{Q}_a(Q_b^{**})$. That is, secondary market liquidity is high when there is excess secondary market demand when Q_a is at its lower bound. Since a_f is strictly increasing in w_f, secondary market liquidity is high when w_f is large relative to w_b. An implication is that financiers receive all rents from secondary markets when they are small relative to bankers, while bankers and financiers share secondary market rents when financiers are large. I show below that the allocation of intermediation rents across bankers and financiers will crucially determine the evolution of net worth.

To determine whether bankers will monitor in equilibrium, the key question is whether the equilibrium secondary market price is below the upper bound given in Proposition 1. Given that $\bar{Q}_a(k_{R,b}, \underline{a}_b) = \hat{y}_R - m\frac{k_{R,b}}{k_{R,b}-\underline{a}_b}$ is a function of the banker's portfolio, the optimal banker portfolio places bounds on $\bar{Q}_a$.

Corollary 3 (Bounding the Upper Bound)**.**
$\bar{Q}_a(k_{R,b}, \underline{a}_b) =\in [\hat{y}'_R, \hat{y}_R - m]$ in any equilibrium.

Proof. No banker promises to sell more claims than is optimal when secondary market liquidity is high. Moreover, bankers never short-sell risky assets. Hence $0 \leq \underline{a}_b \leq \tilde{m}k_{R,b}$. Evaluating $\bar{Q}_a(k_{R,b}, \underline{a}_b) = \hat{y}_R - m\frac{k_{R,b}}{k_{R,b}-\underline{a}_b}$ at $(k_{R,b}, \tilde{m}k_{R,b})$ and $(k_{R,b}, 0)$ gives the result. □

Bankers thus shirk for sure when the secondary market price exceeds $\hat{y}_R - m$. The crucial implication is that $\frac{\hat{y}_R}{\hat{y}_R - m} > \bar{y}_S$. That is, the return on a high-quality claim purchased on secondary markets is strictly higher than the return of the safe technology *even if* the secondary market price is high enough to induce shirking. But this means that sufficiently wealthy financiers may bid up secondary market prices enough to render full-monitoring equilibria unsustainable. The next step therefore is to characterize shirking equilibria.

There are two types of bankers in a shirking equilibrium – those who shirk and those who monitor. Equilibrium strategies are now symmetric within type. I denote the equilibrium portfolio of the high type by superscript H and that of the low type by L. Because the equilibrium features pooling in the funding market, high-type and low-type bankers issue the same amount of bonds, make the same asset-sale promises and invest the same amount of capital in the risky technology. Moreover, all bankers are symmetric in terms of their investment opportunities and net worth. The only way to sustain the coexistence of the two types is for all bankers to be indifferent between shirking and monitor. As a result, he secondary market price must exactly equal the upper bound $\bar{Q}_a$ defined in Proposition 1. Shirking equilibria can therefore also be interpreted as bankers' playing a mixed strategy.

Lemma 5 (Secondary Market Price in Shirking Equilibrium)**.**
In a shirking equilibrium, the implementability constraint is just binding and the secondary market price

satisfies

$$Q_a = \bar{Q}_a(k^H_{R,b}, \underline{a}^H_b) = \hat{y}_R - m\left(\frac{k^H_{R,b}}{k^H_{R,b} - \underline{a}^H_b}\right).$$

The upper bound $\bar{Q}_a$ is a function of the high-type's portfolio. In particular, it is decreasing in $\underline{a}_b$. This leaves open the possibility that the high-type banker will withdraw assets from secondary markets so as to receive a higher price. Yet precisely because prices are bounded above by a continuous function of $\underline{a}_b$, any such deviation cannot lead to a discrete price increase, even if the banker receives the highest possible price after the deviation. Moreover, withdrawing assets from secondary markets will typically lead to higher excess demand on secondary markets. As will become clear, higher excess demand implies more shirking in equilibrium, and thus strengthens the key results. For simplicity, I therefore focus on shirking equilibria in which the high-type banker chooses the same portfolio as in an effort equilibrium. This can be supported in an equilibrium by financiers offering the same Q_a in all sub-markets with $\underline{a}_b \leq \underline{a}^*_b$.[9] Because the low type differs only in the amount of assets sold ex-post and the monitoring decision, equilibrium portfolios are therefore as follows.

Proposition 4 (Optimal Banker Portfolios in Shirking Equilibrium)**.**
The high type's optimal portfolio is

$$k^H_{R,b} = k^*_{R,b}, \quad b^H_b = b^*_b, \quad a^H_b = \underline{a}^H_b = \underline{a}^*_b.$$

The low type's optimal portfolio is

$$k^L_{R,b} = k^H_{R,b}, \quad b^L_b = b^H_b, \quad \underline{a}^L_b = \underline{a}^H_b, \quad a^L_b = k^L_{R,b}.$$

The fraction of low-quality claims traded on secondary markets is

$$\phi = \frac{\Phi a^L_b}{\Phi a^L_b + (1-\Phi)a^H_b},$$

where $\phi \geq \Phi$ because the low type sells more assets than the high type.

The optimal portfolios of all agents are linear in net worth. This permits straightforward aggregation. The market clearing conditions are as follows.

(i) In a **full-monitoring equilibrium with high secondary market liquidity**, the market clearing conditions are:

$$\text{Primary Market}: \quad \frac{Q_a \tilde{m} w_b}{1 - Q_b Q_a \tilde{m}} + \frac{\gamma y_R(l) w_f}{Q_a - Q_b \gamma y_R(l)} = \frac{w_d}{Q_b}$$

$$\text{Secondary Market}: \quad \frac{\tilde{m} w_b}{1 - Q_b Q_a \tilde{m}} = \frac{w_f}{Q_a - Q_b \gamma y_R(l)}$$

[9] More generally, an open question is how prices are constructed upon deviations to inactive sub-markets. In the model, prices are determined by market tightness. Yet because there is no free entry, off-equilibrium market tightness cannot be determined by a zero-profit condition as in Guerrieri, Shimer, and Wright (2010). I sidestep this issue by focusing on the benchmark equilibrium that appropriately minimizes the degree of equilibrium shirking, and thus understates the key results of the paper.

(ii) In a **full-monitoring equilibrium with low secondary market liquidity**, the market clearing conditions are:

$$\begin{aligned}\text{Primary Market}: &\quad \frac{y_R(l)\tilde{m}w_b + (\underline{Q}_a - y_R(l))a_b}{1 - Q_b y_R(l)\tilde{m}} + \frac{\gamma y_R(l) w_f}{\underline{Q}_a - Q_b\gamma y_R(l)} = \frac{w_d}{Q_b} \\ \text{Secondary Market}: &\quad a_b = \frac{w_f}{\underline{Q}_a - Q_b\gamma y_R(l)},\end{aligned}$$

where $Q_a = \underline{Q}_a(Q_b)$.

(iii) In a **shirking equilibrium with high secondary market liquidity**, the market clearing conditions are:

$$\begin{aligned}\text{Primary Market}: &\quad \frac{\bar{Q}_a\tilde{m}w_b}{1 - Q_b\bar{Q}_a\tilde{m}} + \frac{\gamma(1-\phi)y_R(l)w_f}{\bar{Q}_a - (1-\phi)Q_b\gamma y_R(l)} = \frac{w_d}{Q_b} \\ \text{Secondary Market}: &\quad \Phi\left(\frac{w_b}{1 - Q_b\bar{Q}_a\tilde{m}}\right) + (1-\Phi)\left(\frac{\tilde{m}w_b}{1 - Q_b\bar{Q}_a\tilde{m}}\right) = \frac{w_f}{\bar{Q}_a - (1-\phi)Q_b\gamma y_R(l)}\end{aligned}$$

(iv) In a **shirking equilibrium with low secondary market liquidity**, the market clearing conditions are:

$$\begin{aligned}\text{Primary Market}: &\quad \frac{y_R(l)\tilde{m}w_b + (\underline{Q}_a - y_R(l))a_b}{1 - Q_b y_R(l)\tilde{m}} + \frac{\gamma y_R(l) w_f}{\underline{Q}_a - (1-\phi)Q_b\gamma y_R(l)} = \frac{w_d}{Q_b} \\ \text{Secondary Market}: &\quad \Phi\left(\frac{w_b + Q_b(\underline{Q}_a - y_R(l))a_b^H}{1 - Q_b y_R(l)\tilde{m}}\right) + (1-\Phi)a_b^H = \frac{w_f}{\underline{Q}_a - (1-\phi)Q_b\gamma y_R(l)},\end{aligned}$$

where $Q_a = \underline{Q}_a(Q_b) = \bar{Q}_a(k_{R,b}^H, a_b^H)$.

The fraction of shirking bankers Φ affects market clearing in two ways. First, it impacts the number of assets sold on secondary markets because low-type bankers sell more risky claims than high-type bankers. All else equal, increased shirking thus pushes down secondary market prices. Second, because $y_R'(l) = 0$, financiers cannot use low-quality claims as collateral for bonds. They thus borrow only against the fraction of high-type loans $(1-\phi)$ that they receive on secondary markets. This effect reduces the demand for risky assets and shrinks the supply of risk-free bonds. To economize on notation going forward, I use the following definition.

Definition 6 (Aggregate Leverage Ratios)**.**
*The **aggregate leverage ratios** of financiers and bankers are, respectively,*

$$\lambda_f \equiv \frac{1}{Q_a - (1-\phi)Q_b\gamma} \qquad \textit{and} \qquad \lambda_b \equiv \frac{\Phi + (1-\Phi)\tilde{m}}{1 - Q_b Q_a\tilde{m}}$$

The next step is to characterize equilibrium outcomes. I focus on how the distribution of net worth shapes the volume and efficiency of investment. Throughout, I denote the distribution of net worth by $\mathbf{w} \equiv (w_d, w_b, w_f)$, the **relative net worth of financiers** by $\tilde{\mathbf{w}} = \frac{w_f}{w_b}$, and **expected output** by $\hat{Y} \equiv [\Phi\hat{y}_R' + (1-\Phi)\hat{y}_R]\, k_{R,b}$. I first show how changes in financier net worth affect output and investment within each class of equilibrium. I then show that only shirking equilibria exist when financier net worth exceeds a threshold. A large financier sector therefore leads to falling investment efficiency.

Proposition 5 (Financier Net Worth and Equilibrium Outcomes)**.**
Assume that (w_d, w_b) *is such that investment is inefficient in the absence of secondary markets. Then:*

(i) In a full-monitoring equilibrium with high secondary market liquidity, the secondary market price Q_a*, total investment* $k_{R,b}$ *and expected aggregate output* $\hat{Y}$ *are* ***increasing*** *in financier net worth* w_f

(ii) In a full-monitoring equilibrium with low secondary market liquidity, the secondary market price is increasing in w_f*. Aggregate investment* $k_{R,b}$ *and aggregate expected output* $\hat{Y}$ *are strictly increasing in* w_f *if the financial system is highly constrained.*

(iii) In a shirking equilibrium, the share of shirking bankers Φ *is strictly increasing in* w_f*. Aggregate expected output is strictly decreasing in* w_f *if the financial system is highly constrained, or if financiers weakly prefer to not borrow.*

Proof. See appendix. □

The intuition behind the first part of the proposition is straightforward. As financier net worth increases, so does the demand for secondary market assets. Secondary market prices appreciate. When prices increase, bankers receive more collateral per risky claim sold. Borrowing and investment increase. Since all bankers monitor, expected aggregate output also increases. The difference between the first and the second part of the proposition is that, in a low-liquidity equilibrium, financiers receive all rents from secondary market trading. In this region of the state space, increases in financier wealth may increase the total supply of bonds more than banker's borrowing capacity, leading to drop in bond prices that crowds out banker borrowing. Nevertheless, increased secondary market demands leads to increase in the secondary market price and, as financiers grow even larger, investment volumes grow again. As the next proposition shows, the social benefits of increased financier net worth can be large.

Corollary 4 (The Social Value of Financier Net Worth)**.**
Fix a full-monitoring low-liquidity equilibrium with a highly constrained financial system. Then if $\underline{y}_S < \bar{y}_S$ *there exists a* $\Delta > 0$ *such that re-allocating* Δ *units of net worth from bankers to financiers strictly increases investment and expected output.*

Proof. See appendix. □

Reallocations of net worth towards financiers spur investment disproportionately when the financial sector is highly constrained $\left(Q_b^* = \frac{1}{\underline{y}_S}\right)$ and when intermediation is very valuable to depositors $\left(\underline{y}_S < \bar{y}_S = 1\right)$. Financiers are able to lever more than bankers when bond prices are high because they are not subject to the moral hazard problem at the investment stage. This advantage more than outweighs the direct costs of reducing bank net worth when bond prices are high. As a result, putting net worth in the hands of financiers allows for more investment in the aggregate than putting it in the hands of bankers. Growing secondary market volumes can thus trigger a credit boom that is larger than if bank net worth were to grow instead. This *social value* of financier net worth is reflected in the returns on equity earned by financiers and bankers.

Proposition 6 (Expected Return on Intermediary Equity)**.**
Fix a full-monitoring equilibrium with low-liquidity. The expected return on equity earned by bankers and financiers, respectively, is

$$R\hat{O}E_b = \frac{\hat{y}_R - y_R(l)\tilde{m}}{1 - Q_b^* y_R(l)\tilde{m}} \qquad \textit{and} \qquad R\hat{O}E_f = \frac{\hat{y}_R - y_R(l)}{\underline{Q}_a(Q_b^*) - Q_b y_R(l)}$$

Moreover, $R\hat{O}E_f > R\hat{O}E_b$ if $Q_b^* > 1$ and $ROE_f \leq ROE_b$ if $Q_b^* \leq 1$.

Proof. See appendix. □

That is, financiers receive higher returns on equity than bankers when the financial system is highly constrained and depositors pay a premium for intermediation services. This is the case even though financiers cannot invest in the risky technology *directly*, and thus are technologically inferior to bankers. In the dynamic model in Section 3, I show that the large rents earned by financiers when the aggregate net worth of intermediaries is low leads financiers to grow disproportionately when they are small initially.

The downside of increased financier net worth is that appreciating secondary market prices eventually induce some bankers to shirk.

Corollary 5 (Excessively Large Financier Net Worth)**.**
Suppose that the net worth of depositors and bankers (w_d, w_b) is such that bond price in the absence of secondary market Q_b^0 is such that $\underline{Q}_a(Q_b^0) < \hat{y}_R - m$. Then there exists a threshold level of financier net worth $\bar{w}_f(w_d, w_b) \geq 0$ such that the competitive equilibrium is a full-monitoring equilibrium if $w_f \leq \bar{w}_f(w_d, w_b)$ and a shirking equilibrium if $w_f > \bar{w}_f(w_d, w_b)$.

Why does the price adjustment mechanism break down in a shirking equilibrium? The implementability constraint (IMP) now restricts the appreciation of secondary market prices. Banker would prefer to sell and shirk if Q_a were to grow further. But if all bankers continue to monitor, Q_a cannot stay constant either – financier wealth is increasing, and so secondary markets would no longer clear. The solution is to have an increasing number of bankers shirk. Because low-type bankers sell more claims than high types, markets can clear at a constant price. This is the intuition behind the third part Proposition 5. Note that equilibrium shirking occurs even there is no financier irrationality or differential beliefs. Because financiers earn intermediation rents when purchasing assets from high-type bankers, they continue to earn rents even when some bankers shirk.

