Savings Gluts and Financial Fragility*

Patrick Bolton
Columbia University and NBER

Tano Santos
Columbia University and NBER

Jose A. Scheinkman
Columbia University, Princeton University and NBER

December 20, 2017

Abstract

Originators produce higher quality assets at a private cost. These assets can either be bought by informed intermediaries or sold in a pool with low quality assets. Savings gluts diminish *origination incentives* because they compress the spread between the price paid for high quality assets and the price paid for the pool. The narrowing of the spreads relaxes borrowing constraints, which results in higher leverage. Thus savings gluts generate *financial fragility* - the sensitivity of financial intermediaries' equity to unforeseen contingencies. The model offers a coherent narrative of the run up to the crisis.

*We thank Hyun Shin and Chris Sims for insightful comments on an earlier version of the paper. We are grateful to Tobias Adrian and Tomasz Piskorski for the data they have provided to construct Figures 2 and 3.

"Large quantities of liquid capital sloshing around the world should raise the possibility that they will overflow the container." Robert M. Solow page vii in Foreword of *Manias, Panics, and Crashes: A History of Financial Crises* by Charles P. Kindleberger and Robert Aliber (2005)

1 Introduction

Large changes in capital flows have long been linked to financial crises (Kindleberger and Aliber, 2005). The typical narrative is that capital inflows ('hot money') boost asset prices and set in motion a lending and real estate boom. Eventually, when capital flows stop and real estate values decline, a debt crisis ensues (see e.g. Aliber, 2011, and Calvo, 2012). But, as compelling as the historical evidence is, the microeconomic mechanisms that bring about financial fragility as a result of large capital inflows are still poorly understood.

Our paper focuses on three main effects of easy financial conditions: (i) risk-spread compression caused by a savings glut undermines asset origination incentives; (ii) higher asset prices and lower expected returns increase leverage in the financial sector, which; (iii) leads to greater financial fragility.

Three broad stylized facts can be explained by the microeconomic mechanism in our paper. The first is the negative correlation between risk-spreads and the growth in US corporate savings[1] during the years that preceded the great recession, as illustrated in Figure 1.[2] The second broad fact is the deterioration of origination standards for some classes of asset backed securities. Figure 2 Panel A shows the rise in the percentage of private label mortgage loans with poor or no documentation in the years prior to the Great Recession. Panel B in turn shows the average cumulative default paths of non-agency securitized loan vintages from 2000 to 2007. Although default rates at first declined from 2000 to 2003, they rose substantially after 2003.[3] The third stylized fact is that in the run up to the crisis the balance sheets of many

[1]The risk-spread variable we have selected is the U.S. high yield option-adjusted spread, which is a measure of the pricing of risk in credit markets. The savings variable we chose is cash holdings or savings by U.S. corporations, which is the largest component of aggregate U.S. savings. As Karabarbounis and Neiman (2012) emphasize 80% of US savings is by corporations.

[2]This link was also a concern to financial analysts; see Loeys, Mackie, Meggesye and Panigirtzoglou (2005).

[3]See also Keys, Piskorski, Seru and Vig (2013) Figure 4.4. We thank Tomasz Piskorski for providing the

financial intermediaries grew and this growth was mostly driven by increased leverage. Figure 3 displays a regularity first noted by Adrian and Shin (2010): US brokers-dealers grew their balance sheets principally through increased leverage, with repo financing in particular. Shin (2012) documents a similar pattern for large European banks. In addition, in the run-up to the crisis many of these intermediaries accumulated positions in asset backed securities with low quality collateral.

In our model originators can produce assets of low or high quality; it takes additional effort to produce high quality. Originated assets are distributed to investors in two different markets: A securities market where assets are pooled, and another market where the assets are individually purchased by an informed intermediary, who has superior but noisy information about the quality of each individual asset. In analogy to the distinction between futures and forward contracts, we shall refer to the first market as an exchange and to the second market as an over-the-counter (OTC) market.

Our analysis focuses on the incentives for good asset origination financial markets provide, and how these market incentives are affected by changes in savings or capital flows. The market exchange cannot provide any incentives, since investors in this market are unable to distinguish between high and low quality assets. Market incentives for origination of good assets must come from informed investors who are willing to pay more in the OTC market for assets they identify as high quality. A central result of our analysis is that origination incentives at first improve with capital inflows but eventually the savings glut narrows spreads so much that origination incentives are undermined. We focus on the situation where capital in the hands of informed investors is relatively scarce and therefore informed traders obtain a higher rate of return in the OTC market in equilibrium. They can however increase their positions through leverage by borrowing from uninformed investors. Therefore, informed traders act as *financial intermediaries* in equilibrium. We show that intermediary leverage increases monotonically as more savings pour into capital markets, and that it is highest when origination incentives are at their lowest. Since intermediaries make mistakes and face a worse pool of assets, their balance sheets increasingly contain low quality assets.

quarterly data used to construct Figure 2 Panels A and B.

Our definition of financial fragility is inspired by the observation in Gerardi, Lehnert, Sherlund and Willen that “had market participants anticipated the increase in defaults on subprime mortgages originated in 2005 and 2006 the ... extent of the [financial crisis] ... would [have been] very different.”[4] We thus define financial fragility as the sensitivity of the balance sheet to unforeseen contingencies. In our model, the higher leverage leads to a higher sensitivity of the balance sheet of financial intermediaries to unforeseen contingencies: the size of the unexpected shock that would cause a substantial shortfall in the equity of a financial intermediary decreases with the magnitude of the savings glut.

The rising financial fragility prior to the crisis is often attributed to changes in investors' psychology (Minsky 1992). According to Minsky financial fragility is simply due to the growing risk appetite of investors. As we argue below, increased risk-appetite could substitute for an increase in savings in our model. The critical variable is the aggregate risk-weighted capital flowing into asset markets. Thus, our analysis can be also be seen as a microfoundation of a Minsky-type explanation of the lending boom and associated financial fragility.

We conduct much of our analysis under the convenient assumption that there is *cash-in-the-market pricing* of assets in equilibrium in the market exchange (Allen and Gale, 1999). In other words, asset prices are determined by the ratio of aggregate savings and the volume of distributed assets. In this way, a savings glut directly feeds through to higher asset prices. The aggregate savings that investors deploy in the market are exogenously given and we compare equilibria for different levels of aggregate savings, keeping constant the distribution of capital across informed and uninformed investors. We do not take a stand on whether the capital share of uninformed investors grew in the run-up to the crisis.[5]

A key feature of our analysis is that asset prices in the OTC market rise less than proportionately with aggregate savings, so that the overall effect of the rise in aggregate savings is to compress the spread between the price of high quality assets traded in the OTC market and asset prices in the market exchange. In addition, the balance sheets of financial intermediaries grow with aggregate savings. This growth is mostly financed with debt though there is also

[4] See also Cheng, Raina and Xiong (2014).

[5] However we show in an extension that an increase in the capital share of uninformed investors accentuates the results of our model.

some growth in book equity. The end result is that leverage amongst financial intermediaries grows as the savings glut increases. The reason is that, as in Kiyotaki and Moore (1997), the rise in asset prices relaxes leverage constraints, thereby increasing the debt capacity of financial intermediaries.

Since financial intermediaries have information on the quality of originated assets and are willing to pay more for high quality assets, the increase of their balance sheets ought to result in better incentives for originators.[6] However, there is a countervailing effect through the narrowing of spreads, which undoes origination incentives. The effect of savings on origination incentives has an inverted U-shape. Initially, increases in savings improve incentives but eventually they lead to a deterioration of origination standards, as is reflected in Panel B of Figure 2. This is an unexpected result: having a larger fraction of funds in the hands of informed investors does not result in better origination incentives. The reason is that higher asset prices in the market exchange are undoing the effect of more informed funds on origination incentives.

We consider several extensions to our basic model. Behavioral biases most likely did play a role in the run up to the crisis and we explore the effects of the addition of an overconfidence bias of informed investors by allowing for the possibility that these agents exaggerate the precision of the signals they receive about the quality of the assets up for sale, as in Scheinkman and Xiong (2003). When intermediaries overestimate the precision of their signals, they bid more aggressively for assets when they receive a positive signal about asset quality. Obviously, their overconfidence could lead them to overbid and thereby stimulate an asset price bubble. This effect is well understood. But a less obvious and novel insight from our model is that the book leverage of overconfident intermediaries (the ratio of debt to book value of assets) is lower than the book leverage of unbiased intermediaries. What is more, in our model an overconfident intermediary *reports a lower book leverage and has more assets likely to underperform*. This simple but striking observation has implications for the prudential regulation of banks: it illustrates the potentially misleading nature of book leverage as a measure of financial fragility. What appears to be a financially healthy bank by its low reported

[6]Since by construction savings of informed and uninformed traders grow equally, as leverage increases the share of funds at the disposal of informed intermediaries rises.

leverage, could in fact be a bank that has unwittingly taken too much risk.

As we already mentioned, a more conventional but somewhat more involved way of modeling cash-in-the-market pricing would be to have a model with risk-averse investors with risk aversion that decreases with wealth. The empirical asset pricing literature documents that compensation for risk is countercyclical: high during troughs and low at the peak of the cycle.[7] Under this formulation asset prices on the exchange are affected by both the fundamental quality of assets and the risk attitudes of uninformed investors. The key mechanism is then that a reduction in the discount rate leads to a reduction in the spread between the price for high quality assets in the OTC market and the uninformed price on the market exchange. The Allen and Gale notion of cash-in-the-market pricing should therefore be thought of more generally as simply a more tractable way of modeling the effects of decreasing risk aversion when savings rise.