Why the caveat that $\underline{Q}_a(Q_b^0) < \hat{y}_R - m$? If this inequality were not satisfied, then bankers would never sells assets at a price that does *not* induce shirking if secondary markets were active. But if bankers shirk as soon as there is trade on secondary markets, then financiers do not buy assets in the first place. To see why, recall that bankers only sell assets if $\bar{Q}_a \geq \underline{Q}_a(Q_b)$, while $\bar{Q}_a \leq \hat{y}_R - m$. Because Q_b^0 is decreasing in w_b and $\underline{Q}_a(Q_b)$ is decreasing in Q_b, this implies that there are no equilibria with secondary market trading when w_b is large relative to w_d. That is, bankers use secondary markets only if there is sufficiently strong depositor demand.

The next result shows that the harmful effects of excessive financier net worth may be severe in equilibria with active secondary markets.

Corollary 6 (Pareto-improving Reductions in Financier Net Worth)**.**
Consider a shirking equilibrium in which either (i) financiers do not borrow, or (ii) the financial system is highly constrained. Then a partial destruction of financier net worth w_f is Pareto-improving.

Proof. See appendix. □

A pecuniary externality is at work. Increased financier wealth grows secondary market demand, but secondary market prices cannot appreciate beyond the upper bound $\bar{Q}_a$. When prices reach this upper bound, markets clear through quantities, inducing some bankers to shirk. Financiers impose an externality on each other by contributing to a decline in the average quantity of assets traded on secondary markets. This negative externality is severe enough that financiers can be made better off by a uniform destruction of their wealth. Moreover, falling asset quality exposes the financial system to more risk, creating financial fragility.

Figure 2 provides a numerical illustration of the equilibrium effects of w_f. Growing financier net worth initially increases investment and expected output, but gradually induces bankers to shirk. As a result, investment efficiency falls. Throughout the figure, red corresponds to a shirking equilibrium, and blue to a full-monitoring equilibrium. The top left panel depicts asset prices, with the upper line representing the risky claim price Q_a and the lower the bond price Q_b. As financier net worth increases, the secondary market price increases. In the full-monitoring equilibrium, this leads to a boom in investment and increases in expected output. As Q_a continues to rise, however, the full-monitoring equilibrium can no longer be sustained. In the shirking equilibrium, investment is now flat but expected output declines, as a larger fraction of bankers begins to shirk. This can be seen in the figure on the bottom left. The top right panel also 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* net worth of an intermediary at the end of the period, with the dotted lines corresponding to a high and low aggregate shock, respectively. As w_f increases, financiers take on more and more risk. Ultimately, financiers are the only agents in the economy exposed to any risk. For financiers, expected net worth is equivalent to expected utility. In a shirking equilibrium, increased financier wealth therefore *reduces* expected financier utility.

Overall, secondary markets play a dual role. If financier wealth is not too large, then secondary market trading expands investment by *more* than banker net worth. If instead financier wealth is large, then secondary market trading leads to deteriorating investment efficiency with potentially severe welfare consequences. The question is whether financier net worth might end up large enough to harm investment efficiency. The next question provides an affirmative answer

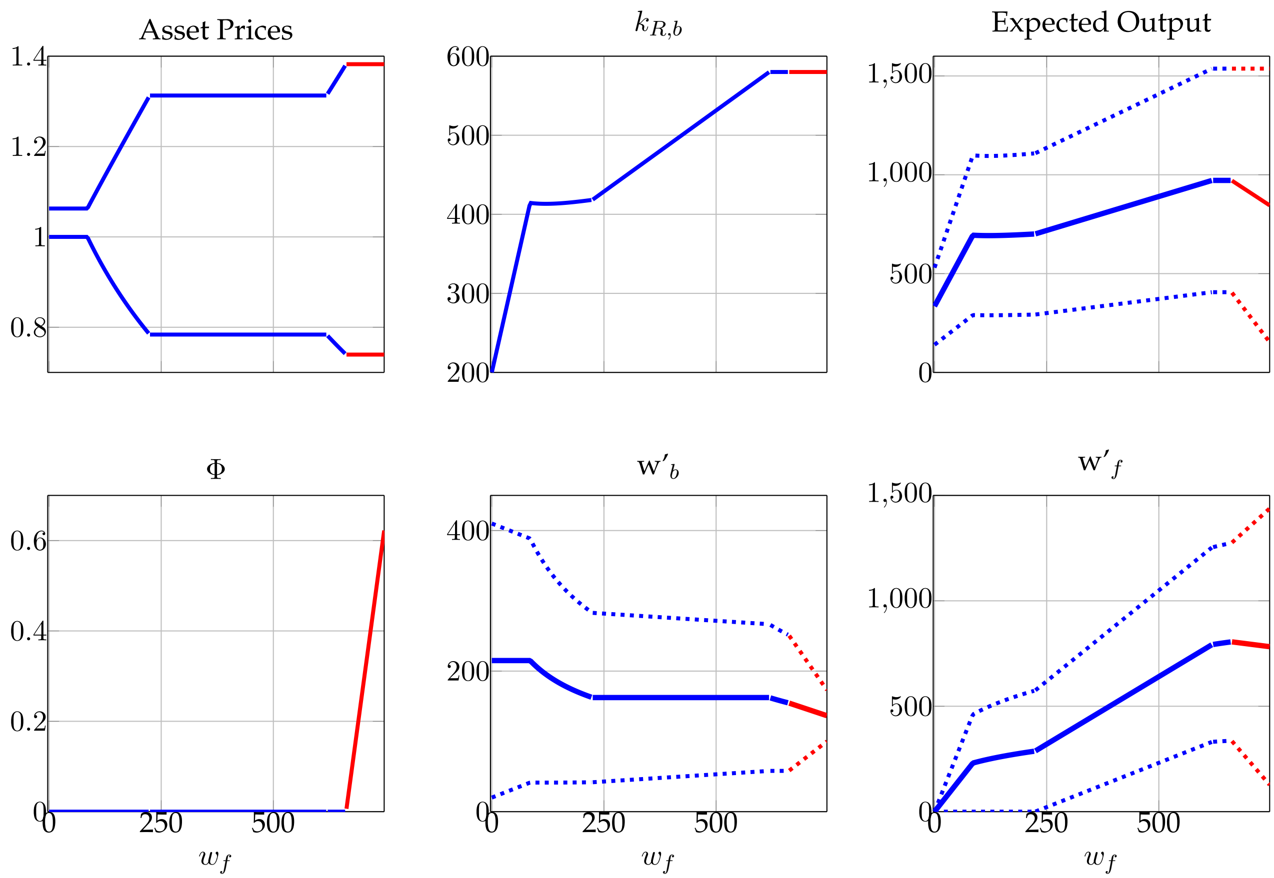

Figure 2: Equilibrium Outcomes as a Function of Financier Net Worth w_f.

in a dynamic framework.

3 A Dynamic Model of Secondary Markets

I now incorporate the static model into a dynamic setting to study the endogenous evolution of the net worth distribution. The key question is whether the net worth distribution moves towards regions of the state space in which only shirking equilibria can be sustained, even when starting out in the full-monitoring region.

Time is discrete and runs from 0 to $T \leq \infty$. A generic period is indexed by t. The model economy is populated by overlapping generations of financiers and bankers, each of whom lives for two periods, and short-lived depositors, each of whom lives 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 net worth good. Only the consumption good can be consumed. Intermediary net worth is special in that bankers and financiers must use net worth 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 net worth. When young, intermediaries and depositors play the intermediation game described in the static model. Before consuming their end-of-period net worth, they have the opportunity to sell it to the young. In this intergenerational net worth 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.[10] I assume that young generation makes a take-it-or-leave to the old. The old therefore always receive the consumption value of their equity capital.[11]

This construction implies that the objective function of a young intermediary is to maximize expected end-of-life net worth. The dynamic model then is equivalent to repeating the static model period-by-period, with the evolution of the net worth distribution linking equilibrium outcomes across periods. This allows me to parsimoniously illustrate the key forces that shape the evolution of net worth. In Appendix C I consider a variant of the model in which the old make take-it-or-leave-it offers to the young, and show that it can generate the same qualitative dynamics as the baseline model. I use the following definitions.

Definition 7 (Credit Booms)**.**
A credit boom of length t *is a sequence of t periods in which the total net worth of the financial sector $w_f + w_b$ and the total amount of risky investment $k_{R,b}$ increases every period. A* secondary market credit boom *is a credit boom in which the relative net worth of financiers $\tilde{\mathbf{w}}$ increases in every period. A* destabilizing *secondary market credit boom is a secondary market credit boom in which the economy transitions from a full-monitoring equilibrium to a shirking equilibrium.*

A necessary condition for credit booms to arise is a sequence of good aggregate shocks. Only when the net worth of the financial system grows can credit volumes increase. Whether the *relative* size of financiers increases – giving rise to a secondary market credit boom – will depend on the equilibrium allocation of risk. The next proposition characterizes the equilibrium evolution of relative financier net worth.

Proposition 7 (Equilibrium Evolution of Relative Net Worth)**.**
Let $\tilde{\mathbf{w}}'(z|\mathbf{w})$ denote relative net worth of financiers tomorrow conditional on productivity shock z and today's wealth distribution $\mathbf{w}$.

(i) The net worth of financiers and bankers increases upon a good aggregate shock and decreases upon a bad aggregate shock in any equilibrium.

*(ii) If $\mathbf{w}$ is such that today's equilibrium is **full monitoring** with **high** secondary market liquidity, then:*

$$\tilde{\mathbf{w}}'(h|\mathbf{w}) = \left[\frac{\tilde{m}(y_R(h) - \gamma(\mathbf{w})y_R(l))}{(1-\tilde{m})y_R(h)}\right].$$

[10] The role of the market maker is solely to make sure that each intermediary in every generation starts out with the same net worth. 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.

[11] I assume that if a generation of intermediaries has zero net worth 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.

*(iii) If $\mathbf{w}$ is such that today's equilibrium is **full monitoring** with **low** secondary market liquidity, then:*

$$\tilde{\mathbf{w}}'(h|\mathbf{w}) = \frac{(y_R(h) - \gamma(\mathbf{w})y_R(l))\lambda_f(\mathbf{w})\tilde{\mathbf{w}}}{\frac{y_R(h)-y_R(l)\tilde{m}}{1-Q_b\tilde{m}y_R(l)} - \frac{(1-\tilde{m})y_R(l)(y_R(h)-\hat{y}_R)}{\hat{y}_R-y_R(l)\tilde{m}}\lambda_f(\mathbf{w})\tilde{\mathbf{w}}}$$

*(iv) If $\mathbf{w}$ is such that today's equilibrium is **shirking** with **high** secondary market liquidity, then:*

$$\tilde{\mathbf{w}}'(h|\mathbf{w}) = \left[\frac{\Phi y_R(h) + (1-\Phi)\tilde{m}(y_R(h) - \gamma(\mathbf{w})y_R(l))}{(1-\tilde{m})(\Phi\hat{y}'_R + (1-\Phi)y_R(h))}\right]$$

*(v) If $\mathbf{w}$ is such that today's equilibrium is **shirking** with **low** secondary market liquidity, then:*

$$\tilde{\mathbf{w}}'(h|\mathbf{w}) = \frac{(\phi y'_R(h) + (1-\phi)(y_R(h) - \gamma(\mathbf{w})y_R(l)))\,\lambda_f(\mathbf{w})\tilde{\mathbf{w}}}{\frac{\Phi\underline{Q}_a+(1-\Phi)y_R(h)-y_R(l)\tilde{m}}{1-Q_b\tilde{m}y_R(l)} - \frac{(1-\tilde{m})y_R(l)(\Phi\underline{Q}_a+(1-\Phi)y_R(h)-\hat{y}_R)}{\hat{y}_R-y_R(l)\tilde{m}}\lambda_f(\mathbf{w})\tilde{\mathbf{w}}}$$

Proof. Follows directly from the optimal intermediary portfolios and exploiting the secondary market clearing condition to cancel out $k_{R,b}$ when liquidity is high. □

It is easy to verify that the evolution of net worth in a full-monitoring equilibrium is equal to that in shirking equilibrium when $\Phi = 0$. Moreover, the relative net worth of financiers *grows faster* after a good shock in a shirking equilibrium, holding secondary market liquidity fixed. The reason is that bankers sell off more risk exposure in a shirking equilibrium. The next examples provide some intuition as to these results.

Example 1 (High Liquidity and No Borrowing by Financiers)**.**
In a full-monitoring equilibrium in which secondary market liquidity is high and financiers do not borrow, the law of motion for relative net worth is:

$$\tilde{\mathbf{w}}'(z|\mathbf{w}) = \frac{\tilde{m}}{1-\tilde{m}} \qquad \textit{and} \qquad g(\tilde{\mathbf{w}}) = 0.$$

When secondary market liquidity is and financiers do not borrow, the relative wealth of financiers is fully pinned down by the severity of the banker's moral hazard problem. Specifically, financiers take on a fraction of $\tilde{m}$ of aggregate risk exposure, and bankers take on the remaining $(1-\tilde{m})$. A given aggregate shock therefore scales the wealth of bankers and financiers up or down while leaving relative net worth unaltered. Perhaps contrary to intuition, financiers end up being relatively wealthy when the banker's moral hazard problem is not too tight. The intuition is that collateral is valuable when bank's can issue a large quantity of bonds per dollar of collateral. Bankers thus have strong incentives to sell assets to financiers so as to increase borrowing capacity.

Example 2 (High Liquidity and Fully Leveraged Financiers)**.**
In a full-monitoring equilibrium in which secondary market liquidity is high and financiers are fully levered, the law of motion for relative net worth is:

$$\tilde{\mathbf{w}}'(z|\mathbf{w}) = \frac{\tilde{m}(y_R(z) - y_R(l))}{(1-\tilde{m})y_R(z)} \qquad \textit{and} \qquad g(\tilde{\mathbf{w}}) = 0.$$

When financiers are fully leveraged, they are exposed to more risk. To borrow, they pledge $y_R(l)$ to bondholders, and must repay this amount in any state of the world. As a result, their relative net worth is lower than when they are not levered. Because bond prices decrease during booms, financier leverage declines as well. Hence, the relative net worth of financiers grows during credit booms given that secondary market liquidity is high. Of course, secondary market liquidity is high only when financiers are sufficiently large to begin with. The question then is whether financiers may grow to be large even when they are small at first. To this end, I now characterize conditions under which relative financier net worth grows in a low-liquidity equilibrium.