Related literature. We provide microfoundations for one of the leading hypotheses on the origins of the crisis of 2007-09: a savings glut combined with a balance-sheet expansion and greater leverage of financial intermediaries. Other commentators, most notably Bernanke (2005), Shin (2012), Gourinchas (2012), and Borio and Disyatat (2011) have argued that the rise in global liquidity more than global imbalances was the major cause of the crisis. Bernanke (2005) famously argued that a 'global savings glut' was the main cause of low long-term interest rates before the crisis of 2007-09. With low interest rates, households could afford bigger mortgages, which, in turn, fuelled real-estate price inflation. Bernanke's thesis is sometimes narrowly interpreted as meaning that the rise of Chinese savings caused the fall in US interest rates, but this narrow interpretation is not completely borne out in the data.[8] The savings glut hypothesis, however, is *not* dependent on a particular source of savings. Any surge in savings, whatever the sector, yields similar implications.

Shin (2011, 2012) points to the role of rising gross capital flows across countries, channelled through an increasingly globalized banking industry, as a major source of financial fragility. He attributes this increase in cross-border flows to the transition to Basel II capital regulation, which allowed banks to lever up, and to the advent of the Euro, which eliminated

[7] The literature on stock return predictability is large. See Cochrane (2008) for a survey and summary.

[8] See Bernanke, Bertaut, Pounder DeMarco, and Kamin (2011).

currency risk within the Eurozone.[9] With higher leverage and greater reliance on fickle foreign capital inflows, Shin argues that banks became fundamentally more fragile. It would not take much to put their viability at risk with a liquidity crunch. Borio and Disyatat (2011) also emphasize the critical role played by the financial sector in the crisis. They argue that the main cause of the crisis was the dynamic expansion of the balance sheets of large, complex, financial institutions in response to the savings glut. The 'excess elasticity' of bank balance sheets, the term they employ, was the main cause of the crisis in their analysis. In effect, banks acted as a multiplier of the savings glut, pouring vast quantities of new fuel on the fire.

More systematic evidence that financial crises are preceded by above-trend credit growth was found by Gourinchas, Valdés and Landerretche (2001) for the period of the 1960s to 2000. More recently Jordà, Schularick and Taylor (2011) have looked at a much longer time-series (140 years) and have found that abnormally high credit growth and low interest rates precede global financial crises. Moreover, they find that credit growth is a better predictor of financial crises than global imbalances, the latter playing an additional role only before WWII.

Several other theories linking savings gluts to financial instability have been proposed. An early model by Caballero, Fahri and Gourinchas (2008) links rising global imbalances to low interest rates, through a limited global supply of safe assets. However, they do not explore the effects of these imbalances on origination standards, leverage, and financial fragility. In independent related work, Martinez-Miera and Repullo (2016) propose a model of intermediation similar to Holmstrom and Tirole (1997), where banks' incentives to monitor are affected by a savings glut. In their model, safe projects are financed by non-monitoring (distributing) banks, while riskier projects are held on the balance sheet of traditional monitoring banks. Their theory is built on a different economic mechanism than ours and makes somewhat different predictions. In particular, any effect of the savings glut on intermediary leverage is absent from their model. A related theory by Neuhann (2016) points to the additional fragility caused by excessive 'risk transfer' from originating banks to 'financiers'. This risk transfer gradually undermined origination incentives of commercial banks as more and more risk was transferred to a rising non-bank sector. Boissay, Collard and Smets (2016) consider a dynamic

[9] See Santos (2017) for an analysis of capital flows into the Spanish banking sector in the years leading up to the Eurozone crisis.

model of the interbank market, where banks borrow from other lenders. Borrowing is limited by adverse selection problems: Lenders don't know whether they are lending to banks with good or bad investment opportunities. As interbank rates rise only the banks with the best investment opportunities borrow in the interbank market, so that the volume of interbank loans increases and the pool of borrowing banks improves. In other words, there is a positive correlation between interbank rates and leverage in their model. A crisis occurs when there is a sudden increase in savings, which causes interbank rates to drop, and consequently leads to a deterioration of the pool of borrowing banks, together with a collapse in interbank lending volume. In contrast, in our model intermediaries borrow in a repo market and leverage and repo rates are negatively related. This seems to better match the stylized facts in the years leading up to the Great Recession, when investment banks and broker-dealers increased the size of their balance sheet size and leverage, increasingly relying on repo financing.

An alternative theory of financial fragility that does not depend on savings gluts is developed by Gennaioli, Shleifer and Vishny (2012 and 2013). In their model some agents underestimate or neglect tail risk, perhaps because they extrapolate from recent history. Financial intermediaries that identify this misperception can exploit it through securitizations to build up more risk. In addition, to control their residual risk, financial intermediaries purchase claims against other intermediaries which generates excessive financial instability. We consider the possibility of misperceptions by financial intermediaries instead of by final investors in Section 6.4 and show how it can lead to additional financial fragility by further undermining origination standards.

2 The Model

The model we develop focuses on the financial market mechanism linking the pricing of assets in financial markets and incentives of originators to supply high quality assets. In particular, our analysis centers on the question of how this mechanism is affected by changes in aggregate liquidity flowing into financial markets. Accordingly, our model must comprise at least two classes of agents, asset originators and investors, interacting over two periods.

2.1 Agents

We assume that each class of agents is of fixed size (we normalize the measure of each class to 1), and that both originators and investors have risk-neutral preferences.

Originators. In period 1 each originator can generate one asset that produces payoffs in period 2, which are either $x_h > 0$ or $x_l = 0$. An asset can be interpreted to mean a business or consumer loan, a mortgage, or other assets. The quality of an originated asset depends on the amount of effort $e \in [\underline{e}, 1)$ exerted by the originator, where we assume that $\underline{e} > 0$. Without loss of generality we set the probability that an asset yields a high payoff x_h equal to the effort e. Asset payoffs are only revealed in period 2, so that the only private information originators have in period 1 is their choice of effort. Originators only value consumption in period 1 and they incur a disutility cost of effort e, so that their utility function takes the form:

$$u(e, c_1) = -\psi(e - \underline{e}) + c_1, \tag{1}$$

where c_1 stands for consumption in period 1. We assume that the disutility of effort function $\psi(z)$ satisfies the following properties: (i) $\psi(0) = \psi'(0) = 0$; (ii) $\psi'(z) > 0$ if $z > 0$; (iii) $\psi'(1-\underline{e}) > x_h$ and (iv) $\psi''(z) >> 0$. Given that $\psi(\underline{e}) = 0$, originators always (weakly) prefer to originate an asset as long as they can sell this asset at a non-negative price in period 1.

Investors. Each investor has an initial endowment of K units of capital in period 1 and a utility function

$$U(c_1, c_2) = c_1 + c_2,$$

where again $c_\tau \geq 0$ denotes consumption at time $\tau = 1, 2$. Since investors are indifferent between consumption in period 1 and 2, they are natural buyers of the assets that originators would like to sell in period 1. Our model captures in a simple way changes in aggregate savings by varying K.

There are two types of investors. A first group, which we refer to as *uninformed investors,* are unable to identify the quality of an asset for sale. We denote by M the fraction of uninformed investors and assume that $0 < M \leq 1$. The second group, which we label as *informed investors* are better able to determine the quality of assets and can identify those assets that are more likely to yield a high payoff x_h. We let $N = 1 - M$ denote the fraction of informed intermediaries.

2.2 Financial Markets

There are three different financial markets in which agents can trade: 1) an opaque, over-the-counter (OTC) market where originators can trade assets with informed investors; 2) An organized, competitive, transparent and regulated exchange where originators can sell their asset to uninformed investors; and 3) A secured debt market, where informed investors can borrow from uninformed investors. The dual market structure for assets builds on Bolton, Santos and Scheinkman (2012 and 2016). A key distinction between these two types of markets is how buyers and sellers meet and how prices are determined. In the organized exchange all price quotes are disclosed, so that effectively asset trades occur at competitively set prices. In the private market there is no price disclosure and all transactions are negotiated on a bilateral basis between one buyer and one seller.[10]

2.2.1 OTC Market

Originators are willing to trade with informed investors in the OTC market despite the lack of competition among intermediaries and the lack of transparency in the hope that their asset will be identified as a high quality asset. Informed intermediaries observe a signal that is correlated with the quality of the asset and are willing to pay a higher price for high quality assets than the price at which a generic asset is sold on the organized exchange. Even though informed traders are free to buy in any market, we show that when informed capital is scarce and uninformed capital is not too scarce, they only operate in the OTC market in equilibrium, and they only purchase assets that they judge to be high quality. As we detail below, although informed traders can borrow from uninformed investors, they have a limited borrowing capacity, due to

[10] In an unpublished version of Bolton, Santos and Scheinkman (2016) we developed a version of our framework where instead of two markets we have a single Grossman-Stiglitz exchange. The results are unaffected. We retain the double market structure for convenience but also realism. As we have argued in Bolton, Santos and Scheinkman (2012), a feature of modern financial markets is the split of risks across private, informational intensive markets where the costs of participation are high due to regulatory barriers or infrastructure needs and public markets, which are less informationally intensive and where the costs of participation are low.

the collateral requirements. Even after exhausting their borrowing capacity, informed traders may not have sufficient capital to purchase all available high quality assets. In this case, any asset that informed traders are not able to purchase will be sold on the exchange.

Each informed trader observes a signal $\sigma \in \{\sigma_h, \sigma_l\}$ on the quality of any asset offered for sale such that

$$\text{prob}\,(\sigma_h|x_h) = 1 \qquad \text{and} \qquad \text{prob}\,(\sigma_h|x_l) = \alpha \in [0, 1). \tag{2}$$

The parameter α captures possible valuation mistakes of informed trader. Conditional on observing σ_h the probability that the asset yields a payoff x_h is:

$$g := \text{prob}\,(x_h|\sigma_h) = \frac{e}{e + \alpha(1-e)}. \tag{3}$$

Note that when $\alpha = 0$ we have $g = 1$; in other words, σ_h is a perfectly informative signal for the payoff x_h. The higher is α the less informative is the signal σ_h, and when $\alpha = 1$ there is no additional information conveyed by σ_h.

Conditional on observing σ_l the investor knows that the asset will pay 0 in period 2. In this case, the investor would offer to pay no more than zero for the asset, so that the originator prefers to sell the asset on the anonymous exchange. In sum, all OTC trades with informed investors involve high quality assets with signal σ_h.

We denote by p^d the price at which an asset with a signal σ_h trades in the OTC market, and by p the competitive price for a generic asset on the exchange. As in Bolton, Santos and Scheinkman (2012 and 2016) we assume that a bilateral trade on the OTC market takes place at negotiated terms via Nash bargaining[11] and that

$$p^d = \kappa g x_h + (1-\kappa)p, \tag{4}$$

[11] In Bolton, Santos and Scheinkman (2016) we offered a discussion of the Nash bargaining solution based on ideas in Binmore, Rubinstein and Wolinsky (1986). Briefly, the asymmetric Nash bargaining solution that we use in our framework can be justified as an equilibrium outcome of a strategic model of bargaining with uncertain termination time. In this class of bargaining games, the threat point is given by the outcome in the event that the bargaining process break down. Our assumption is that in the case of a breakdown of the bargaining game the informed investor consumes his endowment and the originator sells her asset in the exchange. In Binmore et al. (1986) the parameter κ reflects the relative subjective probabilities that each party attributes to the breakdown of the bargaining process. In the sequential strategic justification for the asymmetric Nash bargaining solution developed by BRW, the number of dealers relative to originators does not affect κ.

where $\kappa > 0$ denotes the bargaining strength of the originator. Note that once an informed investor offers any positive price, the originator learns that the signal associated with her asset is σ_h. The bargaining parties are then symmetrically informed. The pricing formula (4) is the *Nash bargaining solution* of a bargaining game between an informed investor and an originator for the purchase of an asset with signal σ_h, where the bargaining weights of the originator and investor are respectively κ and $(1-\kappa)$, and where the *disagreement point* is given by a sale of the asset on the exchange at price p. This OTC bargain only takes place if $gx_h \geq p$, which is a condition satisfied by the equilibrium price on the exchange p, as we show below.

Let q^i denote the number of assets acquired by each informed trader on the OTC market. The probability that an originator with an asset with signal σ_h sells her asset to an informed trader in a symmetric equilibrium is then given by the ratio of the total number of assets purchased by informed intermediaries to the total number of assets with signal σ_h (provided that this ratio does not exceed 1):

$$m := \min\left\{\frac{Nq^i}{e+(1-e)\alpha}; 1\right\}. \tag{5}$$

2.2.2 Organized Exchange

All originated assets that are not cream-skimmed in the OTC market end up being distributed on the organized exchange. Therefore, the volume of high-quality assets traded on the exchange in a symmetric equilibrium is equal to $e\,(1-m)$. To see this, observe first that originators produce assets with payoff x_h with probability e. The fraction of assets x_h in a symmetric equilibrium is then e, of which a fraction m is bought by financial intermediaries. Second, the volume of low-quality assets sold on the exchange is $(1-e)-(1-e)\alpha m$, so that the total volume of assets distributed on the exchange is $1-em-(1-e)\alpha m$. The expected value of an asset traded on the exchange is then:

$$p^f = \frac{e(1-m)x_h}{1-em-(1-e)\alpha m}, \tag{6}$$

where the superscript f stands for “fair value”. Expression (6) highlights that the fair value of assets traded on the exchange depends on the fraction of assets that are cream-skimmed by informed investors.

In Bolton, Santos and Scheinkman (2016) we assume that risk-neutral uninformed investors are always able to pay the fair value p^f and focus on the implications of cream-skimming of high quality assets in the OTC market. Here, we generalize the model to allow asset prices to respond to changes in aggregate liquidity. Specifically, if the total stock of liquidity brought by investors to the exchange is T, then we assume that asset prices on the exchange must satisfy:

$$p \leq p^{\text{cim}} := \frac{T}{1 - em - (1-e)\alpha m}, \tag{7}$$

so that:

$$p = \min\{p^f; p^{\text{cim}}\} = \min\left\{\frac{e(1-m)x_h}{1 - em - (1-e)\alpha m}; \frac{T}{1 - em - (1-e)\alpha m}\right\}, \tag{8}$$

since $p \leq p^f$, $p \leq gx_h$ or $p \leq p^d$.

The superscript "cim" stands for *cash-in-the-market pricing*, which obtains whenever the total pool of liquidity is less than $e(1-m)x_h$.

It is helpful to introduce two further pieces of notation: We denote by R the return from buying an asset with signal σ_h on the OTC market, and by r^x the return from investing on the exchange, when $p > 0$. That is:

$$R = \frac{gx_h}{p^d} \qquad \text{and} \qquad r^x = \frac{e(1-m)x_h}{p(1 - em - (1-e)\alpha m)}. \tag{9}$$

2.2.3 The Secured Credit Market

Investors can borrow from other investors in the form of risk free collateralized loans akin to *repo* contracts. Under a typical loan an investor borrows at the rate r against the assets it acquires. Since informed traders have access to all trades available to uninformed traders we may assume without loss of generality that in equilibrium only informed traders act as borrowers. Formally, we assume that an informed intermediary can borrow against purchased assets at the exchange market value p of these assets with a haircut $\eta > 0$. More precisely, if an intermediary purchases y units it can borrow any amount D that satisfies the constraint

$$D \leq (1-\eta)\, py. \tag{10}$$

Note that the valuation of the collateral in constraint (10) is determined using the exchange price p. In section 6.2 we justify using the exchange price p as the reference price, by assuming an intermediary period between periods 1 and 2, when lenders hit by a liquidity shock have the option not to rollover a loan. If the loan is not rolled over, the lender ends up as the owner of an asset that he can only sell in the exchange.[12] Constraint (10) also features a haircut η, which, in practice, reflects lenders' beliefs about the risk with respect to selling the asset in the secondary market.

The secured credit market plays a critical role in what follows. In particular, although the fraction of capital endowed to informed investors is invariant to the per-capita endowment K, the fraction of funds deployed by informed investors varies with K in equilibrium. Let

$$f^i := \frac{N\left(K+D\right)}{\left(N+M\right)K} = \frac{N\left(K+D\right)}{K} \qquad \text{and} \qquad f^u := \frac{MK-ND}{K} = 1-f^i, \tag{11}$$

denote the shares of total funds deployed by informed and uninformed investors, respectively. If $D>0$ the fraction f^i exceeds the fraction of capital initially held by informed investors, N. We will establish how the equilibrium shares react to changes in K, that is how the distribution of deployable funds with respect to knowledge varies with changes in aggregate endowment.

2.3 Expected Payoffs

Having described how originated assets are distributed and priced in period 1, we are now in a position to specify originator and investor payoffs. Originators sell assets to investors in period 1, who hold them until maturity. Replacing c_1 in (1) with the expected price of the originated asset we obtain the following expression for an originator's expected payoff:

$$-\psi\left(e-\underline{e}\right)+e\left(mp^d+(1-m)p\right)+(1-e)\left[\alpha\left(mp^d+(1-m)p\right)+(1-\alpha)p\right]. \tag{12}$$

An originator who chooses origination effort e is able to generate an asset with a payoff x_h with probability e. In this case he is matched with an informed investor with probability m and

[12] Kiyotaki and Moore (1997, equation (3), page 218), take the price of the asset at time $t+1$ to value the collateral in their borrowing constraint. Our specification reflects the market practice for repo transactions, in which valuation of the collateral is based on the current market price. Nevertheless, constraint (10) also incorporates the main feedback effect of Kiyotaki and Moore (1997), whereby increases in asset prices can relax collateral constraints for secured loans.

obtains a price p^d in the OTC market, whereas if he is not matched with an informed investor he sells the asset on exchange for a price p. If he originates an asset with a low payoff, he is still able, with probability α, to obtain a price p^d if matched to an informed investor who obtains a good signal σ_h.

Uninformed investors do not have access to the OTC market. They can acquire assets on the exchange or lend to informed intermediaries. Let $q^u \geq 0$ be the quantity of assets bought on the exchange and $D^u = q^u p - K$ the amount borrowed by uninformed investors (if $K > qp^u$ uninformed investors are net lenders in the repo market). Then, an uninformed investor's expected payoff is given by:

$$V^u\left(q^u\right) := q^u p r^x - D^u r. \tag{13}$$

Feasible choices for D^u satisfy the leverage constraint:

$$\eta p q^u \leq K. \tag{14}$$

Informed investors choose an amount $q^i \geq 0$ to purchase in the OTC market and an amount $y - q^i \geq 0$ to purchase on the exchange. Informed investors borrow the amount $D^i = p^d q^i + p(y - q^i) - K$ (or lend if this amount is negative). An informed investor's expected payoff is thus given by:

$$\begin{aligned} V^i\left(q^i, y\right) &= q^i p^d R + \left(y - q^i\right) p r^x - D^i r \\ &= q^i p^d (R - r) + p(y - q^i)(r^x - r) + Kr \end{aligned} \tag{15}$$

Again feasible choices for D^i satisfy the leverage constraint:

$$\eta p y \leq K - (p^d - p) q^i. \tag{16}$$

We will be especially interested in situations where $R > r^x = r$. In such situations informed intermediaries only trade in OTC markets, so that $q^i = y$ and the leverage constraint takes the form:

$$D^i = p^d q^i - K \leq (1 - \eta) p q^i. \tag{17}$$

Note that the constraint (17) features two margins. The first is the standard haircut in secured transactions as captured by the parameter η. The second is slightly more subtle and arises because informed investors acquire assets in private transactions at a price p^d, but the collateral value of those assets is $p < p^d$. In effect, informed equity capital is required to acquire assets in OTC markets.