Corollary 7 (Growth Rate of Relative Financier Net Worth With Low Liquidity)**.**
When secondary market liquidity is low, the growth rate of relative financier net worth in a full-monitoring equilibrium is strictly positive upon a positive shock if and only if

$$\underbrace{(1-\tilde{m})y_R(l)\left(\frac{y_R(h)-\hat{y}_R}{\hat{y}_R-y_R(l)\tilde{m}}\right)\lambda_f\tilde{\mathbf{w}}}_{\textit{Risk Transfer}} \geq \underbrace{\frac{y_R(h)-y_R(l)\tilde{m}}{1-Q_b y_R(l)\tilde{m}}}_{=ROE_b(h)} - \underbrace{(y_R(h)-\gamma(\mathbf{s})y_R(l))\lambda_f(\mathbf{w})}_{=ROE_f(h)}.$$

This inequality holds strictly for any $\tilde{\mathbf{w}}$ *if* $Q_b \geq 1$.

Proof. Follows directly from Proposition 6. □

The left-hand side of the inequality is the degree of risk transfer from bankers to financiers. It is increasing in relative financier net worth and financier leverage because the ability of financiers to take on credit is risk is limited by their wealth scaled by leverage. The right hand side is the difference between the state-h return on equity achieved by bankers and financiers, respectively. Proposition 6 showed that the right-hand side is strictly negative. Moreover, the left-hand side is strictly positive. No matter how small w_f is initially, the relative net worth of financiers thus grows as long as Q_b is sufficiently large.

More generally, relative financier net worth thus grows in a low-liquidity equilibrium when the aggregate net worth of the financial system is small relative to depositor net worth, or when financiers are not too small to begin with. Moreover, the relative net worth of financiers grows even when w_f is vanishingly small so long as interest rates are sufficiently low. The reason is that financiers can leverage more than bankers when interest rates are low. A financier can pledge the full worst-case payoff a risky loan $y_R(l)$, while a banker can only pledge $\tilde{m}y_R(l)$. This borrowing advantage translates into a disproportionate advantage for financiers in acquiring aggregate risk exposure. For any (w_f, w_b), there exists a w_d large enough such that $Q_b \geq 1$ in equilibrium. Increased demand for financial intermediation may therefore spur secondary market booms. Caballero, Farhi, and Gourinchas (2008) and Krishnamurthy and Vissing-Jorgensen (2015) argue that this pressure existed in the run-up to the 2008 financial crisis. Figures 3 and 4 depict the importance of initial conditions graphically. I plot the evolution of financier and banker net worth after a sequence of positive aggregate shocks. In both figures, the left panel depicts a baseline scenario in which financier net worth is smaller than bank net worth 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 net worth no longer catches up with banker net worth.

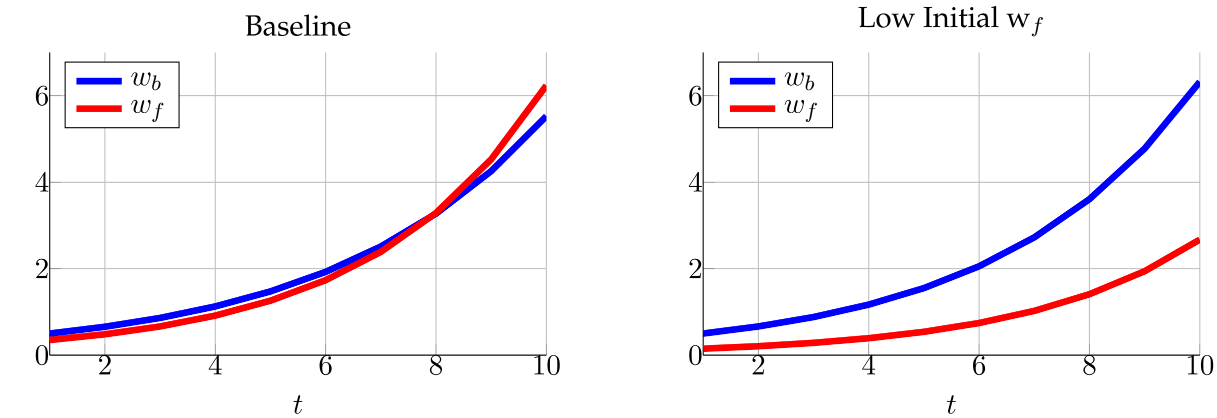

Figure 3: The effects of initial conditions – reduction in initial financier net worth w_f^0. Baseline parameter values: $\pi_h = 0.8, y_R(l) = 0.5, y_R(h) = 1.2, \tilde{m} = 0.82$. Initial net worth distribution: $(w_d, w_b^0, w_f^0) = (25, 0.5, 0.35)$. Comparative static: w_f^0 from 0.3 to 0.15.

Figure 3 shows the effect of a reduction in depositor net worth. Lower depositor net worth 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

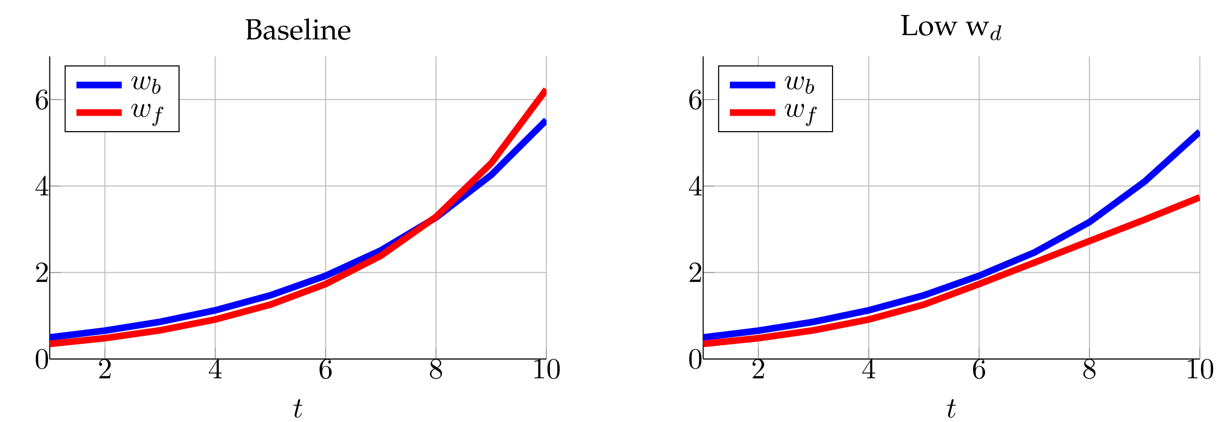

Figure 4: The effects of initial conditions – reduction in depositor net worth w_d. Baseline parameter values: $\pi_h = 0.8, y_R(l) = 0.5, y_R(h) = 1.2, \tilde{m} = 0.82$. Initial net worth distribution: $(w_d, w_b^0, w_f^0) = (25, 0.5, 0.35)$. Comparative static: w_d from 25 to 5.

suitable initial conditions, the model can therefore give rise to secondary market credit booms – period of credit growth during which financier net worth grows disproportionately. Furthermore, starting out in an equilibrium in which financiers borrow and moving to one in which they do not, relative net worth must grow upon a good shock. It follows that a series of good shocks pushes the economy towards the shirking region, with growing secondary market prices leading credit standards to deteriorate, even when starting out in the full-monitoring region. That is, given appropriate initial conditions, the model admits secondary market credit booms.

The next question is whether the model admits *destabilizing* secondary market credit booms.

To answer this question, the next proposition provides conditions under which a positive shock to the net worth of both bankers and financiers leads to an increase in secondary market prices and/or the fraction of shirking bankers. The general theme is that this is the case when the growth rate of relative financier net worth $g(\tilde{\mathbf{w}})$ is sufficiently large.

Proposition 8 (Relative Net Worth and Equilibrium Outcomes)**.**
Let $Q_b^*(\Phi, Q_a, \mathbf{w})$ denote the bond pricing function that clears bond markets for a given fraction of shirking bankers Φ, secondary market price Q_a and net worth distribution $\mathbf{w}$. Consider a positive shock ξ to financier and banker net worth.

1. *If secondary market liquidity is high and financiers are fully levered, then Q_a (in a full-monitoring equilibrium) and Φ (in a shirking equilibrium) are increasing in ξ if and only if*

$$g(\tilde{w}) \geq \left(\frac{Q_b^*}{\partial \xi}\right) \lambda_f \left(Q_a^* \tilde{w} - y_R(l)\right).$$

2. *If secondary market liquidity is high and financiers do not borrow, then Q_a (in a full-monitoring equilibrium) and Φ (in a shirking equilibrium) are increasing in ξ if and only if*

$$g(\tilde{w}) \geq \left(\frac{Q_b^*}{\partial \xi}\right) \tilde{w}$$

3. $\frac{Q_b^*}{\partial \xi} \leq 0$ *in any full-monitoring equilibrium.*

4. *If secondary market liquidity is low, Q_a is increasing in ξ in any full-monitoring equilibrium. If secondary market liquidity is low and the financial system is highly constrained, then Φ is increasing in ξ if*

$$g(\tilde{w}) \geq 0.$$

Proof. See appendix. □

Because $\frac{Q_b^*}{\partial \xi} \leq 0$ the secondary market price appreciates during booms with high-liquidity as long as the relative net worth grows weakly and $\tilde{w} \geq \frac{y_R(l)}{Q_a^*} \in (0,1)$. When secondary market liquidity is low, in turn, the secondary market price always appreciates. Secondary markets may thus be *destabilizing* in that the economy transitions into a shirking equilibrium over time.

3.1 Characteristics of Credit Booms

The last section showed that the model admits destabilizing secondary market credit booms. I characterize the properties of such booms in this section. I do so by computing equilibrium outcomes as a function of a time path for the exogenous shock z and the initial wealth distribution $\mathbf{w}^0 = (w_d^0, w_b^0, w_f^0)$. 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 destabilizing secondary market credit boom. I simulate the economy for 11 periods – an initial period, 8 positive shocks, a single negative shock, and then another positive shock. Financiers and bankers each start out with 0.5 units of net worth. Initial conditions are such that the economy starts out in a full-monitoring equilibrium. The left panel plots the evolution of net worth over time. There is a rapid build-up of net worth in the aggregate, with financiers growing faster. When a negative shock occurs, financier net worth collapses sharply because financiers are disproportionately exposed to risk. Banker net worth drops only moderately because financiers provide partial insurance to bankers. The middle panel plots the evolution of investment – equivalently, credit – over time. The blue line depicts total investment and the red line depicts the fraction of investment that goes to low-quality projects because bankers do not monitor. Initially all bankers monitor and there is no investment in low-quality projects. Over time, however, continued financier growth pushes the economy into a shirking equilibrium. As a result, the fraction of investment that flows to low-quality increases and grows steadily during the boom. In the aftermath of the crisis, investment falls. Yet because banker net worth only falls some, credit volumes recover quickly. 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 bankers are forced to monitor. During the boom phase, output increases steadily. During the crisis, output collapses sharply. As the comparison between the solid and dashed lines shows, almost one third of the drop is accounted for by falling credit quality over the course of the boom. Growing secondary markets can therefore generate credit booms that end in sharp crisis.

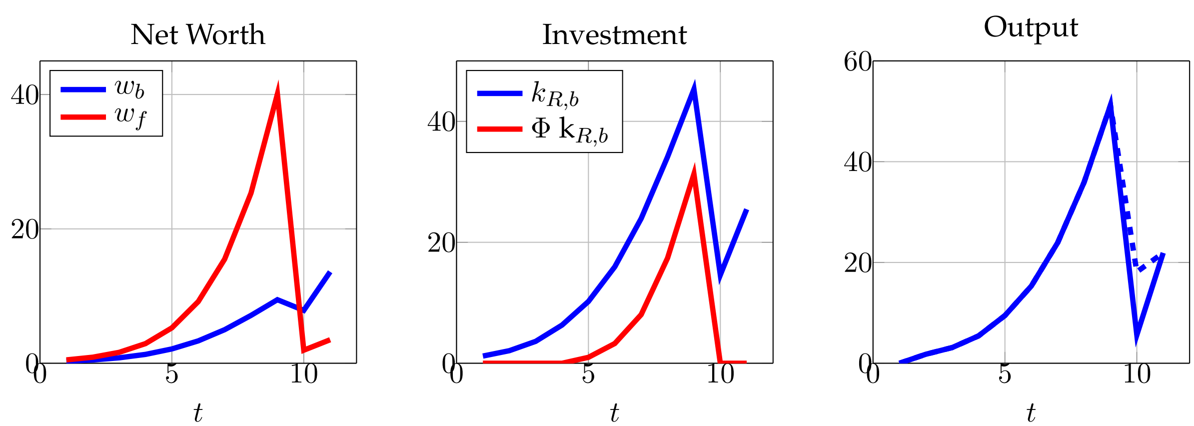

Figure 5: Secondary Market Credit Boom - low $\tilde{m}$. Parameter values: $\pi_h = 0.65, y_R(l) = 0.4, y_R(h) = 1.5, \tilde{m} = 0.82$. Initial net worth distribution: $(w_b^0, w_f^0) = (0.5, 0.5)$. Depositor wealth: $w_d = 450$.

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 banker's moral hazard problem is less severe and bankers can leverage each unit of net worth more. The time path of aggregate shocks and initial conditions are the same for both simulations. Three observations stand out. First, financier net worth grows *faster* when $\tilde{m}$ is large. This is perhaps counterintuitive given that increases in $\tilde{m}$ principally allow *bankers* to lever more. In equilibrium, however, the portfolios of bankers and financiers are intertwined. When $\tilde{m}$ is high, the shadow value of collateral is high for bankers. When banker net worth is scarce, bankers increase their collateral position by selling claims on secondary markets. Increases in $\tilde{m}$ thus boost supply and reduce prices in secondary markets. This allows financiers to lever more and purchase more risk exposure on secondary markets. In equilibrium, increased *scope* for bank may lead actual financier leverage to increase by more than actual bank leverage. Conditional on a good shock, financiers thus grow faster than bankers.

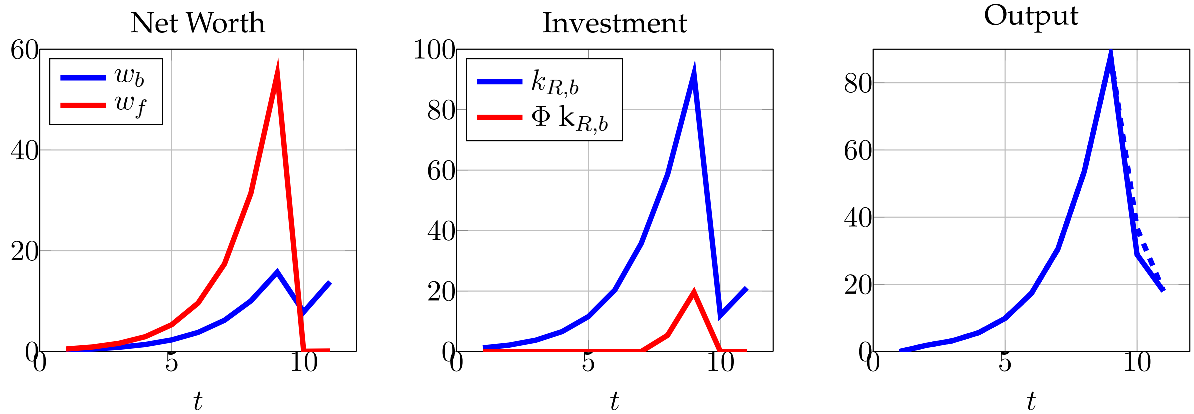

Figure 6: Secondary Market Credit Boom - high $\tilde{m}$. Parameter values: $\pi_h = 0.65, y_R(l) = 0.4, y_R(h) = 1.5, \tilde{m} = 0.85$. Initial net worth distribution: $(w_d^0, w_b^0, w_f^0) = (450, 0.5, 0.5)$.