2.4 Cash-in-the-Market (CIM) Equilibrium

We parameterize our economy by (K, N, α) and determine under what conditions a unique equilibrium exists. We then characterize how the equilibrium varies with (K, N, α). Given (K, N, α), a vector of equilibrium prices (p, p^d, r, r^x, R) and quantities ($e, g, q^u, D^u, q^i, y, D^i, m$) is such that:

1. g, p^d and m satisfy equations (2)-(4),

2. p satisfies (8) when $T = p(q^u + y - q^i)$,

3. $(e, q^u, D^u, q^i, y, D^i)$ solve the maximization problems of respectively originators, uninformed investors, and informed intermediaries and furthermore the following market clearing conditions hold:

$$Mq^u + Ny = 1 \qquad \text{and} \qquad MD^u + ND^i = 0.$$

4. $rD \leq q^i g x_h$, that is, collateralized loans are indeed risk free.

A first immediate observation is that there is no equilibrium where $r > r^x$. The reason is that if $r > r^x$ uninformed investors strictly prefer to lend all their savings to informed intermediaries and the latter also prefer to lend rather than purchase assets on the exchange. It follows that we must then have $p = 0$. But then informed investors cannot borrow, as the leverage constraint always binds, so that $MD^u + ND^i < 0$.

A second immediate observation is that if $p > 0$ and $R > r$, then it is optimal for informed investors to borrow as much as possible and earn the spread $(R - r)$ on every dollar they borrow, so that their leverage constraint binds. Moreover, when $R > r$ an equilibrium can exist only if $r = r^x$. If $r < r^x$ then both informed and uninformed investors want to borrow, so that D^i and D^u are positive and $MD^u + ND^i > 0$. In sum, in any maximal leverage equilibrium we must have $r = r^x$.

Definition 1: A *fair value equilibrium* is such that $r^x = 1$.

Definition 2: A *CIM equilibrium* is such that $r^x > 1$, and a *strict CIM equilibrium* is a CIM equilibrium such that $R > r > 1$ and $p > 0$.

To establish that a candidate configuration is indeed is a strict CIM equilibrium it suffices to verify 1-3 above. Condition 4 follows because

$$rp = r^x p = \frac{e(1-m)x_h}{1 - em - (1-m)\alpha m} < ex_h < gx_h$$

and thus

$$rD \leq \frac{D}{p} gx_h \leq (1-\eta)\, qgx_h < qgx_h,$$

where we have used (10) to obtain the second inequality.

Our analysis focuses mostly on strict CIM equilibria. In a strict CIM equilibrium there is market segmentation, with informed investors trading only in the OTC market (that is $q^i = y$) and uninformed investors trading on the exchange. In addition, informed investors act as intermediaries and borrow from uninformed investors, as $R > r$ and $r^x = r$. Since $R > r$, informed investors exhaust their leverage constraint. We illustrate the financial flows that obtain in a strict CIM equilibrium in Figure 4.

Strict CIM equilibria are of special interest because changes in aggregate liquidity affect asset prices in both markets in equilibrium. That is, in a strict CIM equilibrium assets sell below their fair value on the exchange, where $r^x > 1$, and the capital that informed intermediaries can garner is not sufficient to purchase all high quality assets. Intermediaries are constrained by their borrowing capacity, so that $m < 1$ and $p > 0$.

In what follows we show that strict CIM equilibrium exist for certain configurations of parameters (K, N, α). We then derive the main comparative statics results with respect to K and use these to interpret the main stylized facts on asset origination and leverage of financial intermediaries that preceded the financial crisis of 2007-09. We also derive necessary and sufficient conditions for an equilibrium such that $R > r$, which we report in Appendix A.1.

3 Strict CIM Equilibrium

In Appendix A.1 we show that we can completely determine all prices and quantities in a strict CIM equilibrium, once we know the vector $(p; e)$. For this reason we often refer to a strict CIM equilibrium $(p; e)$ in what follows. We can characterize the necessary conditions for existence of a strict CIM equilibrium as the solution to a single pair of equations $f(p, e) = 0$ given in Appendix A.1. We show that the converse result also holds: Starting from a solution (p, e) to $f(p, e) = 0$, we can construct a unique candidate equilibrium with $r = r^x$ that is a strict CIM equilibrium provided that $R > r^x > 1$. The system of equations $f(p, e) = 0$ also allows us to characterize the main properties of strict CIM equilibria.

3.1 Existence and Uniqueness of a strict CIM equilibrium

We begin the analysis by establishing the existence of a strict CIM equilibrium in an economy where the measure N of informed intermediaries is small. When this measure is small enough, the aggregate capital available in the OTC market will be too small to absorb all originated high quality assets. To save on notation we write for each $\tilde{\alpha} \geq 0$, the minimum probability that a high quality asset (with signal σ_h) yields the payoff x_h as:

$$\tilde{g} := g(\tilde{\alpha}, \underline{e}) = \frac{\underline{e}}{\underline{e} + \tilde{\alpha}(1 - \underline{e})}.$$

Our main existence and uniqueness result is then as follows.

Proposition 1 *Fix $\alpha \leq \bar{\alpha} < 1$. Then there exist $K_a < K_b$ such that for $K_a < \tilde{K} < K_b$ there is a pair $(p(K, N), e(K, N))$ that is the unique equilibrium corresponding to parameters (K, N, α) in the neighborhood $|K - \tilde{K}| < \epsilon$ and $0 < N < \epsilon$ where $\epsilon = \epsilon(K, \alpha) > 0$. Moreover $(p(K, N), e(K, N))$ is a strict CIM equilibrium and $p(K, N)$ and $e(K, N)$ are C^2.*

Proof: See appendix.

This Proposition guarantees that if $K_a < \tilde{K} < K_b$ one can find smooth functions that define the unique equilibrium for K in a neighborhood of $\tilde{K}$ and N small. Furthermore this equilibrium is a strict CIM equilibrium. It is intuitive that one must impose limitations on K

to obtain a strict CIM equilibrium. First, K cannot be too low for otherwise the expected rate of return r^x on the exchange would be so large that informed investors would prefer to deploy their capital in the exchange. Second, K cannot be too high for then there would be so much liquidity in the economy that assets would trade in the exchange at their fair value p^f. Explicit formulas for these bounds are stated as part of the proof in the appendix.

In a strict CIM equilibrium informed intermediaries earn a strictly positive spread on any dollar they borrow, so that they borrow up to their debt capacity ((17) is met with equality). Uninformed investors lend to informed intermediaries and earn the same expected return on the collateralized debt claims they hold as on the assets they purchase on the exchange, so that: $D^i > 0$, $D^u < 0$, and $R > r^x = r > 1$. Accordingly, in a strict CIM equilibrium capital flows across markets as illustrated in Figure 4.

4 Savings gluts, asset prices and origination incentives

We are now in a position to study the central question of our analysis, namely how the strict CIM equilibrium is affected by changes in aggregate savings K. We have already established that for certain regions of parameters (K, N, α) for every α there exist smooth functions $p(K, N)$ and $e(K, N)$ that define a CIM equilibrium that is the unique equilibrium of our model. We now derive key comparative statics properties of these CIM equilibria, holding α and N constant as a function of K, the per-capita amount of capital available to agents. In particular, we characterize how equilibrium asset prices in the exchange, p, origination incentives, e, and the balance-sheet of (informed) financial intermediaries vary with K. For notational simplicity we write the equilibrium functions as $p(K), e(K)$ etc.

4.1 Asset Prices

Consider first the CIM prices in the exchange, given by (7). It is not immediately obvious that an increase in K results in higher prices p, as there is a direct and an indirect effect. The direct effect of an increase in K is that investor savings MK increase, which should result in higher asset prices, except that intermediaries also increase their borrowing ND^i, thereby reducing the savings that investors channel to the exchange. Still, the next proposition shows that, provided NK is small, the net effect of an increase in K is an increase in the price p. Intuitively, if capital in the hands of informed intermediaries is small, so will be the share of incremental savings of uninformed investors that flows to informed intermediaries through the secured debt market. Most of the increase in uninformed capital is then invested in the exchange, thereby pushing up asset prices on the exchange.

Proposition 2 *If NK is small enough then $p(K)$ is an increasing function of K.*

Proof: This follows immediately from Lemma A.5. □

Propositions 1 and 2 are illustrated in Figure 5. As can be seen in Panel A, when $K = 1$ the fair value of assets on the exchange is $p^f \simeq 2.7$ while the CIM price is much lower at $p \simeq 0.8$. As K increases more money flows into asset markets, which pushes up the price p but also results in more cream-skimming by informed intermediaries in the OTC market. There is a direct effect and an indirect effect from this greater cream-skimming, as we explain in greater detail below when we look at the comparative statics of e with respect to K. The direct effect is that, other things equal, a worse quality pool of assets is sold on the exchange as a result of the greater cream-skimming. The indirect effect, which at first dominates, is that the greater cream-skimming improves origination incentives, so much so that the average quality of assets for sale on the exchange net of cream-skimming is also improved. As K rises further, the direct effect dominates at some point, so that the greater cream-skimming results in a deterioration of expected asset quality on the exchange and therefore a reduction in p^f. Eventually, as K is increased even further, $p = p^f$, at which point there is no more cash in the market pricing. When $p = p^f$ additional increases in K only affect asset prices to the extent that they change the average quality of assets for sale on the exchange.

Consider next Panel B, which plots the expected rate of return of informed intermediaries, R, and the expected rate of return of uninformed investors, r^x (which equals the equilibrium interest rate r), as a function of aggregate savings K. At the smallest value of K the two returns are equal, $r^x = R$, at which point a strict CIM equilibrium ceases to exist.[13] As K

[13]Recall that for α small, as p is increases with K, asset prices in the OTC market p^d increase as well (see (4)).

rises beyond this low value informed intermediaries returns R grow larger and larger relative to uninformed investors' returns r^x. But, note that both returns decline as more savings get channeled into asset markets. In sum, Figure 5 plots the strict CIM equilibrium for the entire admissible range of K. At the lowest value for K we have $r^x = R$, and at the highest value for K we have $p = p^f$.