Second, aggregate investment also increases faster because bankers 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 secondary market prices are relatively low. Because bankers sell off more risk exposure when $\tilde{m}$ is high, they suffer less in a crisis, and financier net worth declines disproportionately. Nevertheless, looser banker 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 the only difference being 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 net worth. 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 net worth 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 dynamics of aggregate productivity *during* a crisis episode shape the evolution of net worth 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 net worth, and shows how the model generates the migration of risk exposure back onto

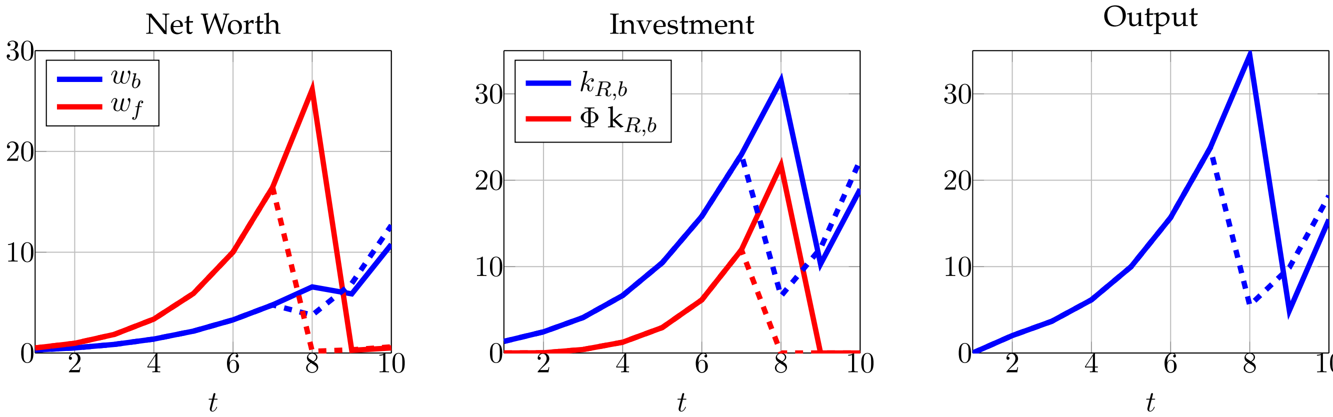

Figure 7: **Effects of increased boom length** - negative shock in period 9 (solid) vs. negative shock in period 8 (dashed).

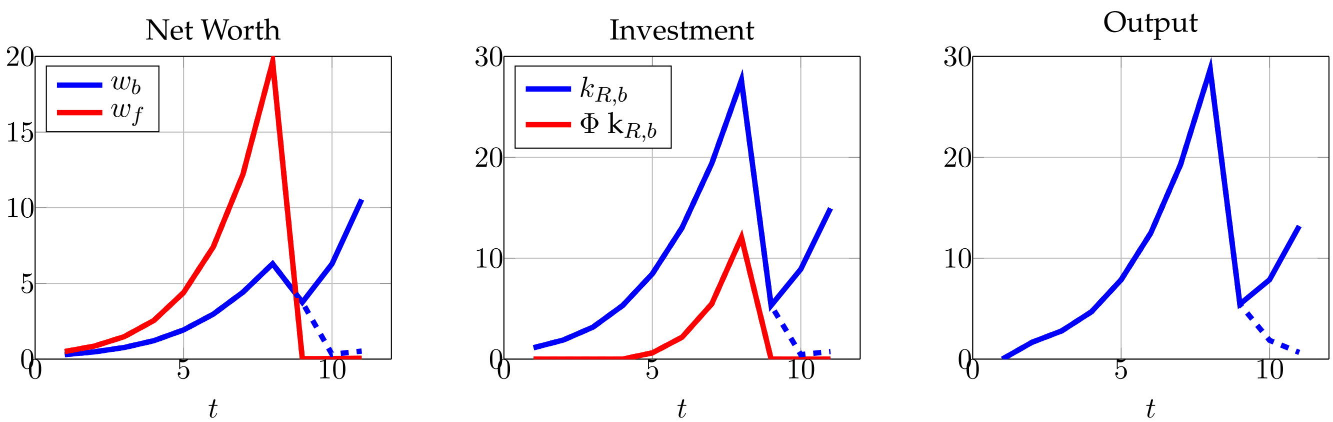

Figure 8: **Effects of increased crisis duration** - negative shock in period 8 (solid) vs. negative shock in periods 8 *and* 9 (dashed).

bank balance sheets once the initial negative shock has depleted financier net worth. In particular, the second negative shock leads to a dramatic fall in bank net worth, 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 bankers 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.

3.1.1 Returns on Equity during Secondary Market Credit Booms

Destabilizing secondary market credit booms are driven by a growing imbalance in the net worth of bankers and financiers. Do bankers 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 bankers. The next proposition shows that the model can generate destabilizing secondary market credit booms even when financiers always receive higher returns on equity than bankers.

Proposition 9 (Secondary Market Booms with High Financier ROE)**.**
There exist parameters such that (i) the model economy transitions from a full-monitoring equilibrium to a shirking equilibrium after a sequence of good shocks, (ii) relative financier net worth $\tilde{\mathbf{w}}$ *grows throughout, and (iii)* $\hat{ROE}_f > \hat{ROE}_f$ *throughout.*

Proof. See appendix. □

The intuition is that the harmful effects of secondary markets arise as a function of the imbalance between bankers 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 depositor net worth w_d such that financiers receive higher rents than bankers. As a result, the model's results are robust to allowing for endogenous equity issuances.

4 Policy

In this section, I ask how policy shapes the likelihood and evolution of secondary market credit booms. Rather than characterizing optimal policy, I evaluate three extant policies – monetary policy as a determinant of short-term interest rates, leverage requirements, and equity injections to kick-start lending in a crisis – from a positive perspective. I then propose a simple tool to eliminate pecuniary externalities in secondary markets.

4.1 Monetary Policy

I begin by studying the role of the monetary policy. I do so in reduced form. Specifically, I assume that monetary policy determines the depositor's investment opportunities outside of the financial system. That is,

$$\underline{y}_S = M(\rho),$$

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 $Q_b \leq \frac{1}{\underline{y}_S}$ 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 leads to a growth in financier net worth even when it is initially small, I assume that secondary market liquidity is low initially. The combination of these two assumptions implies that all bankers monitor.

Proposition 10 (Monetary Policy and Secondary Market Booms)**.**
Fix a full-monitoring equilibrium with low secondary market liquidity. Then a loosening of monetary policy (a reduction in ρ) increases investment and the growth rate of relative financier net worth after a good shock.

Proof. See Appendix. □

Due to asymmetric leverage constraints, loose monetary policy biases the growth rates of intermediary net worth towards financiers, even as both bankers 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 net worth 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 net worth, 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.

The monetary authority may find it difficult to reign in a secondary market boom once it is underway. Corollary 7 showed that the *growth rate* of relative financier net worth is increasing in relative net worth itself: financiers grow faster when they are already large. An initial monetary policy boost to financier net worth 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.

4.2 Capital Requirements

Now consider capital requirements. Specifically, assume that the government puts in place leverage limits $\bar{\lambda}_b$ and $\bar{\lambda}_f$ such that total investment by bankers and financiers cannot exceed a multiple of their net worth,

$$k_{R,b} \leq \bar{\lambda}_b w_b \qquad\qquad a_f \leq \bar{\lambda}_f w_f.$$

To the extent that leverage constraints can be chosen contingent on the state of the economy and are freely enforceable, it is clear that a social planner can enforce any upper bound on market quantities by setting the appropriate capital requirements. Instead of characterizing optimal leverage requirements in such settings, I focus on another extreme: leverage requirements must be set once and for all, and financier leverage constraints cannot be meaningfully enforced. The latter concern arises may arise because financiers are amorphous institutions that may be relatively hard to regulate effectively, such as those operating in the shadow banking sector. I therefore set $\bar{\lambda}_f = \infty$, and study the implications on binding capital requirements on bankers. For simplicity, I focus on regions of the state space where secondary market liquidity is high in the absence of capital requirements.

Define *bank leverage* to be $\lambda_b = \frac{k_{R,b}}{w_b}$. Bank leverage in the absence of secondary markets and capital requirements is

$$\lambda_b^0 = \frac{k_{R,b}^0}{w_b} = \frac{1}{1 - Q_b^0 y_R(l)\tilde{m}}.$$

Bank leverage with highly liquid secondary markets and without capital requirements is

$$\lambda_b^* = \frac{k_{R,b}^*}{w_b} = \frac{1}{1 - Q_b^* Q_a^* \tilde{m}}.$$

For leverage requirements to influence equilibrium outcomes without shutting down secondary markets altogether, we must therefore have

$$\lambda_b^0 < \bar{\lambda}_b < \lambda_b^*.$$

I maintain this assumption going forward.

I now turn to the banker's problem in the presence of leverage constraint. The key observation is that, since banks are prohibited from levering as much as they would like, they sell just enough risky claims to exactly hit the leverage requirement. Recall from Section 2.4.3 that the collateral short-fall of the banker for a given quantity of bonds issued b_b was defined as

$$\bar{a}_b(k_{R,b}, b_b, Q_a) = \frac{b_b}{Q_a}.$$

Risky claim sales relax borrowing constraints by covering this shortfall. In the absence of capital requirements, the banker's optimal portfolio is such that they sell off exactly the amount of risky claims that maximizes their borrowing capacity: $a_b^* = \bar{a}_b$. When capital requirements bind, however, bankers are no longer permitted to exhaust their entire borrowing capacity. Because risky claims trade below par, bankers therefore withdraw assets from secondary markets until the capital requirement just binds. I summarize the degree to which bankers exhaust their borrowing capacity by $\gamma_b \in [0,1]$. The secondary market supply of bankers can then be written as

$$a_b = \gamma_b \left(\frac{b_b}{Q_a} \right).$$

Since the borrowing and budget constraints continue to bind, the optimal portfolio for given prices is

$$k_{R,b} = \left[\frac{(1-\gamma_b)Q_a + \gamma_b y_R(l)}{(1-\gamma_b)Q_a + \gamma_b y_R(l) - Q_a Q_b y_R(l)\tilde{m}} \right] w_b \qquad \text{and} \qquad b_b = \left[\frac{Q_a y_R(l)\tilde{m}}{(1-\gamma_b)Q_a + \gamma_b y_R(l) - Q_a Q_b y_R(l)\tilde{m}} \right] w_b.$$

Accordingly, bank leverage for a given γ_b is

$$\lambda_b(\gamma) = \frac{(1-\gamma_b)Q_a + \gamma_b y_R(l)}{(1-\gamma_b)Q_a + \gamma_b y_R(l) - Q_a Q_b y_R(l)\tilde{m}}.$$

Setting $\lambda_b(\gamma) = \bar{\lambda}_b$ reveals that the degree to which the banker exhausts his borrowing capacity under capital requirements is

$$\gamma_b^*(\bar{\lambda}_b) = \left(\frac{Q_a}{Q_a - y_R(l)}\right)\left[1 - \left(\frac{\bar{\lambda}_b}{\bar{\lambda}_b - 1}\right) Q_b y_R(l)\tilde{m}\right] \in (0,1).$$

All else equal, the supply of risky claims is thus increasing in the capital requirement $\bar{\lambda}_b$, but *decreasing* in the secondary market price Q_a. The reason is that, by virtue of the fixed capital requirement, bankers aim to fill a fixed revenue target on secondary markets. As a result, supply curves are downward sloping. Most importantly, binding capital requirements put upward pressure on the secondary market price by reducing the *supply* of assets on secondary markets. As the next proposition shows, this price pressure may be sufficiently large that a fraction of bankers *must* start shirking.

Proposition 11 (Shirking Due to Capital Requirements)**.**
Assume that the financial system is highly constrained whether or not capital requirements are binding. Suppose that the competitive equilibrium without leverage constraints is full-monitoring with highly liquid secondary markets. Let λ_b^* denote the associated bank leverage. If $\bar{\lambda}_b < \lambda_b^*$, then a strictly positive fraction of bankers must shirk in the equilibrium with capital requirements.

Proof. See Appendix. □

The next corollary then follows immediately.

Corollary 8 (Equilibrium with Binding Capital Requirements)**.**
If the financial system is highly constrained in the absence of capital requirements and the capital requirement is binding, the competitive equilibrium with capital requirements is a shirking equilibrium. The equilibrium secondary market price Q_a^* satisfies

$$Q_a^* = \hat{y}_R - m \cdot \left[\frac{\bar{L}_b(Q_a^* - y_R(l))}{\bar{\lambda}_b(Q_a^* - m') - (\bar{\lambda}_b - 1)}\right]$$

and is strictly decreasing in $\bar{\lambda}_b$. The equilibrium portfolio of the high-type banker is given by $k_{R,b}^H = \bar{\lambda}_b w_b$, $a_b^H = \frac{w_b(\bar{\lambda}_b(1 - y_R(l)\tilde{m}))}{Q_a^* - y_R(l)}$ and $b_b^H = w_b(\bar{\lambda}_b - 1)$. The equilibrium portfolio of the low-type banker is $k_{R,b}^L = k_{R,b}^H$, $a_b^L = k_{R,b}^L$ and $b_b^L = b_b^H$. The fraction of shirking bankers is Φ is determined by the secondary market clearing condition

$$\Phi k_{R,b}^H + (1-\Phi) a_b^H = \frac{w_f}{Q_a^* - (1-\phi) y_R(l)},$$

where $\phi = \frac{\Phi k_{R,b}^H}{\Phi k_{R,b}^H + (1-\Phi) a_b^H}$ and Φ is decreasing in $\bar{\lambda}_b$.

In Figure 9, 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 secondary market sales by the high-type banker decrease as the capital requirement is relaxed. The reason is that bankers cannot lever beyond a fixed multiple of their net worth, and thus only issue sufficiently many risky claims to reach the target leverage. Accordingly, the next two panels show that both bond issuances and the level of

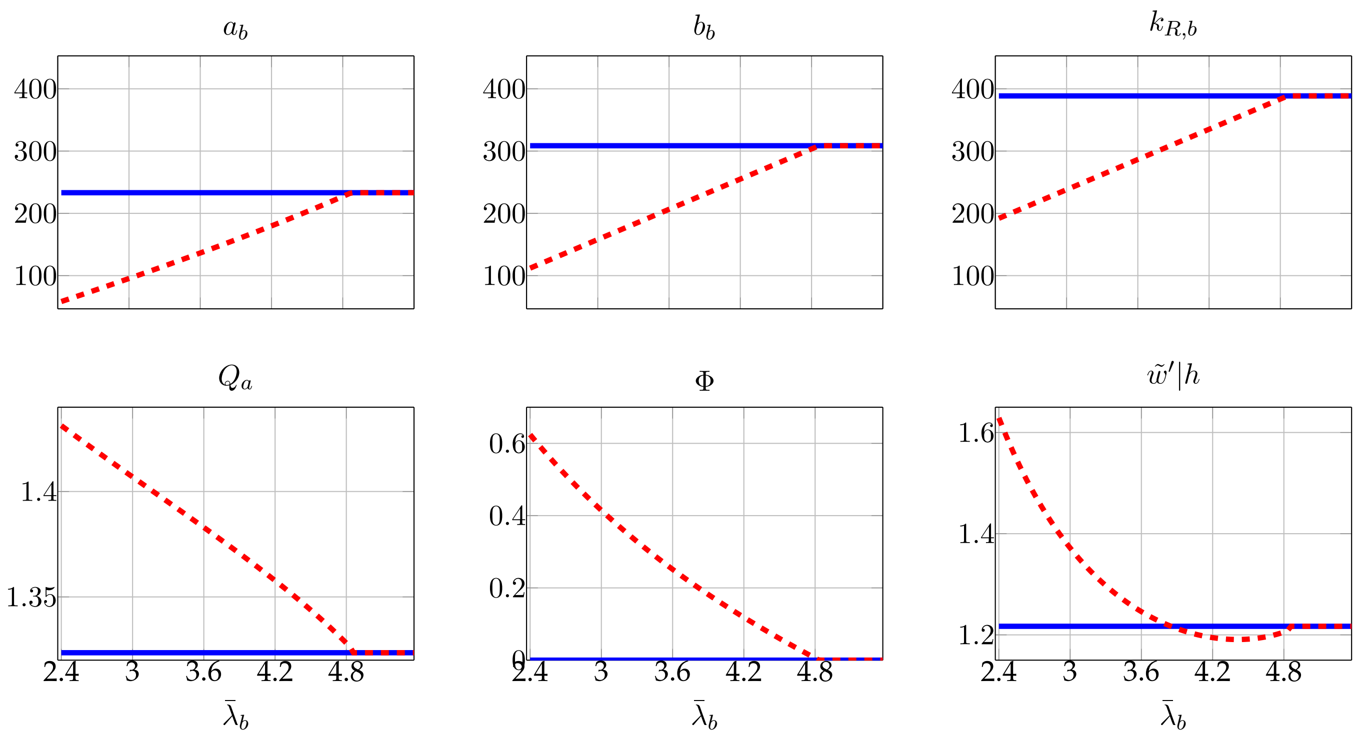

Figure 9: Equilibrium Outcomes as a function of the bank capital requirement $\bar{L}_b$. The unconstrained equilibrium is depicted in solid blue, the equilibrium with capital requirements in dashed red. Leverage in the unconstrained equilibrium is equal to $\lambda_b^* = 4.85$. Parameter values: $\pi_h = 0.5$, $y_R(l) = 0.5$, $\hat{y}_R = 2.65$, $\underline{y}_S = 1$, $m = 0.1$. Wealth distribution: $w_d = 1000$, $w_b = 80$, $w_f = 190$.

investment increase as $\bar{\lambda}_b$ increases. The bottom row plots the secondary market price Q_a, the fraction of shirking bankers Φ, and tomorrow's relative financier net worth after a good shock. As per the proposition, secondary market prices as well as the fraction of shirking bankers decrease in $\bar{\lambda}_b$. The intuition is simple: tighter leverage constraints push banks to withdraw assets from secondary markets, leading to excess demand for risky claims. If only the price were to adjust to clear the market, all bankers would have an incentive to shirk and sell. To satisfy the excess demand without inducing all bankers to shirk, the price increases slightly as the upper bound $\bar{Q}_a$ grows, and more bankers .The bottom-right panel plots the evolution of relative financier net worth $\tilde{w}$ 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 high-type bankers to withdraw assets from secondary markets and *decreases* the amount of risk transferred. Second, an increase in the fraction of shirking bankers leads to an *increase* in risk transfer because low-type banker sell more assets on secondary markets than high-type bankers. When capital requirements are not too tight, the first effect dominates and relative financier net worth grows more slowly in the constrained equilibrium. When capital requirements are tight, the second effect dominates and relative financier net worth grows *faster* in the constrained equilibrium. The simulations reveal a static and a dynamic channel through which capital requirements 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 bankers, inducing further falls in investment quality. It also stands to reason that the first channel is particularly strong when capital requirements are counter-cyclical – secondary market demand is particularly high at the peak of a boom – while the second channel is particularly strong when capital requirements are risk-weighted – if selling assets for cash relieves leverage constraints, then transferring risk on secondary markets is particularly attractive to bankers. 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 bankers used secondary markets to bypass capital requirements. It is also clear that capital requirements on *financiers* may be a useful policy tool to lean against some of the adverse effects of capital requirements on bankers. My results thus highlight why bank capital requirements may be harmful when set independently of financier regulation.

4.3 Post-crisis Interventions and Macro-prudential Regulation

I now summarize the model's implications for post-crisis interventions and macro-prudential regulation. Begin by studying how to best kick-start lending in the aftermath of a financial crisis event. To this end, suppose that the financial system is highly constrained. Corollary 4 showed that increases in financier net worth lead to *more* lending than an equivalent increase in banker net worth when $\underline{y}_S < \bar{y}_S$. Providing equity to financiers may therefore be a more cost-effective way to boost lending. On the downside, Corollary 7 shows that increased financier net worth may set the economy on the path towards a destabilizing secondary market boom.

This suggests a role for macro-prudential policy in regulating the dynamics of secondary market booms more generally. Indeed, Corollary 6 shows that excessively large financier net worth may lead to Pareto-inferior outcomes within a period. A simple policy that eliminates this static inefficiency is to place a cap on the total amount of capital financiers can spend in secondary markets in a given period. To this end, let the regulator choose a $\tilde{w}_f$ such that financiers must invest at least $w_f - \tilde{w}_f$ in the safe technology. This cap on secondary market investment $\tilde{w}_f$ can be chosen such that $\Phi = 0$ whenever the competitive equilibrium in the absence of regulation is a shirking equilibrium in which reductions in financier net worth are Pareto-improving in accordance with Corollary 6. As a result, the policy eliminates within-period inefficiencies. Yet it may also have dynamic benefits. Indeed, it is easy to see that *aggregate* net worth $w_f + w_b$ is larger in any state of the world under this policy. The reason is that the policy eliminates shirking on the equilibrium path, while the conditions in Corollary 6 ensure that total investment is independent of Φ. Aggregate net worth is generally not a sufficient statistic for welfare or total investment. Figure 10 presents an example in which the policy leads to strictly higher investment and, by extension, output in every period.

This suggests that constraints on the asset side of financier balance sheet are a useful macro-prudential policy tool. Three aspects of such a policy are of note. First, financier net worth is harmful only when it is large. Because financiers grow during expansions, the policy is procyclical. 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

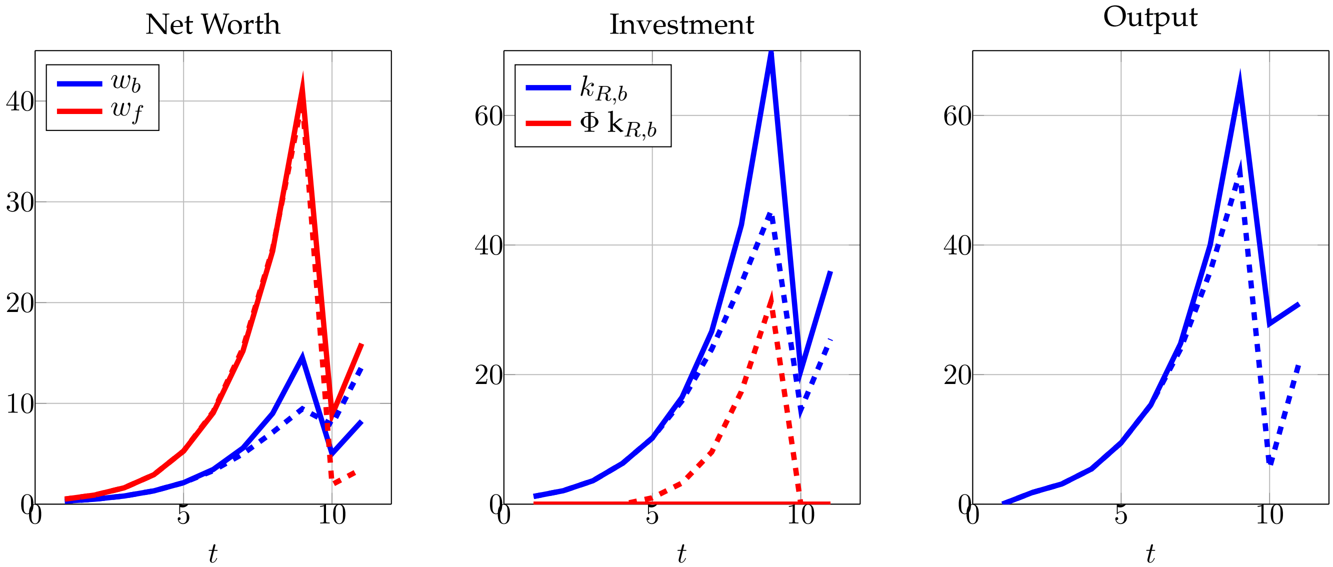

Figure 10: Equilibrium Outcomes when financiers must invest at least $w_f - \tilde{w}_f$ in the safe technology. $\tilde{w}_f$ chosen such that $\Phi = 0$ in every period in which reductions in financier net worth are Pareto-improving as in Corollary 6. Solid lines depict the equilibrium with the policy and dashed lines the equilibrium in the absence of the policy.

relevant concern. The last two points contrast with a regulatory discussion at the Financial Stability Board, which focused on the designating individual asset managers as systemically important because they were fearful of sudden withdrawals from such institutions. In this sense, my results suggest a novel motive for financial regulation.

5 Conclusion

This paper offers a theory of the macroeconomic effects of secondary markets. 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 net worth in the financial system. I distinguish between “bankers” – 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 bankers. Dynamically, the risk transfer that allows credit volumes to expand when financiers are not too large causes financier net worth to grow disproportionately after a sequence of good shocks. Endogenous secondary market credit booms arise that gradually lead to 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. Regarding policy, I show that asymmetric capital requirements on bankers are harmful, and that there is a strong motive for pro-cyclical restrictions on financier's purchases of asset-backed securities.

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 (2009): “Money, Liquidity, and Monetary Policy,” *American Economic Review: Papers & Proceedings*, 99(2), 600–605.

ADRIAN, T., AND H. S. SHIN (2010a): “Financial Intermediaries and Monetary Economics,” in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3. Elsevier.

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

ADRIAN, T., AND H. S. SHIN (2014): “Procyclical Leverage and Value-at-Risk,” *Review of Financial Studies*, 27(2), 373–403.

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.

BERNANKE, B. (2005): “The Global Saving Glut and the U.S. Current Account Deficit,” Remarks by Ben S. Bernanke at the Sandridge Lecture.

BERNANKE, B., AND M. GERTLER (1989): “Agency Costs, Net Worth, and Business Fluctuations,” *American Economic Review*, 79(1), 14–31.

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*.

——— (2015): “Endogenous Liquidity and the Business Cycle,” *American Economic Review*, 105(6), 1883–1927.

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

BRUNNERMEIER, M. K., AND Y. SANNIKOV (2014): “A Macroeconomic Model with a Financial Sector,” *American Economic Review*, 104(2), 379–421.

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*.

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

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

GERTLER, M., AND P. KARADI (2011): "A Model of Unconventional Monetary Policy," *Journal of Monetary Economics*, 58, 17–34.

GERTLER, M., AND N. KIYOTAKI (2010): "Financial Intermediation and Credit Policy in Business Cycle Analysis," in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3, pp. 547–599. Elsevier.

GORTON, G., S. LEWELLEN, AND A. METRICK (2012): "The Safe-Asset Share," *American Economic Review, Papers and Proceedings*, 102(3), 101–106.

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 (2013): "The Supply and Demand for Safe Assets," *NBER Working Paper 18732*.

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

GORTON, G., AND G. G. PENNACCHI (1990): "Financial Intermediaries and Liquidity Creation," *The Journal of Finance*, 45(1), 49–71.

GORTON, G., AND G. G. PENNACCHI (1995): "Banks and Loan Sales: Marketing Nonmarketable Assets," *Journal of Monetary Economics*, 35(3), 389–411.

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*.

GUERRIERI, V., R. SHIMER, AND R. WRIGHT (2010): "Adverse Selection in Competitive Search Equilibrium," *Econometrica*, 78(6), 1823–1862.

HE, Z., AND A. KRISHNAMURTHY (2011): "A Model of Capital and Crises," *Review of Economic Studies*, 79(2), 735–777.

HEBERT, B. (2015): "Moral Hazard and the Optimality of Debt," *Working Paper*.

HOLMSTROM, B., AND J. TIROLE (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112(3), 663–691.

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.

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.

KIYOTAKI, N., AND J. MOORE (1997): "Credit Cycles," *The Journal of Political Economy*, 105(2), 211–248.

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.

——— (forthcoming): "Asset Markets with Heterogenous Information," *Econometrica*.

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

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.

ORDOÑEZ, G. (2013): "Fragility of Reputation and Clustering of Risk-Taking," *Theoretical Economics*, 8(3), 653–700.

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.

PLANTIN, G. (2015): "Shadow Banking and Bank Capital Regulation," *Review of Financial Studies*, 28(1), 146–175.

RAMPINI, A. A., AND S. VISWANATHAN (2010): "Collateral, Risk Management, and the Distribution of Debt Capacity," *The Journal of Finance*, 65(6), 2293–2322.

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.

STEIN, J. C. (2012): "Monetary Policy as Financial-Stability Regulation," *Quarterly Journal of Economics*, 127, 57–95, forthcoming in the Quarterly Journal of Economics Draft Date: Revised May 2011.