Proposition 2 is an admittedly intuitive, yet fundamental, result for our analysis. It establishes under what conditions increases in aggregate savings result in a *glut* that has the effect of increasing asset prices. As we show next, changes in asset prices also affect origination incentives and the expected quality of the assets traded on the exchange.

4.2 Origination Incentives

The asset price changes induced by changes in K in turn affect origination incentives. We will show that at first an increase in aggregate savings improves origination incentives, but eventually, as a savings glut emerges, origination incentives are impaired. As we note in Appendix A.1, the first order condition for effort optimization (see equation (12)) is

$$\psi'(e - \underline{e}) = (1-\alpha)\, m\left(p^d - p\right). \tag{18}$$

This condition makes clear that there are three determinants of origination incentives:

1. The size of the spread $\left(p^d - p\right) = \kappa(gx_h - p)$ naturally affects origination incentives. If p is very close to p^d, there is little point in exerting costly effort to produce good assets, given that the price paid by intermediaries is very close to the price paid by an uninformed investor.

2. The probability of selling a good asset to an (informed) financial intermediary; the higher is m, the higher are the incentives to produce high quality projects.

3. The precision of intermediaries' information about asset quality captured by the term $(1-\alpha)$. The higher the precision $(1-\alpha)$ with which an asset yielding x_h can be identified the higher are the incentives to originate a good asset.

The next proposition shows that as long as (informed) intermediaries' aggregate capital is not too large and the haircut on the collateralized debt is sufficiently small, increases in savings K at first improve origination incentives, but eventually reduce them when savings have reached a critical high level.

Proposition 3 *(Single peakedness of effort) Fix N and α. Suppose $(p(K); e(K))$ are continuous functions that define a strict CIM equilibrium for K in an interval (K_1, K_2) with NK_2 sufficiently small. If $\eta < \kappa$ then (a) $e'(K_1) > 0$ for K_1 small enough; (b) If K_2 is large enough, $e'(K_2) < 0$; and (c) the function $e(K)$ is either monotone or has a single global maximum.*

Proof: This is a consequence of Lemma A.5. □

It is intuitive that κ must be sufficiently large for $e(K)$ to be non-monotonic. For if $\kappa = 0$ then $\left(p^d - p\right) = 0$, so that there would be no origination incentives at all. When the haircut η is larger, intermediaries can borrow less, so that *ceteris paribus* m is lower other things equal. To make up for the lower m, κ must be larger to preserve incentives.

Figure 6 illustrates this single-peakness. Here we assume that $\psi(e) = \theta\frac{e^2}{2}$, with $\theta = 0.25$; $\kappa = 0.15$; $\eta = 0.5$; $M = 0.75$, and $x_h = 5$. Panel B shows that, in this example, $p^d - p$ is decreasing with K while m is increasing. It is natural to expect $p^d - p$ to go down with K since p is increasing in K.[14] The increase in K places more capital in the hands of intermediaries and the decrease in $p^d - p$ allows them to finance a larger share of their purchases. These forces lead to an increase in m. Since effort depends on $m \times (p^d - p)$, the sign of m' dominates when $p^d - p$ is large and the sign of $(p^d - p)'$ dominates when m is large. The direct effect of spread compression is to reduce incentives to bring higher quality assets to the market. This is a reason for savings gluts to undermine origination incentives. This example also illustrates a central point of our analysis. That the reduction in spreads should reduce incentives to originate is obvious. What our analysis emphasizes is that the impact of the reduction in spreads is mediated by the amount of funds (including leverage) that the intermediaries deploy and, as we show in the next section, this amount increases as spreads are reduced.

[14] This ignores the effect on g from a possible increase in effort, which increases $p^d - p$. This effect is second-order near $\alpha = 0$.

Proposition 3 provides an explanation based on economic fundamentals for why *late in a lending boom* origination incentives deteriorate. As Figure 2 shows, the proportion of incomplete or no documentation private label mortgage loans increased consistently in the years up to the Great Recession (Panel A), yet the delinquency rates first improved for the vintages or 2001 to 2003, and only started to deteriorate in the final phase of the run-up to the crisis (Panel B).[15]

The deterioration of origination standards is a phenomenon commonly observed in episodes shortly preceding the onset of a financial crisis, and it has puzzled economic historians (see Kindleberger and Aliber, 2005). Minsky (1992), most notably, offers a popular explanation of this phenomenon rooted in investor psychology. It emphasizes the growing risk appetite, to the point of recklessness, of investors over the lending boom. It is easy to find examples of investor irresponsibility in boom periods, which lend credibility to this explanation. Yet, economists have generally been wary of explanations based on changing investor psychology because of their fundamental circularity: booms occur because of growing investor irresponsibility, and investors become more reckless because of the boom. Hence, one advantage of our theory is that we do not need to appeal to changing psychological factors to explain the decline in asset quality origination. Of course, to the extent that such behavior is prevalent it would reinforce our underlying economic mechanism. Moreover, our mechanism based on economic fundamentals is closely intertwined with two other phenomena also observed before financial crises, the rise in asset demand and the rise in intermediary leverage.

[15]There were, of course, other forces in play around the crisis of 2007-2009, in particular house price dynamics, which led to lower rates of non-performing mortgages as long as prices were rising and to a sharp increase, when house prices collapsed during the crisis. Note however that the increase in delinquencies for mortgages originated in 2004 that were at most 5 quarters old when compared to similarly seasoned mortgages originated in 2002 or 2003 cannot be explained by house-price deterioration since these prices only peaked in the second quarter of 2006.

5 Savings gluts and financial fragility

In this section we show how savings gluts affect financial stability. First, and as observed, we show that large savings gluts result in the growth of the balance sheets of informed financial intermediaries. Borio and Disyatat (2011) use the term “excess elasticity” to refer to the this expansion of the financial system in good times. Moreover as Shin (2012) emphasizes, this growth in balance sheets is accompanied by an increase in leverage, as measure by the growth in the ratio of debt to equity.[16] Our model produces this effect as well. We go further than the existing literature in showing that in addition to the growth in the balance sheets and leverage, savings gluts result in a deterioration of the quality of the assets in the asset side of the balance sheet. Thus when savings gluts are pronounced, the financial sector is characterized by large balance sheets, high leverage and bad assets. Second, the financial sector becomes more sensitive to unforeseen contingencies and thus more fragile the larger the savings glut.

5.1 Intermediaries' Balance Sheets

How do savings gluts affect the balance sheets of intermediaries? In this section we examine the effects of an increase in per-capita K on the asset and liability sides of the balance sheet. We show that increases in K increase leverage and eventually lead to a deterioration of the quality of the assets held by financial intermediaries. As we will argue in Section 5.2, late in the savings-glut cycle intermediaries' balance sheets are particularly sensitive to unforeseen contingencies.

In a strict CIM equilibrium informed intermediaries exhaust their borrowing capacity, so that (17) is met with equality. When aggregate savings K increase so do asset prices $p(K)$, which relaxes the constraint (17), thereby increasing intermediary leverage. More formally, define leverage as:

$$\mathcal{L} = \frac{\text{Total assets}}{\text{Book equity}} = \frac{K + D^i}{K} = 1 + \ell^i \qquad \text{with} \qquad \ell^i := \frac{D^i}{K}.^{17}$$

[16] See in particular Figures 3 and 4 in Shin (2012).

[17] Note that this definition takes the marks of informed intermediaries, $p^d q^i$, to value their assets rather than pq^i. The reason is that otherwise intermediaries would be immediately marking down newly acquired assets,

A straightforward computation yields,

$$\ell^i = \frac{1-\eta}{\frac{p^d}{p} - (1-\eta)}, \tag{19}$$

and thus changes in leverage are determined by changes in the ratio p/p^d. Holding effort (and hence g) constant, an increase in p always results in a increase in p/p^d and thus an increase in leverage ℓ^i. The next proposition shows that changes in effort do not overturn this result.

Proposition 4 *(Leverage) If N is small enough and $(\widetilde{p}; \widetilde{e})$ is a strict CIM equilibrium for parameter values (K, N, α), then $\mathcal{L}$ is an increasing function of K.*

Proof: The result follows directly from Proposition A.7 in the Appendix. □

Since intermediary borrowing is constrained by the market value of its collateral it is obvious that borrowing increases with K. But, the proposition states a much stronger result: As aggregate savings increase, *leverage*, that is the amount of debt per unit of capital, also increases. In other words, intermediary borrowing increases more than proportionally with K. This implication is tested in Adrian and Shin (2010), who stress that there is: "a strongly positive relationship between changes in leverage and changes in the balance sheet size," [Adrian and Shin, page 419, 2010, and as shown in Figure 3]. The following is an immediate consequence of the expression for the share of funds at the disposal of (informed) intermediaries (see equation (11)).

Corollary 5 *(Capital and knowledge) Under the conditions of Proposition 4, the share of funds deployed by informed investors, f^i, is an increasing function of K.*

This simple result has an interesting implication: Having a larger fraction of funds deployed by informed investors does not always result in better origination incentives as shown in Panel A of Figure 6.[18]

In addition, the greater leverage may be accompanied by deterioration of the asset side of intermediaries' balance sheets. This effect is far from obvious. After all, intermediaries

which is obviously counterfactual.

[18]This implication holds even when informed investors do not make mistakes ($\alpha = 0$).

are able to identify high quality assets through their informational advantage. Yet, the fraction $g = \Pr(x_h|\sigma_h)$ of assets paying off x_h acquired by intermediaries (see (3)), is an increasing function of e. In other words, as origination standards deteriorate, the proportion of non-performing assets on the balance sheet of intermediaries also increases, since the fraction of non-performing assets with a signal σ_h bought in the OTC market increases.