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

A Proofs

Proof of Lemma 2

Fix the financier's outstanding debt b_f and assets $k = w_f + Q_b b_f$. By risk-neutrality, the solution must be bang-bang: the financier invests all his assets in either risky claims or the safe technology. By investing k in secondary markets at price $Q_a(\mu^*)$, the financier obtains an expected profit of $v_f^R = \pi_h \left(\frac{y_R(h)}{Q_a(\mu^*)} k - b_b \right) + (1-\pi_h) \max\left\{ \frac{y_R(l)}{Q_a(\mu^*)} k - b_f, 0, \right\}$. Hence $v_f^R \geq \underline{v}_f^R \equiv \pi_h \left(\frac{y_R(h)}{Q_a(\mu^*)} k - b_f \right) + (1-\pi_h) \left(\frac{y_R(l)}{Q_a(\mu^*)} k - b_f \right)$. If the financiers invests k units in the safe technology instead, he receives an expected profit of $v_f^S = \bar{y}_S k - b_b$. It follows from the definitions of k and $\hat{R}^{\text{unlev}}(\mu^*)$ that $\underline{v}_f^R = \hat{R}^{\text{unlev}}(\mu^*) k - b_f$, while $v_f^R > \underline{v}_f^R$ if and only if $b_f > \frac{y_R(l) w_f}{Q_a(\mu^*) - Q_b y_R(l)}$. Accordingly, $v_R^f > v_S^f$ if at least one of the stated conditions is not satisfied. □

Proof of Lemma 3

Fix the banker's outstanding debt b_b and assets $k = w_b + Q_b b_b$. By risk-neutrality, the solution must be bang-bang: the banker invests all his assets in either the risky technology or the safe technology. By investing k in the risky technology, the banker obtains an expected profit of $v_b^R = \pi_h (y_R(h)k - b_b) + (1-\pi_h) \max\{y_R(l)k - b_b, 0\} \geq \pi_h (y_R(h)k - b_b) + (1-\pi_h)(y_R(l)k - b_b) \equiv \underline{v}_b^R$. By investing k in the safe technology instead, the banker receives an expected profit of $v_b^S = \bar{y}_S k - b_b$. Since $\mathbb{E}_z y_R(z) > \bar{y}_S$, $\underline{v}_b^R > v_b^S$ for all k and b_b. □

Proof of Proposition 2

Consider the following candidate equilibrium. Bankers and depositors choose bond and investment quantities as in there were no secondary markets. Every banker sets $\underline{a}_b = a_b = 0$. Every financier bids $a_f(\mu) = 0$ and $Q_a(\mu) = 0$ for all μ and sets $k_{S,f} = w_f$. We want to show that this is an equilibrium. Specifically, we want to show that there no profitable deviations that lead to positive trade on secondary markets. Since $Q_a(\mu) < \underline{Q}_a(Q_b) \leq \hat{y}_R$ for all μ and all $Q_b \in [\frac{1}{\hat{y}_R}, 1]$, no banker has an incentive to sell risky claims at the given prices. Since $\underline{a}_b = 0$ for all bankers, a financier can induce a banker to sell assets under limited commitment only by offering $Q_a = \hat{y}_R$. But doing so yields no greater return than investing in the safe technology. □

Proof of Proposition 5

Begin with the first part of the proposition, and fix a full-monitoring equilibrium with high secondary market liquidity. Assume first that financiers are fully levered ($\gamma = 1$). The optimal banker portfolio satisfies $b_b = Q_a a_b$, while the secondary market clearing condition is $a_b = a_f$. Since the financier portfolio satisfies $b_f = y_R(l) a_f$, the funding market clearing condition in an interior equilibrium can be written as. $Q_a a_f + y_R(l) a_f = \frac{w_d}{Q_b}$ Rearranging gives the bond price as $Q_b(Q_a) = \min\left(\frac{Q_a w_d}{(Q_a + y_R(l)) w_f + y_R(l) w_d}, \frac{1}{\underline{y}_S} \right)$. The secondary market clearing condition in turn gives the secondary market price as $Q_a = \frac{w_f + w_b \tilde{m} y_R(l) Q_b}{\tilde{m} w_b + \tilde{m} Q_b w_f}$. Assume first that $Q_b = \frac{1}{\underline{y}_S}$. Then $Q_a^* = \frac{w_f + w_b \tilde{m} y_R(l) \frac{1}{\underline{y}_S}}{\tilde{m} w_b + \tilde{m} \frac{1}{\underline{y}_S} w_f}$. Differentiating yields that Q_a^* is increasing in w_f if and only if $w_b > \tilde{m} w_b \frac{y_R(l)}{\underline{y}_S}$, which always holds because $\tilde{m} \in (0,1)$ and $y_R(l) < \underline{y}_S$. Moreover, $k_{R,b} = \frac{w_b}{1 - \tilde{m} Q_b Q_a}$ is increasing because Q_a is increasing and Q_b is a constant. Since all bankers monitor, expected output also increases. Now assume that $Q_b < \frac{1}{\underline{y}_S}$. Solving the system of two unknowns generated by the market clearing conditions gives the secondary market price as

$$Q_a^* = \frac{w_f - \tilde{m} y_R(l) w_b + \sqrt{(w_f - \tilde{m} y_R(l) w_b)^2 + 4\tilde{m}(w_b + w_d) y_R(l)(w_f + w_d)}}{2\tilde{m}(w_b + w_d)}$$

which is clearly increasing in w_f. Next, we need to show that $k_{R,b}$ is strictly increasing in w_f. Given the optimal banker portfolio, $k_{R,b} = \frac{1}{\tilde{m}} a_b$. By market clearing, $k_{R,b} = \frac{1}{\tilde{m}} a_f$. Since a_f is strictly increasing in w_f, the result follows. Moreover, expected output is increasing because all bankers monitor. Next, assume that financiers do not borrow ($\gamma = 0$). Then the market clearing conditions yield $Q_b^* = \min\left(\frac{w_d}{Q_a \tilde{m}(w_b+w_d)}, \frac{1}{\underline{y}_S}\right)$ and $Q_a^* = \frac{w_f}{\tilde{m}(w_b+Q_b w_f)}$. If $Q_b^* = \frac{1}{\underline{y}_S}$, then Q_a is clearly increasing in w_f. If $Q_b < \frac{1}{\underline{y}_S}$, then $Q_a^* = \frac{w_f}{\tilde{m}(w_d+w_b)}$ which is again increasing in w_f. Next, note that $Q_b^* Q_a^* = \frac{w_d}{\tilde{m}(w_b+w_d)}$. Hence $k_{R,b} = \frac{w_b}{1-Q_b Q_a \tilde{m}}$ is non-decreasing in w_f. Next, assume that financiers are indifferent between borrowing and lending ($\gamma \in (0,1)$). Then by definition, $Q_a^* = \hat{y}_R Q_b^*$, and $b_f = \gamma y_R(l) a_f$. The secondary market clearing condition is $\frac{\tilde{m} w_b}{1-Q_a^2 \frac{\tilde{m}}{\hat{y}_R}} = \frac{w_f}{Q_a(1-\frac{\gamma y_R(l)}{\hat{y}_R})}$. Suppose for a contradiction that Q_a is decreasing in w_f. Then $a_b = \frac{\tilde{m} w_b}{1-Q_a^2 \frac{\tilde{m}}{\hat{y}_R}}$ is also decreasing in w_f. To maintain market clearing, a_f must be decreasing in w_f, and hence γ, $b_b = Q_a a_b$ and $b_f = \gamma y_R(l) a_f$ must also be decreasing. But if b_b and b_f are decreasing in w_f, then Q_b must be *increasing* in w_f. This is a contradiction with the fact that Q_b must decreasing because $Q_a^* = \hat{y}_R Q_b^*$ and Q_a was presumed to be decreasing. It then follows that $k_{R,b} = \frac{w_b}{1-Q_a^2 \frac{\tilde{m}}{\hat{y}_R}}$ is increasing in w_f. Because all bankers monitor, expected output is increasing in w_f also.

Now turn to the second part of the proposition, and fix a full-monitoring equilibrium with low liquidity. By definition, $Q_a^* = \underline{Q}_a(Q_b) = \frac{\hat{y}_R - y_R(l) + \hat{y}_R Q_b m'}{Q_b(\hat{y}_R - y_R(l) + m')}$ so that Q_a^* is *decreasing* in Q_b. To show that Q_a is increasing in w_f is therefore to show that Q_b is decreasing in w_f. Note first that the proposition is trivial when financiers are indifferent toward leverage. In that case, $Q_a = \hat{y}_R Q_b$ and so both prices are constants. Moreover, $k_{R,b}$ is strictly increasing because $a_b = a_f$ is increasing. If the financial system is highly constrained, then $Q_b^* = \frac{1}{\underline{y}_S}$. Hence Q_a and Q_b are constants, and $k_{R,b}$ is strictly increasing in w_f. It remains to be shown that Q_a increasing in w_f if the financial system is not highly constrained and financiers are either fully levered or do not borrow. To this end, recall financier demand is $a_f = \frac{w_f}{\underline{Q}_a(Q_b) - \gamma y_R(l) Q_b}$. Suppose for a contradiction that Q_b is increasing in w_f. We first show that a_f must strictly increase. Suppose for a contradiction that a_f weakly decreases. Then Q_b must strictly decrease given that w_f increased. But if a_f is weakly decreasing, so is b_f. Similarly, strictly lower Q_b and weakly lower a_f imply that b_b is strictly smaller. But if b_f and b_b both weakly decrease, then Q_b cannot fall, yielding a contradiction. Hence a_f is strictly increasing in w_f. But b_b and b_f are both strictly increasing in a_f. Hence Q_b must fall, and Q_a must increase.

Next, consider shirking equilibria. Given the optimal portfolios of bankers, it is straightforward to show that $Q_a^* = \bar{Q}_a = \hat{y}_R'$ when liquidity is high. The secondary market price is a constant. If the financial system is highly constrained, then Q_b is fixed at $\frac{1}{\underline{y}_S}$, and prices are constants. As a result, $k_{R,b}$ is a constant. To clear secondary markets at fixed prices, Φ must be increasing in w_f. Given that $k_{R,b}$ is constant, expected output must be declining. If the financial system is not highly constrained, then the market clearing conditions are:

$$\frac{\hat{y}_R' \tilde{m} W_b}{1 - Q_b \hat{y}_R' \tilde{m}} + \frac{\gamma(1-\phi) y_R(l) W_f}{\hat{y}_R' - (1-\phi) Q_b \gamma y_R(l)} = \frac{W_d}{Q_b} \qquad \text{and} \qquad \left(\frac{\Phi + (1-\Phi)\tilde{m}}{1 - Q_b \hat{y}_R' \tilde{m}}\right) W_b = \frac{W_f}{\hat{y}_R' - (1-\phi) Q_b \gamma y_R(l)}.$$

Suppose first that financiers are indifferent toward leverage ($\gamma \in (0,1)$). Then $Q_b^* = \frac{Q_a^*}{\hat{y}_R}$ is a constant. It follows immediately that Φ is increasing in w_f, while $k_{R,b}$ is independent of w_f. Hence expected output must decline. Next, suppose financiers do not borrow ($\gamma = 0$). In this case, the bond market clearing condition is independent of w_f, and so Q_b is a constant. Hence $k_{R,b}$ is a constant, and expected output must decline. Finally, assume that financiers are fully levered ($\gamma = 1$). Imposing bond market clearing reveals that Q_b must satisfy

$$Q_{b,PM}^*(\Phi, w_f) = \frac{w_d}{\tilde{m}\left[\hat{y}_R'(w_b + w_d) + (1-\Phi) y_R(l) w_f\right]}.$$

Note that $Q_{b,PM}^*(\Phi, w_f)$ is strictly increasing in Φ and strictly decreasing in w_f. Similarly, the bond price that clears the secondary market is

$$Q_{b,SM}^*(\Phi) = \frac{w_f - w_b \hat{y}_R'(\Phi + (1-\Phi)\tilde{m})}{\tilde{m}\left[(w_b + w_f) y_R(l) + w_b y_R(l) \Phi\right]}.$$

Note that $Q_{b,SM}^*(\Phi, w_f)$ 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_b. But by secondary market clearing, an increase in w_f and a decrease in Φ leads to an *increase* in Q_b. Hence both markets do not clear simultaneously, leading to a contradiction. The result then follows.

Next consider a low-liquidity equilibrium. I will first show that whenever the financial system is highly constrained, or financiers weakly prefer to not borrow, then Q_b^*, Q_a^*, $k_{R,b}^H$ and a_b^H are all invariant to w_f. Suppose first that the financial system is highly constrained, so that $Q_b^* = \frac{1}{\underline{y}_S}$. Then $Q_a^* = \underline{Q}_a(Q_b^*)$ is a constant, and thus invariant to w_f. Since $k_{R,b}^H = \frac{w_b + Q_b^*(Q_a^* - y_R(l))a_b^H}{1 - y_R(l)Q_b^*\tilde{m}}$ in any shirking equilibrium, $k_{R,b}^*$ and $Q_a^* = \bar{Q}_a = \hat{y}_R - m\frac{k_{R,b}^H}{k_{R,b}^H - a_b^H}$ in any shirking equilibrium, $k_{R,b}^H$ and a_b^H 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 $b_b^H = \frac{y_R(l)\tilde{m}w_b + (Q_a - y_R(l))a_b^H}{1 - y_R(l)Q_b^*\tilde{m}} = \frac{w_d}{Q_b}$. Hence a_b^H is fixed conditional on Q_b. Since $Q_a^* = \bar{Q}_a = \hat{y}_R - m\frac{k_{R,b}^H}{k_{R,b}^H - a_b^H}$ and $k_{R,b}$ is fixed once Q_b is determined, the condition $Q_a = \underline{Q}_a(Q_b)$ suffices to pin down Q_b, Q_a, $k_{R,b}^H$ and a_b^H 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 $k_{R,b}$, 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:

$$\frac{y_R(l)\tilde{m}w_b + (\underline{Q}_a - y_R(l))a_b^H}{1 - Q_b y_R(l)\tilde{m}} + \frac{y_R(l)w_f}{\underline{Q}_a - (1-\phi)Q_b y_R(l)} = \frac{w_d}{Q_b}$$

and

$$\Phi\left(\frac{w_b + Q_b(\underline{Q}_a - y_R(l))a_b^H}{1 - Q_b y_R(l)\tilde{m}}\right) + (1-\Phi)a_b^H = \frac{w_f}{\underline{Q}_a - (1-\phi)Q_b y_R(l)},$$

The secondary market price must satisfy $Q_a^* = \underline{Q}_a(Q_b)$, and is thus fixed for given Q_b. The high-type banker's investment is given by:

$$k_{R,b}^H = \frac{w_b}{1 - Q_b y_R(l)\tilde{m}} + \left(\frac{\hat{y}_R - y_R(l)}{\hat{y}_R - y_R(l)\tilde{m}}\right)a_b^H$$

Hence, $k_{R,b}^H$ is a function of a_b^H and Q_b only. Given that we are in a shirking equilibrium, $Q_a = \bar{Q}_a(k_{R,b}^H, a_b^H)$, and so Q_a and Q_b are fixed for a given a_b^H. It is easy to verify that $\bar{Q}_a$ is strictly decreasing in a_b^H. 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_b and/or a_b^H must decrease. Since $Q_a = \underline{Q}_a$, if Q_b falls, then Q_a must increase. Since $Q_a = \bar{Q}_a$ and $\bar{Q}_a$ is strictly decreasing in a_b^H, it follows that Q_b and a_b^H must both decrease. Since b_b and Q_b are decreasing, it follows that $k_{R,b}$ must decrease. Since $k_{R,b}$, a_b^H, 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.

Proof of Corollary 4

Inspecting the optimal intermediary portfolios, it follows that aggregate investment in a full-monitoring low-liquidity equilibrium is proportional to $W_b - \Delta + \frac{Q_b^*\left[\underline{Q}_a(Q_b^*) - y_R(l)\right][W_f + \Delta]}{\underline{Q}_a(Q_b^*) - Q_b^* y_R(l)}$. The coefficient on Δ is equal to zero when $Q_b^* = 1$ and strictly positive when $Q_b^* > 1$. When the financial sector is highly constrained, then $Q_b^* = \frac{1}{\underline{y}_S}$ independent of Δ for Δ sufficiently small. $\square$

Proof of Proposition 6

Fix a full-monitoring low-liquidity equilibrium with a highly constrained financial system. The expected return on equity earned by bankers and financiers, respectively, is

$$\hat{\text{ROE}}_b = \frac{\hat{y}_R - y_R(l)\tilde{m}}{1 - Q_b^* y_R(l)\tilde{m}} \qquad \text{and} \qquad \hat{\text{ROE}}_f = \frac{\hat{y}_R - y_R(l)}{\underline{Q}_a(Q_b^*) - Q_b y_R(l)}$$

Moreover, $\hat{\text{ROE}}_f > \hat{\text{ROE}}_b$ if $Q_b^* > 1$ and $\text{ROE}_f \leq \text{ROE}_b$ if $Q_b^* \leq 1$.

Since liquidity is low, the secondary market price is given by $Q_a^* = \underline{Q}_a(Q_b^*)$. Moreover, bankers receive no rents from secondary market trading by definition. Bankers' expected return on equity is therefore equal to bankers' expected return on equity in an equilibrium without secondary markets. This gives the first result:

$$\hat{\text{ROE}}_b = \frac{\hat{y}_R - y_R(l)\tilde{m}}{1 - Q_b^* y_R(l)\tilde{m}}$$

Next, turn to financiers' return on equity. Since financiers are fully levered, the expected utility of financiers is

$$v_f = \hat{y}_R a_f - b_f = \frac{\hat{y}_R - y_R(l)}{\underline{Q}_a(Q_b^*) - Q_b^* y_R(l)} w_f$$

Hence $\hat{\text{ROE}}_f = \frac{\hat{y}_R - y_R(l)}{\underline{Q}_a(Q_b^*) - Q_b y_R(l)}$. To show the remaining results, I begin by showing that $\hat{\text{ROE}}_f = \hat{\text{ROE}}_b$ if $Q_b = 1$. To this, write the expected returns on equity of both intermediaries at $Q_b = 1$ and for a generic Q_a as:

$$\hat{\text{ROE}}_b = \frac{\hat{y}_R - y_R(l)\tilde{m}}{1 - y_R(l)\tilde{m}} \qquad \text{and} \qquad \hat{\text{ROE}}_f = \frac{\hat{y}_R - y_R(l)}{Q_a - y_R(l)}$$

Algebra reveals that $\hat{\text{ROE}}_b = \hat{\text{ROE}}_f$ at $Q_b = 1$ if and only if

$$Q_a = \frac{\hat{y}_R - y_R(l) + (1 - \tilde{m}) y_R(l) \hat{y}_R}{\hat{y}_R - y_R(l)\tilde{m}} = \underline{Q}_a(1).$$

To show that $\hat{\text{ROE}}_f > \hat{\text{ROE}}_b$ when $Q_b > 1$ and $\hat{\text{ROE}}_f \leq \hat{\text{ROE}}_b$ otherwise, it then suffices to show that $\frac{\partial \hat{\text{ROE}}_f}{\partial Q_b} > \frac{\partial \hat{\text{ROE}}_b}{\partial Q_b}$. To this end, note first that $\hat{\text{ROE}}_f = (\hat{y}_R - y_R(l))\lambda_f$ and $\hat{\text{ROE}}_f = (\hat{y}_R - y_R(l)\tilde{m})\lambda_b$, where $\lambda_f = \frac{1}{\underline{Q}_a - Q_b y_R(l)}$ and $\lambda_b = \frac{1}{1 - Q_b y_R(l)\tilde{m}}$ denotes the leverage of financiers and bankers, respectively. Since $\tilde{m} \in (0, 1)$, it follows immediately that $\lambda_f > \lambda_b$ if $Q_b = 1$. That is, financiers have higher leverage than bankers when $Q_b = 1$. Finally, note that

$$\frac{\partial \hat{\text{ROE}}_f}{\partial Q_b} = \left(y_R(l) - \frac{\partial \underline{Q}_a(Q_b)}{\partial Q_b} \right) \lambda_f^2 \qquad \text{and} \qquad \frac{\partial \hat{\text{ROE}}_b}{\partial Q_b} = y_R(l)\tilde{m}\lambda_b^2$$

Since $\frac{\partial \underline{Q}_a(Q_b)}{\partial Q_b} < 0$ and $\tilde{m} \in (0, 1)$, the result follows. □

Proof of Corollary 5

Fix a full-monitoring equilibrium. Note that for any Q_b, Q_a, $a_f^* \geq \frac{w_f}{Q_a}$, while $\bar{Q}_a(k_{R,b}, a_b) \leq \hat{y}_R - m$. Because financiers strictly prefer risky claims to the safe technology at $\bar{Q}_a$, it is sufficient to show that there exists a w_f' such that there is excess demand on secondary markets at price $\hat{y}_R - m$, given that all bankers monitor. This is the case whenever $\frac{w_f'}{\hat{y}_R - m} > \frac{\tilde{m} w_b}{1 - Q_b \tilde{m}(\hat{y}_R - m)}$. Since the RHS is bounded, there always exists a w_f' large enough.

Proof of Corollary 6

Suppose first that secondary market liquidity is *high*. In a high-liquidity shirking equilibrium, $Q_a^* = \hat{y}_R'$, which is a constant. Furthermore, Q_b^* 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 Q_a^* is a constant), or because the financial system is highly constrained so that $Q_b^* = \frac{1}{\underline{y}_S}$ and further reductions in w_f increase excess demand in the bond market. Hence $k_{R,b}^*$ is invariant to w_f. Since Q_b^* is constant, so is depositor utility. By construction, banker utility is $(1 - \tilde{m})k_{R,b}^*$, which is constant. If financiers do not borrow, financier utility is $(\Phi \hat{y}_R' + (1 - \Phi)\tilde{m}\hat{y}_R)k_{R,b}^*$, and is strictly decreasing in w_f.If financiers do borrow, financier utility is $(\Phi \hat{y}_R' + (1 - \Phi)\tilde{m}(\hat{y}_R - y_R(l))k_{R,b}^*$, which is again strictly decreasing in w_f.

Now suppose that secondary market liquidity is *low*. I will first show that whenever the financial system is highly constrained, or financiers do not borrow, then Q_b^*, Q_a^*, $k_{R,b}^H$ and a_b^H are all invariant to reductions in w_f. Suppose first that the financial system is highly constrained, so that $Q_b^* = \frac{1}{\underline{y}_S}$. Then $Q_a^* = \underline{Q}_a(Q_b^*)$ is a constant, and thus invariant to reductions in w_f. Since $k_{R,b}^H = \frac{w_b + Q_b^*(Q_a^* - y_R(l))a_b^H}{1-y_R(l)Q_b^*\tilde{m}}$ in any shirking equilibrium, $k_{R,b}^*$ and $Q_a^* = \bar{Q}_a = \hat{y}_R - m\frac{k_{R,b}^H}{k_{R,b}^H - a_b^H}$ in any shirking equilibrium, $k_{R,b}^H$ and a_b^H 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 $b_b^H = \frac{y_R(l)\tilde{m}w_b + (Q_a - y_R(l))a_b^H}{1-y_R(l)Q_b^*\tilde{m}} = \frac{w_d}{Q_b}$. Hence a_b^H is fixed conditional on Q_b. Since $Q_a^* = \bar{Q}_a = \hat{y}_R - m\frac{k_{R,b}^H}{k_{R,b}^H - a_b^H}$ and $k_{R,b}$ is fixed once Q_b is determined, the condition $Q_a = \underline{Q}_a(Q_b)$ suffices to pin down Q_b, Q_a, $k_{R,b}^H$ and a_b^H independently of w_f. Given these preliminaries, we can now show that reductions in w_f are Pareto-improving. First, note that the utility of depositors is given by $v_d = \frac{w_b}{Q_b^*}$ if $Q_b^* > \frac{1}{\underline{y}_S}$ and $v_d = \underline{y}_S w_d$ otherwise. Since high-type bankers are indifferent towards selling assets on secondary markets in a low-liquidity equilibrium, their utility is unchanged by the presence of secondary markets: $v_b^H = \frac{\hat{y}_R - \tilde{m}y_R(l)}{1-Q_b^* y_R(l)\tilde{m}}$. By construction, the utility of the low-type banker satisfies $v_b^L = v_b^H = v_b$. Since Q_b^* 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 $v_f = (\phi\hat{y}_R' + (1-\phi)(\hat{y}_R - y_R(l)))a_f^*$. When they do not borrow, it is $v_f = (\phi\hat{y}_R' + (1-\phi)\hat{y}_R)a_f^*$. Market clearing requires that $\Phi a_b^{H*} + (1-\Phi)k_{R,b}^{H*} = a_f = \frac{w_f}{Q_a^* - (1-\phi)y_R(l)Q_b}$, where we have established that Q_b^*, Q_a^*, $k_{R,b}^{H*}$ and a_b^{H*} are all constant. By the definition of ϕ, the utility of financiers then is $v_f = \Phi\hat{y}_R' k_{R,b}^{H*} + (1-\Phi)\hat{y}_R a_b^{H*}$ when they do not borrow, and $v_f = \Phi\hat{y}_R' k_{R,b}^{H*} + (1-\Phi)(\hat{y}_R - y_R(l))a_b^{H*}$ when they do. Since $k_{R,b}^{H*}$ 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 8

Fix an equilibrium with high secondary market liquidity. The bond market clearing condition uniquely determines Q_b as a function of the wealth distribution $\mathbf{w} = (w_d, w_b, w_f)$, the secondary market price Q_a and the fraction of shirking bankers Φ. Hence we can write $Q_b = Q_b^*(\mathbf{w}, Q_a, \Phi)$. The first step is to show that Q_a (in a full-monitoring equilibrium) and Φ (in a shirking equilibrium) are increasing in the wealth shock ξ if and only if

$$\underbrace{\lambda_f \frac{\partial W_f}{\partial \xi} + W_f\left(\frac{\partial \lambda_f}{\partial Q_b}\right)\left(\frac{Q_b^*}{\partial \xi}\right)}_{\text{Change in Financier Secondary Market Demand}} \quad \geq \quad \underbrace{\lambda_b \frac{\partial W_b}{\partial \xi} + W_b\left(\frac{\partial \lambda_b}{\partial Q_b}\right)\left(\frac{Q_b^*}{\partial \xi}\right)}_{\text{Change in Banker Secondary Market Supply}} \tag{5}$$

To this end, define the leverage of bankers and financiers as $\lambda_b(Q_a, Q_b) = \frac{\Phi + (1-\Phi)\tilde{m}}{1-Q_bQ_a\tilde{m}}$ and $\lambda_f(Q_a, Q_b) = \frac{1}{Q_a - \gamma(1-\phi)Q_b y_R(l)}$, respectively. Then the secondary market clearing condition is $\lambda_b(Q_b, Q_a)w_b = \lambda_f(Q_a, Q_b)w_f$. Start by fixing a full-monitoring equilibrium. Then $\Phi = \frac{\partial \Phi}{\partial \xi} = 0$. Totally differentiating the secondary market clearing condition and rearranging yields

$$\frac{\partial Q_a}{\partial \xi} \cdot \underbrace{\left[\frac{\partial \lambda_b}{\partial Q_a}w_b + \frac{\partial \lambda_b}{\partial Q_b}\frac{\partial Q_b^*}{\partial Q_a}w_b - \left(\frac{\partial \lambda_f}{\partial Q_a}w_f + \frac{\partial \lambda_f}{\partial Q_b}\frac{\partial Q_b^*}{\partial Q_a}w_f\right)\right]}_{\equiv A} = \left[\lambda_f w_f + \frac{\partial \lambda_f}{\partial Q_b}\frac{\partial Q_b^*}{\partial \xi}w_f - \left(\lambda_b w_b + \frac{\partial \lambda_b}{\partial Q_b}\frac{\partial Q_b^*}{\partial \xi}w_b\right)\right].$$

where A is the excess supply on secondary markets induced by a marginal increase in Q_a. Since banker supply is increasing in Q_a and financier supply is decreasing in Q_a, $A > 0$ and the result follows. Next, consider a shirking equilibrium. Now $Q_a = \hat{y}_R'$ and so $\frac{\partial Q_a}{\partial \xi} = 0$. Totally differentiating the secondary market clearing condition and rearranging now yields

$$\frac{\partial \Phi}{\partial \xi} \cdot \underbrace{\left[\frac{\partial \lambda_b}{\partial \Phi}w_b + \frac{\partial \lambda_b}{\partial Q_b}\frac{\partial Q_b^*}{\partial \Phi}w_b - \left(\frac{\partial \lambda_f}{\partial \Phi}w_f + \frac{\partial \lambda_f}{\partial Q_b}\frac{\partial Q_b^*}{\partial \Phi}w_f\right)\right]}_{\equiv B} = \left[\lambda_f w_f + \frac{\partial \lambda_f}{\partial Q_b}\frac{\partial Q_b^*}{\partial \xi}w_f - \left(\lambda_b w_b + \frac{\partial \lambda_b}{\partial Q_b}\frac{\partial Q_b^*}{\partial \xi}w_b\right)\right].$$

where B is the excess supply on secondary markets induced by a marginal increase in Φ. Since banker supply is increasing in Φ and financier supply is decreasing in Φ, $B > 0$ and the result follows. The next step is to show that condition (5) is equivalent to the condition stated in the text. To this end, divide (5) through by w_f and impose the market clearing condition $\lambda_b w_b = \lambda_f w_f$. This yields the condition $\lambda_f \frac{\partial \frac{w_f}{w_b}}{\partial \xi} + \frac{\partial \lambda_f}{\partial Q_b} \frac{\partial Q_b}{\partial \xi} \frac{w_f}{w_b} \geq \frac{\partial \lambda_b}{\partial Q_b} \frac{\partial Q_b}{\partial \xi}$. Now suppose that the financier is fully leveraged. Then, by definition, $\frac{\partial \lambda_b}{\partial Q_b} = Q_a \lambda_b^2$, and $\frac{\partial \lambda_f}{\partial Q_b} = y_R(l)\lambda_f^2$. Moreover, secondary market clearing implies that $\lambda_b = \lambda_f \frac{w_f}{w_b}$. Imposing these conditions gives the first part of the corollary. Next, suppose that the financier does not borrow. Then $\lambda_f = \frac{1}{Q_a}$ and thus $\frac{\partial \lambda_f}{\partial Q_b} = 0$. Canceling out Q_a and rearranging gives the second part of the corollary. □