Panel A in Figure 7 shows that the asset side of intermediaries' balance sheet mirrors the non-monotonicity of origination effort as a function of K. Panel B illustrates Proposition 4: Intermediaries' leverage increase as aggregate savings K increase. Thus, for a while there is a positive correlation between leverage and asset quality on intermediaries' balance sheets, but once aggregate capital is large enough, eroding origination incentives produce a negative correlation. Finally, intermediary leverage is higher the less precise is their information about asset quality (the higher is α). Other things equal, an increase in α lowers the expected payoff of acquired assets, and therefore the price intermediaries are willing to pay, p^d. This, in turn, results in a narrower spread, $p^d - p$, which economizes the equity capital that intermediaries need to hold, thereby increasing their leverage, $\ell^i = D^i/K$.

5.2 Financial fragility

We show next that financial fragility increases with the magnitude of the savings glut. In particular we address the following question: What is the size of an unforeseen loss that results in an equity shortfall L, and how does the size of the shock depend on K? If the size of this shock *decreases* with K then we conclude that savings gluts increase the fragility of financial intermediaries.

We analyze how a loss affects bank balance sheets in a simple "steady state" version of our model. In this steady-state reformulation time is continuous and originators supply one asset per unit of time. An asset originated at time t yields a return at time $t + 1$. Moreover, investors maintain a constant level of capital K over time and pay out incremental earnings. In particular, informed intermediaries pay out dividends so as to maintain a constant leverage over time.[19] We simplify the analysis by setting $\alpha = 0$ since mistakes in identifying asset

[19]The evidence on banks' payout policy suggests that, unless they are barred from doing so by regulators,

quality play no essential role in the argument linking greater financial fragility to savings gluts.[20] Finally, we restrict attention to the parameter subset where K and N are such that a CIM equilibrium obtains and the assumptions on Propositions 2 and 4 hold.

In this reformulation of the model we introduce the following unforeseen contingency: with some small probability a fraction $\epsilon\left(K\right)$ of banks' assets are revealed to be *non-performing*. We then ask what shock size $\epsilon\left(K\right)$ wipes out banks' equity buffer and how $\epsilon\left(K\right)$ depends on K? The minimum loss that entirely wipes out a bank's equity, $\underline{\epsilon}\left(K\right)$, is given by the solution to:

$$(1-\eta)\,pq-\left(1-\underline{\epsilon}\left(K\right)\right)p^{d}q=0. \tag{20}$$

That is, $\underline{\epsilon}\left(K\right)$ is the minimum fraction of non-performing loans (NPLs) yielding no return that entirely wipes out bank equity. The central result linking greater bank fragility to savings gluts is stated in the following corollary.

Corollary 6 $\underline{\epsilon}\left(K\right)$ *is decreasing in* K.

Proof: See the Appendix. □

The corollary states that the larger is the savings glut the lower is the size of the loss needed to wipe out the equity of an informed financial intermediary. This key observation is strengthened by the next corollary that links the probability of lenders incurring a loss of a given size to savings gluts. It states that if we take the distribution of unforeseen losses as given then an increase in K increases the probability of lenders facing a loss of a given size $\mathsf{L}>0$, and that the increase in probability is larger for larger L. Let $\epsilon\left(K,\mathsf{L}\right)$ be defined as the solution to:

$$(1-\eta)\,pq-\left(1-\epsilon\left(K,\mathsf{L}\right)\right)p^{d}q=\mathsf{L} \tag{21}$$

and note that $\underline{\epsilon}\left(K\right)=\epsilon\left(K,0\right)$. Then the following corollary can be established.

banks do effectively maintain a constant leverage ratio and pay out any excess realized earnings (see e.g. Bebchuck, Cohen and Spamann, 2010, and Shin, 2012). Indeed, banks' reluctance to cut dividends in an impending crisis to shore up their equity capital buffer is the reason why the *capital conservation buffer* requirement was introduced under Basel III.

[20] In section 6.4 we investigate the effect on financial stability of asset quality misperceptions related to α. In addition, in our numerical examples Corollary 6 obtains even for $\alpha>0$.

Corollary 7 *(Financial fragility)* $\epsilon(K, \mathsf{L})$ *satisfies*

$$\frac{\partial^2 \epsilon}{\partial K \partial \mathsf{L}} < 0,$$

so that

$$\frac{\partial \epsilon}{\partial K}(K, \mathsf{L}') < \frac{\partial \epsilon}{\partial K}(K, \mathsf{L}) < \frac{\partial \epsilon}{\partial K}(K) < 0,$$

for $\mathsf{L}' > \mathsf{L} > 0$.

Proof: See the Appendix. $\square$

Corollary 7 states that an increase in K increases the probability of lenders facing a loss L and that the increase in probability is larger when we consider larger losses. It is in this sense that savings gluts are accompanied by increasing financial fragility.

Our observations on the link between savings gluts and financial fragility rely on the unforeseen nature of a loss. It is the realization of unforeseen losses that generally trigger a crisis. The crisis of 2007-2009 is no exception. Indeed, there is substantial evidence that the subprime MBS losses that materialized in 2007 and 2008 were not foreseen by market participants, nor were the across-the-board asset price declines that could be caused by such losses.[21] As Gerardi, Lehnert, Sherlund and Willen (2008, page 70) argue, "[H]ad investors known the future trajectory of home prices, they would have predicted large increases in delinquency and default and losses on subprime MBSs roughly consistent with what has occurred." The widespread adoption of stress testing by bank supervisors after the financial crisis is due precisely to the realization of the need to determine the solvency of intermediaries on the basis of unforeseen large shocks independently of their probability of occurrence.[22]

[21] A striking example of such lack of foresight is recounted by Timothy Geithner in his book on the crisis: "In December 2004, the New York Fed's research division produced a paper titled "Are Home Prices the Next Bubble?" that concluded they weren't. It conceded the possibility of corrections along the coasts, but noted "regional price declines in the past have not had devastating effects on the broader economy." Six months later, in a presentation to the FOMC, Fed economists projected that even if there were a 20 percent nationwide decline in housing prices, it would cause only about half the economic damage of the bursting of the dot-com bubble." [in Timothy Geithner, *Stress Test: Reflections on Financial Crises,* 2014, pp 167]

[22] Interestingly, and as documented in Schuermann (2014, page 721), Brian Peters, at the time head of risk in bank supervision at the New York Fed, noticed at an industry conference in March of 2007 "that no firm had a

5.3 Additional Comparative Statics

Our exclusive focus so far has been on comparative statics with respect to K. But our model yields two other important comparative statics results with respect to N and η, which we characterize below.

5.3.1 More capital in the hands of uninformed investors

How are asset prices, leverage, and origination incentives affected by changes in the fraction of capital held by uninformed investors, M? We can address this question by looking at the comparative statics with respect to N, the measure of informed traders: by decreasing N we increase the share of the endowment in the hands of uninformed investors. The next proposition establishes the key comparative statics results with respect to N.

Proposition 8 *Let N_2 and K be such that $N_2 K < \bar{\epsilon}$ and $N_1 < N_2$. Suppose that continuous functions $p(N)$ and $e(N)$ exist for $N \in [N_1, N_2)$ such that ($p(N); e(N)$) is a strict CIM equilibrium with parameters (K, N, α). Then $p(N)$ is decreasing, $e(N)$ is increasing in N and $\mathcal{L}(N)$ is decreasing in N.*

Proof: Follows from Lemma A.2. □

In words, when the fraction of uninformed capital increases asset prices on the exchange increase and leverage of financial intermediaries also increases, while origination incentives decrease. Overall, as one would expect based on our previous analysis, the financial system becomes more fragile when uninformed savings increase.

5.3.2 Haircuts and incentives for good origination

Could excessively low repo haircuts have been a contributing factor in worsening the fragility of financial intermediaries before the crisis? To address this question we characterize the comparative statics of the CIM equilibrium with respect to η. As the proposition below establishes,

---

fully developed program of integrated stress testing that captured all major financial risks on a firm-wide basis." As Schuermann (2014) notes the key ingredient in a stress tests is precisely how severe (and for how long) should the scenario considered be.

a lower haircut allows informed intermediaries to borrow more, resulting in lower asset prices on the exchange and higher origination incentives.

Proposition 9 *Let N and K be such that $NK < \bar{\epsilon}$, and suppose that continuous functions $p(\eta)$ and $e(\eta)$ exist over the interval $\eta \in [\eta_1, \eta_2]$ that correspond to strict CIM equilibria for parameter values (K, N, α, η). Then $p(\eta)$ is increasing and $e(\eta)$ is decreasing in η.*

Proof: Follows directly from Lemma A.3. □

In other words, an increase in the haircut η limits the amount of liquidity flowing to informed intermediaries through the repo market. More liquidity gets channelled to the exchange, resulting in higher asset prices $p(\eta)$. Thus, an unintended consequence of a policy seeking to strengthen financial stability by imposing higher haircuts η for secured loans is to undermine origination incentives and thereby to adversely affect the quality of assets distributed in financial markets.

6 Robustness

6.1 Endogenous origination volume and distribution

For simplicity we have assumed that the total volume of originated assets is fixed and is price inelastic. But, when a savings glut pushes up asset prices, isn't it natural to expect a supply response and an increase asset origination? A striking example of such a response was the large increase in the float of dot.com IPOs following the expiration of lock up provisions, which contributed to the bursting of the dot.com bubble (Ofek and Richardson, 2003). Similarly, the rise in real estate prices up to 2007 gave rise to a construction and securitization boom. But this increased origination volume was not sufficient to quell the market's thirst for new mortgage-backed securities, so that it was further augmented at the peak of the cycle by a larger and larger volume of synthetically created assets.

Our model can be extended to allow for a price-elastic origination volume and our results are robust as long as the supply of assets is not too price elastic. Indeed, if origination volume were perfectly price elastic there could not be a savings glut. A particularly interesting way of introducing a price-elastic origination volume is to let originators choose whether to distribute or retain the asset they originated until maturity and to allow for different demand for liquidity across originators. Then, in equilibrium, for any given K, a fraction $\lambda(p)$ of originators would prefer to distribute their asset, and the fraction $(1-\lambda(p))$ to hold on to their asset until maturity. The fraction $\lambda(p)$ would be increasing in p, thus giving rise to a price-elastic volume of distributed assets. In this more general formulation of the model, a savings glut would doubly undermine origination incentives. Not only would spread compression reduce market incentives of originators who intend to distribute their assets, but also a smaller fraction of originators would have skin in the game by retaining their assets to maturity. This more general version of the model could also account for the deterioration of origination standards in the run-up to the crisis of 2007-09 that was caused by lower skin-in-the-game incentives.[23]

6.2 Motivating the Collateral Constraint

Our analysis builds on the assumption that intermediary borrowing is limited by constraint (10). One way of motivating this constraint is to introduce the possibility that the lender may change her mind in period 1 and refuse to roll over the loan in period 1. That is, suppose that period 1 is broken into two sub-periods, 1_- and 1_+. In sub-period 1_- assets are originated, borrowing and lending occurs, and the OTC market is operating. In period 1_+ only the exchange is open for trade and all other markets (OTC and Repo) are closed. However, any loan agreed to in period 1_- is callable in period 1_+, or in other words, the repo loan extended in period 1_- is open to a roll-over decision in period 1_+. Suppose, in addition, that a small set of uninformed agents learns in period 1_+ that they must consume immediately in this period and for this reason need to call their loans. The intermediaries who borrowed from them must then surrender their collateral to the lender, who subsequently pays an "iceberg cost" η before selling the collateral in the only market which is open - the exchange. If lenders understand that they may face this liquidity risk in period 1_+, and want to make sure that they can consume

[23] See on this issue, for instance, Bord and Santos (2012), Keys, Mukherjee, Seru and Vig (2010) and Purnanandam (2011). There are of course additional issues associated with distribution such as active misrepresentation by originators as in Piskorski, Seru and Witkin (2015).

at least their endowment in period 1_+, then they will impose constraint (10) on the borrower. In other words, the collateral constraint (10) is determined by the rollover risk lenders face in period 1_+ and is not driven by concerns about the creditworthiness of the borrower.

Alternatively, the collateral requirement could be driven by credit risk and the concern that the borrower might engage in a strategic default. More precisely, consider the situation where the borrower could at any time abscond with a fraction η of the assets after the loan has been agreed, but before date 2 cash flows are realized. In this case the borrower expects to obtain an expected payoff of $\eta q g x_h$. To prevent such a move the lender must make sure that the borrower is better off realizing the returns on his portfolio in period 2 and repaying his total debt obligation D. In other words, the following incentive constraint must hold:

$$qgx_h - D \geq \eta q g x_h$$

or,

$$D \leq (1 - \eta)\, qgx_h. \tag{22}$$

Moreover, to obtain a finite D it must be the case that in equilibrium, $p^d > (1 - \eta)\, gx_h$. Preliminary analysis shows that the main qualitative results of our model are robust to the replacement of leverage constraint (10) with (22).

There is no obvious way to derive a constraint analogous to (10) but with p^d replacing p. Often, accounting conventions define the haircut as a function of the book-price of an asset. This, however, does not insure that modeling the haircut as a constant proportion of book-price is adequate. Each of the two economic motives for a haircut that we presented would imply a haircut as a proportion of p^d, but this proportion would vary with market conditions and p.

6.3 Variation in risk premia or cash-in-the-market pricing?

We have modeled the savings glut and its effect on asset prices, spreads and origination incentives as an aggregate liquidity asset pricing phenomenon captured in simple terms by cash-in-the-market pricing. But, an alternative account of the effects of an increased pool of savings is possible based on the more classical notions of discount rates and "risk adjusted capital." Indeed, there is ample evidence coming from the asset pricing literature that discount rates are countercyclical, high during troughs and low at the peak of the cycle. Under this alternative interpretation, asset prices on the exchange are affected by both the fundamental quality of assets and the risk appetite of uninformed investors. Indeed, Shin (2012) develops a model in which bondholders' risk appetite is pro-cyclical, inducing banks to lever up more in booms. In our model, when the discount rate of uninformed investors moves counter-cyclically, so will the excess return $r^x - 1$. A reduction in discount rates then leads to higher asset prices on the exchange, smaller price spreads $p^d - p$ and consequently weaker origination incentives. It also results in higher leverage of financial intermediaries, yielding the pro-cyclical leverage patterns identified by Adrian and Shin (2010). This alternative account also matches the evidence that leverage is highest, and the worst assets are originated at the peak of the cycle, which leads to maximal financial fragility just when the economy is booming.

6.4 Risk management and market misperceptions

As shown in Corollary 6, savings gluts result in increased financial fragility, to the extent that bank balance sheets are more sensitive to unforeseen losses in the middle of a savings glut. In this section we examine the effects of asset value misperceptions, either by intermediaries or the market on bank fragility. Specifically, suppose that informed investors have a screening technology with $\alpha = .2$, but that a single individual informed investor mistakenly believes that her $\alpha = 0$. This optimistic investor takes as given the equilibrium price on the exchange p^* and interest rate r^*. The quantity of assets bought and leverage of the optimistic investor are then:

$$\widetilde{q} = \frac{K}{\kappa \left(x_h - p^*\right) + \eta p^*} \qquad \text{and} \qquad \widetilde{D} = \frac{\left(1-\eta\right) p^* K}{\kappa \left(x_h - p^*\right) + \eta p^*},$$

which should be compared with (A.14) and (A.15).

A first observation is that $q^* > \widetilde{q}$, as the optimistic investor bids more for assets in the OTC market than the other informed intermediaries, who have the correct assessment of α. Indeed the optimistic investor bids

$$\widetilde{p}^d = \kappa x_h + \left(1-\kappa\right) p^*,$$

which is higher than the price offered by the other intermediaries (see (4)). As a result $D^* >$

$\widetilde{D}$, and thus $\mathcal{L}^* > \widetilde{\mathcal{L}}$. In other words, book leverage of the optimistic intermediary is lower than that of intermediaries who have an accurate estimate of the precision of the signal.

The optimistic intermediary believes his expected equity in period 2 is

$$E_o^2 = \widetilde{q}x_h - r^*\widetilde{D},$$

whereas the true level of capital is

$$E_{true}^2 = \widetilde{q}g^*x_h - r^*\widetilde{D} < E_o^2.$$

Second, note that the optimistic investor's true average level of capital is lower than the average level of capital of investors with an accurate assessment of α, E^2, which is

$$E^2 = q^*g^*x_h - r^*D^* = \left(\frac{\kappa\left(x_h - p^*\right) + \eta p^*}{\kappa\left(g^*x_h - p^*\right) + \eta p^*}\right)E_{true}^2 > E_{true}^2.$$

In sum, optimistic investors appear to take on less leverage, but in truth have less equity capital at $\tau = 2$ than investors with an accurate estimate of α. This simple example remarkably illustrates how a bank supervisor focusing on book leverage would be misled to believe that the optimistic intermediary is the safer one. This is illustrated in Panel A of Figure 8 where the top line represents the equity capital under the wrong beliefs and the bottom line represents the true equity capital. Leverage, simply put, does not equal risk exposure.

As we have already noted, during the credit bubble many financial intermediaries, although fully cognizant of the link between home price appreciation (HPA) and MBS values, did not consider possibility of a sharp nationwide negative scenario of HPA such as the one that actually came to pass. Analysts most extreme scenarios at a major investment bank did not go beyond a 5% correction in housing prices, far less than what was experienced (see Gerardi, Lehnert, Sherlund and Willen, 2008). We could also model a situation of aggregate optimism by assuming that *all* informed intermediaries believe that $\alpha_o = 0$ even though $\alpha > 0$. Because all intermediaries are excessively optimistic about their signal, they bid aggressively for "good" assets beyond what their true expected payoff warrants. As above, book leverage would be relatively low in this situation and the average period 2 equity capital would be lower than anticipated. Panel B of Figure 8 shows the level of capital under the wrong beliefs ($\alpha_o = 0$) and the actual level of capital under the true value of α, for $\alpha = .4$.

Of course, some financial institutions, such as Lehman Brothers, Merrill Lynch and Bear Stearns, were more aggressive than others in expanding their balance sheets, to the point where they ultimately failed. One intriguing possibility in terms of our model is that these institutions had underestimated their own α. They were overly relying on information on their past returns on the assets they purchased to assess their own ability to control risks. If they were unaware of the deterioration in origination incentives, as seems plausible, they may have unwittingly added a larger proportion of non-performing assets to their balance sheets than they could handle (see Foote, Gerardi and Willen, 2012). When larger losses than predicted by their own risk models materialized and these financial institutions realized that the proportion of good assets on their balance sheet was lower than estimated it was too late. To capture this behavior, we could extend the model to allow for the possibility of an endogenous α. We could add to the model that informed intermediaries are engaged in costly endogenous screening of assets and that they determine their screening effort based on the past history of non-performing assets they acquired. Then, as origination standards improve ($e(K)$ increases) intermediaries would respond by cutting their screening effort, which, in turn, could amplify their financial fragility at the peak of the savings glut and lead to their insolvency.

6.5 Double Margin

Our key single-peakedness of effort result (Proposition 3) is driven by the interplay of two countervailing forces. A first force is that when savings K increase the spread $p^d - p$ narrows, undermining incentives to originate good assets. The second force is that when savings K increase the price p increases and the return r^x decreases, relaxing the borrowing constraint of informed intermediaries. As a result, more informed capital can be deployed to purchase good assets, increasing origination incentives. The basic economic logic behind Proposition 3 is that at low levels of savings the latter effect dominates, but at higher levels of savings the former effect dominates, giving rise to our single-peakedness result.