Proof of Proposition 9

By construction. Consider a low-liquidity equilibrium in which the financial system is highly constrained. Such an equilibrium always exists for w_d large enough and w_f small enough. In such an equilibrium, $Q_b^* = \frac{1}{\underline{y}_S}$ and $Q_a^* = \underline{Q}_a(Q_b^*)$. Let $\lambda_f = \frac{1}{Q_a^* - Q_b^* y_R(l)}$ and $\lambda_b = \frac{1}{1 - Q_b^* y_R(l)\tilde{m}}$ denotes the equilibrium leverage of financiers and bankers in a full-monitoring equilibrium, respectively. From the optimal banker portfolio and the definition of $\underline{Q}_a(Q_b)$, it follows that

$$k^*_{R,b} = \lambda_b w_b + \chi a_f^*$$

where $\chi = \left(\frac{\hat{y}_R - y_R(l)}{\hat{y}_R - y_R(l)\tilde{m}}\right) \in (0,1)$ and $a_f^* = \lambda_f w_f$. Hence, the upper bound on the secondary market price stemming from the implementability constraint (IMP) is

$$\bar{Q}_a = \hat{y}_R - \tilde{m}\left(\frac{\lambda_b w_b + \chi \lambda_f w_f}{\lambda_b w_b - (1-\chi)\lambda_f w_f}\right) = \hat{y}_R - \tilde{m}\left(\frac{\lambda_b + \chi \lambda_f \tilde{\mathbf{w}}}{\lambda_b - (1-\chi)\lambda_f \tilde{\mathbf{w}}}\right)$$

Note that $\bar{Q}_a = \hat{y}_R - m$ if $\tilde{\mathbf{w}} = 0$ and that $\bar{Q}_a$ is strictly decreasing in $\tilde{\mathbf{w}}$. It follows that as long as $Q_a^* < \hat{y}_R - m$ there exists, for small enough w_f, a *full-monitoring* low-liquidity equilibrium in which the financial system is highly constrained. This parametric condition is equivalent to

$$\underline{y}_S < \frac{1}{\chi}(\hat{y}_R - m) - \frac{(1-\tilde{m})y_R(l)\hat{y}_R}{\hat{y}_R - y_R(l)} \tag{6}$$

Next, note that $\bar{Q}_a \geq \hat{y}'_R$ because $a_b \leq \tilde{m}k_{R,b}$. For a shirking equilibrium to exist for sufficiently large w_f, we therefore require that $Q_a^* \geq \hat{y}'_R$. This parametric condition is equivalent to

$$\underline{y}_S > \frac{1}{\chi}\hat{y}' - \frac{(1-\tilde{m})y_R(l)\hat{y}_R}{\hat{y}_R - y_R(l)} \tag{7}$$

It is easy to see that there exist parameters such conditions (6) and (7) are jointly satisfied. For example, set $\hat{y}'_R = \underline{y}_S - \epsilon$ for ϵ and m sufficiently small. Hence there exist parameters such that $Q_a \in [\hat{y}'_R, \hat{y}_R - m)$. Assume a set of such parameters from now on, and choose initial financier net worth w_f^0 such that the economy is initially in a full-monitoring equilibrium. We now want to show that the economy may transition into a shirking equilibrium after a sufficiently long sequence of large shocks. Note first that because intermediary net worth is bounded after any finite sequence of good aggregate shocks, there always exists a level of depositor net worth 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 $Q_b^* \geq 1$ because $\underline{y}_S \leq \bar{y}_S = 1$. By Proposition 7, relative financier net worth $\tilde{\mathbf{w}}$ thus grows after a good shock for any $\tilde{\mathbf{w}}$. By the parametric condition (7), a sufficiently long sequence of good aggregate shocks therefore triggers a shirking equilibrium. We then only need to show that there exist parameters such that *expected* return on equity is higher for financiers than for bankers throughout. Recall from Proposition (6) that, in a full-monitoring equilibrium, $\widehat{\text{ROE}}_f > \widehat{\text{ROE}}_b$ for $Q_b > 1$. Hence $\widehat{\text{ROE}}_f > \widehat{\text{ROE}}_b$ in a full-monitoring equilibrium when the financial system is highly constrained if $\underline{y}_S < 1$. Next, turn to a shirking equilibrium. By construction, the return on equity of bankers is independent of the fraction of shirking bankers 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 full-monitoring equilibrium, there exists a Φ^* such that the return on equity is also strictly higher in a shirking equilibrium in which Φ^* bankers shirk. This is the case when $\underline{y}_S < 1$. $\square$

Proof of Proposition 10

Begin with the growth rate of relative financier net worth after a good shock. In the given equilibrium, it is given by:

$$\frac{\tilde{\mathbf{w}}'}{\tilde{\mathbf{w}}} = \frac{\frac{y_R(h)-y_R(l)}{\underline{Q}_a(Q_b^*)-Q_b^* y_R(l)}}{\frac{y_R(h)-\tilde{m}y_R(l)}{1-Q_b\tilde{m}y_R(l)} - \frac{m'(y_R(h)-\hat{y}_R)}{\hat{y}_R-\tilde{m}y_R(l)}\frac{\tilde{\mathbf{w}}}{\underline{Q}_a(Q_b^*)-Q_b^* y_R(l)}}.$$

Let the *relative levered return* $\tilde{LR} \equiv \left[\frac{y_R(h)-y_R(l)]}{y_R(h)-\tilde{m}y_R(l)}\right]\left[\frac{1-Q_b\tilde{m}y_R(l)}{\underline{Q}_a(Q_b)-Q_b y_R(l)}\right]$ denote the ratio of financier and banker levered returns on equity. Let $RT \equiv \frac{m'(y_R(h)-\hat{y}_R)}{\hat{y}_R-\tilde{m}y_R(l)}\frac{\tilde{\mathbf{w}}}{\underline{Q}_a(Q_b^*)-Q_b^* y_R(l)}$ denote the degree of risk transfer from bankers to financiers. RT is increasing in Q_b and thus decreasing in ρ: lower bond prices allow financiers to expand borrowing and purchase more risky claims. Hence, the growth rate of relative financier net worth after a good shock is increasing in risk transfer. For $\frac{\tilde{\mathbf{w}}'}{\tilde{\mathbf{w}}}$ to be decreasing in ρ for all $\tilde{\mathbf{w}} > 0$, we thus require that it to be decreasing in ρ even when risk transfer RT is close to zero (i.e. when $\tilde{\mathbf{w}}$ is close to zero). Hence, we require $\tilde{LR}$ to be increasing in Q_b. Differentiating $\tilde{LR}$ w.r.t to Q_b reveals that $\frac{\partial Q_b \tilde{LR}}{\partial Q_b} \geq 0$ if and only if $\frac{\tilde{m}y_R(l)}{-Q_a'+y_R(l)} \leq \tilde{LR}$ where Q_a' denotes the derivative of $\underline{Q}_a(Q_b)$ w.r.t. Q_b. Since $Q_a' < 0$ and $\tilde{m} < 1$, the LHS is strictly less than unity. It remains to be shown that $\tilde{LR} \geq 1$, i.e. financiers lever more than bankers in a low-liquidity equilibrium in which the financial system is highly constrained. Rearranging $\tilde{LR}$ implies that $\tilde{LR} \geq 1$ if and only if $Q_b y_R(l)(1-\tilde{m}) \geq \underline{Q}_a(Q_b) - 1$. The LHS is strictly increasing in Q_b, while the RHS is strictly decreasing in Q_b. Since $Q_b \geq 1$ when the financial system is highly constrained, the result follows if $y_R(l)(1-\tilde{m}) \geq \underline{Q}_a(1) - 1$. This always holds under the assumption $y_R(l) < \bar{y}_S < 1$.

Next, turn to investment. In a low-liquidity full-monitoring equilibrium with low liquidity, it is given by $K_{R,b} = \left[W_b + \frac{Q_b^*\left[\underline{Q}_a(Q_b^*)-y_R(l)\right]W_f}{\underline{Q}_a(Q_b^*)-Q_b^* y_R(l)}\right][1-Q_b^* y_R(l)\tilde{m}]^{-1}$. Since Q_b^* is decreasing in ρ, a sufficient condition for the desired result is that $\chi_0 \equiv \frac{Q_b\underline{Q}_a(Q_b)-Q_b y_R(l)}{\underline{Q}_a(Q_b)-Q_b y_R(l)}$ is decreasing in Q_b. Differentiating χ_0 with respect to Q_b implies that χ_0 is increasing in Q_b if $(Q_a + Q_b Q_a' - y_R(l))(Q_a - Q_b y_R(l)) > (Q_a' - y_R(l))(Q_b Q_a - Q_b y_R(l))$, where it is understood that $Q_a = \underline{Q}_a(Q_b)$ and Q_a' denotes the derivative w.r.t. Q_b. By definition of $\underline{Q}_a$, it follows that $0 > Q_a + Q_b Q_a' - y_R(l) > Q_a' - y_R(l)$ while $Q_b Q_a - Q_b y_R(l) \geq Q_a - Q_b y_R(l) > 0$ for $Q_b \geq 1$. Since $\underline{y}_S \leq 1$, $Q_b \geq 1$ and the result follows. $\square$

Proof of Proposition 11

Assume for a contradiction that all bankers monitor. Let Q_a^* denote the equilibrium secondary market price. Since all bankers monitor, $Q_a^* < \hat{y}_R$. Since the financial system is highly constrained, $Q_b^* = 1$ with and without leverage constraints.. As a result, financiers are fully levered, and the demand for risky claims is $a_f = \frac{w_f}{Q_a^*-y_R(l)}$. From the optimal banker portfolios, the supply of risky assets is $a_b = \frac{w_b\left(\bar{L}_b(1-y_R(l)\tilde{m})\right)}{Q_a^*-y_R(l)}$. Since secondary markets clear in the absence of capital requirements, $a_f > a_b$ for any Q_a^* if $\bar{L}_b < L_b^*$. 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 $Q_a = \hat{y}_R$. $\square$

B Ruling out Separating Equilibria with Active Secondary Markets

I now state and prove a claim from Section 2.5.2 regarding separating equilibria.

Proposition (No Separation)**.**
If financier bids satisfy bid consistency, then there does not exist a separating equilibrium in which the high-type banker sells a strictly positive amount of risky claims on secondary markets.

Proof. Begin with the first claim. Suppose for a contradiction that both low-type and high-type bankers sell assets on secondary markets ($\underline{a}_b^H, \underline{a}_b^L > 0$) but issue different bond quantities. As a result, the two types of bankers trade on separate secondary sub-markets - mu^H and μ^L, say. Let Q_a^H and Q_a^L denote asset prices on the respective sub-markets. Because low-quality assets cannot be levered against by financiers (their low-state payoff is zero), financiers receive rate of return $R^L = \frac{\hat{y}'_R}{Q_a^L}$ when they buy claims from the low type. When they buy assets from the high type, they receive a return of $R^H = \max\left\{\frac{\hat{y}_R}{Q_a^H}, \frac{\hat{y}_R - y_R(l)}{Q_a^H - Q_b y_R(l)}\right\}$. No arbitrage requires that $R^H = R^L$. The implementability condition IMP implies that $Q_a^H < \hat{y}_R$ – else, the high-type banker would prefer to sell and shirk. It follows that $Q_a^L < \hat{y}'_R$. If this is the case, ex-post optimality in asset sales requires that the low-type banker sells exactly $\underline{a}_b^L$ in risky assets – i.e. he sells no more than he needs to do because asset prices are below the expected value of claims. When setting the asset sale commitment $\underline{a}_b^L$, he thus promises to sell no more than is required to guarantee that he shirks in equilibrium. As a result, the incentive constraint for shirking must hold with equality, and the low-type banker is *exactly* indifferent between shirking and monitoring at $\underline{a}_b^L$. Now consider a deviation by the low-type banker to an asset sale commitment $\underline{a}_b^{L'} = \underline{a}_b^L - \epsilon$ for small but positive ϵ. Conditional on this deviation, the low-type banker strictly prefers to monitor, and trades on sub-market μ'. Since all bankers on sub-market μ' have incentives to monitor, *bid consistency* implies that the secondary market price must satisfy $Q'_a = Q_a^H$. Because $\underline{a}_b^{L'}$ is an arbitrarily small deviation from $\underline{a}_b^L$, the deviating banker must only scale back borrowing and investment by an arbitrarily small amount. Yet the secondary market price he obtains after a deviation is strictly higher for any ϵ, and it applies to all infra-marginal asset sales. Since $y_R(l) > y'_R(l)$, he also maintains solvency in all states of the world. Hence, there always exists a profitable deviation for a low-type banker.

C A Dynamic Model with Endogenous Risk Aversion

In this section, I study a variant of the dynamic model in Section 3 in which old intermediaries have full bargaining power ($\theta = 1$). This choice of parameters implies that the old appropriate the entire *value* of their end-of-life stock of net worth. Generically, this value of net worth 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 net worth, 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 net worth 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 net worth 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 net worth. □

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 bankers 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 net worth 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 banker weakly prefers to sell the asset at price Q_a if and only if

$$\frac{\alpha_b(l)}{\alpha_b(h)} \geq \frac{\pi(y_R(h) - Q_a)}{\pi_l(Q_a - y_R(l))}$$

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

$$\frac{\alpha_f(l)}{\alpha_f(h)} < \frac{\pi(y_R(h) - Q_a)}{\pi_l(Q_a - y_R(l))}.$$

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 Q_a such that financiers are willing to purchase the risky asset and bankers 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 3.

Proposition.
If, after any shock, the competitive equilibrium in the final period is a full-monitoring equilibrium with low secondary market liquidity and a highly constrained financial system then $\sigma_f = \sigma_b = 1$.

Proof. In a low-liquidity equilibrium in which the financial system is highly constrained we have $Q_b^* = \frac{1}{\underline{y}_S}$ and $Q_a^* \underline{Q}_a(Q_b^*)$. By Proposition 6, $\hat{\text{ROE}}_f = \frac{\hat{y}_R - y_R(l)}{\underline{Q}_a(Q_b^*) - Q_b^* y_R(l)}$ and $\hat{\text{ROE}}_f = \frac{\hat{y}_R - y_R(l)\tilde{m}}{1 - Q_b^* y_R(l)\tilde{m}}$. Given that the financial system is highly constrained after any shock, the result follows. □

Proposition 9 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 full-monitoring equilibrium. What remains to be shown is that the economy also transitions into a shirking equilibrium after a sequence of good shocks.

Proposition.
If the competitive equilibrium in the final period is a full-monitoring equilibrium *with low secondary market liquidity and a highly constrained financial system after a bad shock, and a* shirking equilibrium *with low secondary market liquidity and a highly constrained financial system after a good shock, then*

$$\sigma_b = 1 \qquad \textit{and} \qquad \sigma_f = \frac{\frac{\phi \hat{y}'_R + (1-\phi)(\hat{y}_R - y_R(l))}{Q_a^* - (1-\phi) Q_b^* y_R(l)}}{\frac{\hat{y}_R - y_R(l)}{Q_a^* - Q_b^* y_R(l)}} \leq 1$$

where $Q_b^* = \frac{1}{\underline{y}_S}$ and $Q_a^* = \underline{Q}_a(Q_b^*)$.

Proof. For bankers, 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 shirking equilibrium. □

It follows that the economy with endogenous risk aversion must also transition into a shirking equilibrium. To see this, suppose first that the economy with endogenous risk aversion does *not* transition into a shirking 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 shirking equilibrium, yielding a contradiction. Note that $\underline{Q}_a$ 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.

□