It turns out that this result depends on the presence of a double margin, the $p^d - p$ margin and the collateral haircut η. When there is no haircut ($\eta = 0$) it turns out that

$$\frac{\partial e}{\partial K} = \frac{N}{\psi_e + e\psi_{ee}} > 0.$$

This can be immediately observed by setting $\eta = 0$ in equations A.17 to A.19, and computing $\frac{\partial e}{\partial K}$ by totally differentiating with respect to e and K in equation A.17.[24] The fact that $\frac{\partial e}{\partial K}$ always has a positive sign when $\eta = 0$ is not entirely surprising. In effect, a positive haircut acts like a speed-bump is reducing the flow of capital from the uninformed exchange to the informed dealer market. This speed-bump reduces the strength of the countervailing force on incentives through a greater share of informed capital. It is only when the speed-bump is entirely removed that this countervailing force is always dominant. Paradoxically, our analysis suggests that financial fragility can by lessened by lowering the haircut on repo loans. A reduction in η increases the flow of savings to informed intermediaries and results in a lower uninformed price p, thereby improving origination incentives.

7 Conclusion

We have developed a simple yet rich model of the origins of the financial crisis, in which asset prices, spreads, origination incentives and leverage are driven by aggregate liquidity conditions. The notion of cash-in-the-market pricing, first introduced by Allen and Gale (1998), is central for tractability and for a clear analysis of potentially complex effects. The other central building block is the dual representation of financial markets as in Bolton, Santos and Scheinkman (2016), with an organized exchange where uninformed investors trade, and an OTC market, where informed traders cream-skim the best assets. The third essential element is a repo market where agents can borrow against collateral.

A key economic mechanism in our analysis centers on origination incentives, which arise from the ability of informed investors to identify the better assets, and their offer of a price improvement relative to the exchange for those higher quality assets. There are two effects of aggregate liquidity on origination incentives. First, the equilibrium price improvement for higher quality assets narrows as liquidity surges, which weakens origination incentives. Second, as aggregate liquidity increases some of it will find its way to the balance sheets of informed investors, who then can buy more high-quality assets, which is good for origination

[24] Thus to explain the deterioration of standards while maintaining the assumption that $\eta = 0$ one would need to assume that the capital increase is biased towards uniformed investors (see Proposition 8 above).

incentives. A central result in our model is that the latter effect dominates when the level of aggregate liquidity is low, whereas the former effect dominates when it is high. An important implication of this result is that origination incentives eventually deteriorate with increasing aggregate liquidity.

Another basic result is that the balance-sheets of financial intermediaries become more fragile as liquidity increases. There are two reasons why financial fragility increases. First, unless the screening abilities of informed investors are perfect, the fraction of good assets in intermediaries' balance sheets is an increasing function of the fraction of good assets originated. Thus, as origination standards deteriorate and the fraction of originated high-quality assets falls, the balance sheets of informed intermediaries necessarily absorb an increasing fraction of non-performing assets. Second, a further fragility is induced because of leverage. As liquidity becomes abundant, there are worse assets on the intermediaries balance sheets with more leverage. Our model thus offers a particularly simple account of the origins of the Great Recession, based on a straightforward economic mechanism.

References

Adrian, Tobias and Hyun Shin (2010) "Money, Liquidity and Monetary Policy", *American Economic Review, Vol. 99, pp. 600-605.*

Adrian, Tobias and Hyun Shin (2010) "Liquidity and Leverage", *Journal of Financial Intermediation*, 19, 418-437.

Aliber, Robert (2011) "Financial turbulence and international investment", *BIS working paper No* 58.

Allen, Franklin and Douglas Gale (1998) "Optimal Financial Crisis," *Journal of Finance*, 53, 1245-1284.

Bebchuck, Lucian A., Alma Cohen, and Holger Spamann (2010) "The Wages of Failure: Executive Compensation at Bear Sterns and Lehmann 2000-2008," Yale Journal on Regulation, vol. 27, issue 2.

Bernanke, Ben (2005) "The Global Saving Glut and the U.S. Current Account Deficit," Remarks by Governor Ben S. Bernanke at the Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia.

Bernanke, Ben S., Carol Bertaut, Laurie Pounder DeMarco, and Steven Kamin (2011), "International Capital Flows and the Returns to Safe Assets in the United States, 2003-2007,"International Finance Discussion Paper 1014, Board of Governors of the Federal Reserve System.

Bernanke, Ben and Mark Gertler (1989), "Agency Costs, Net Worth and Business Fluctuations," *American Economic Review, Vol. 79, pp. 14-31.*

Binmore, Ken, Ariel Rubinstein and Asher Wolinsky (1986) "The Nash Bargaining Solution in Economic Modeling,"*Rand Journal of Economics*, XVII, 176-88.

Boissay, Frédéric, Fabrice Collard and Frank Smets (2016) "Booms and Banking Crises," *Journal of Political Economy.*

Bolton, Patrick, Tano Santos, and Jose Scheinkman (2011), “Outside and Inside Liquidity”, *Quarterly Journal of Economics*, 126(1): 259-321.

Bolton, Patrick, Tano Santos, and Jose Scheinkman (2012), “Shadow Finance” in *Rethinking the Financial Crisis*, Alan S. Blinder, Andrew W. Lo, and Robert M. Solow (eds.), Russell Sage Foundation and The Century Foundation, New York

Bolton, Patrick, Tano Santos, and Jose Scheinkman (2016), “Cream skimming in financial markets, ”*Journal of Finance*, April, vol. 71, no.2, 709-736.

Bord, Vitaly M. and Joao A. C. Santos (2012) “The rise of the Originate-to-Distribute Model and the role of Banks in Financial Intermediation,” FRBNY economic policy Review, July, pp. 21-34.

Borio, Claudio and Piti Disyatat (2011) “Global imbalances and the financial crisis: Link or no link?,” BIS Working Papers, no. 346, May.

Caballero, Ricardo, Emmanuel Farhi and Pierrel-Olivier-Gourinchas (2008) “An equilibrium model of “Global Imbalances” and Low Interest Rates,” *American Economic Review,* 98:1, 358-393.

Calvo, Guillermo (2012), “On Capital Inflows, Liquidity and Bubbles,”mimeo, Columbia University

Cochrane, John (2008), “The Dog That Did Not Bark: A Defense of Return Predictability,”*Review of Financial Studies,* vol. 21, no. 4, pp. 1533-1575

Copeland, Adam, Darrell Duffie, Antoine Martin, and Susan McLaughlin (2012) “Key Mechanics of The U.S. Tri-Party Repo Market”, FRBNY Economic Policy Review

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong (2014) “Wall Street and the Housing Bubble,” *American Economic Review,* 104 (9), 2797-2829.

Coval, Joshua, Jacob Jurek, and Erik Stafford, (2009) “Economic Catastrophe Bonds,” American Economic Review 99(3): 628-666.

Diamond, Douglas and Raghuram Rajan (2001) "Liquidity Risk, liquidity Creation and Financial Fragility: A Theory of Banking," *Journal of Political Economy*, vol. 109, no. 2, 287-327.

Gennaioli, Nicola, Andrei Shleifer and Robert W. Vishny (2012) "Neglected risks, financial innovation, and financial fragility," *Journal of Financial Economics,* 104, 452-468.

Gennaioli, Nicola, Andrei Shleifer and Robert W. Vishny (2013)"A Model of Shadow Banking," *The Journal of Finance,* vol. 68 (4), August, 1331-1363.

Gerardi, Kristopher, Andreas Lehnert, Shane M. Sherlund and Paul Willen (2008), "Making Sense of the Subprime Crisis," *Brookings Papers on Economic Activity*, Fall, pp. 69-145.

Gorton, Gary and Andrew Metrick (2009) "Securitized Banking and the Run on Repo", NBER Working Paper 15223

Gourinchas, Pierre-Olivier (2012) "Global imbalances and Global Liquidity," in Asia's Role in the Post-Crisis Global Economy, Asian Economic Policy Conference, R. Glick and M. Spiegel eds., Federal Reserve Bank of San Francisco, February.

Gourinchas, Pierre-Olivier, Rodrigo Valdés and Oscar Landerretche (2001) "Lending Booms: Latin America and the World," *Economía* 1:47-99.

Holmstron, Bent and Jean Tirole (1997) "Financial Intermediation, loanable Funds and the Real Sector," *Quarterly Journal of Economics,* 112, 1239-1283.

Jordà, Òscar, Moritz Schularick and Alan M. Taylor (2011) "Financial crises, Credit Booms and External Imbalances: 140 Years of Lessons," IMF Economic Review, vol. 59, no. 2.

Karabarbounis, Loukas and Brent Neiman (2012) "Declining labor shares and the global rise of corporate savings," NBER WP18154.

Keys, Ben, Tanmoy Mukherjee, Amit Seru and Vikrant Vig (2010), "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *Quarterly Journal of Economics*

Keys, Benjamin, Tomasz Piskorski, Amit Seru and Vikrant Vig (2013), "Mortgage financing in housing boom and bust," in E. Glaeser and T. Sinai *Housing and the Financial Crisis*, NBER, The University of Chicago Press, Chicago and London.

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

Kindleberger, Charles and Robert Aliber (2005), *Manias, Panics, and Crashes: A History of Financial Crises*

Krishnamurty, Arvind and Tyler Muir (2016) "Credit Spreads and the Severity of financial Crises," manuscript, Stanford University.

Loeys, Jan, David Mackie, Paul Meggyesi and Nikolaos Panigirtzoglou (2005) "Corporates are driving the global savings glut," Global Issues, June 24th, JPMorgan.

Martinez-Miera, David and Rafael Repullo (2015) "Search for Yield," manuscript, CEMFI.

Neuhann, Daniel (2016) "Macroeconomic Effects of Secondary Market Trading", Working Paper, University of Pennsylvania.

Ofek, Eli and Matthew Richardson (2003) "DotCom Mania: The Rise and Fall of Internet Stock Prices," Journal of Finance, vol. LVIII, no. 3, June, pp. 1113-1137.

Piskorski, Tomasz, Amit Seru, and James Witkin (2015) "Asset quality misrepresentation by financial intermediaries: Evidence from RMBS market," *Journal of Finance*.

Pozsar, Zoltan, (2011), "Institutional Cash Pools and the Trifflin Dilemma of the U.S. Banking System", IMF Working Paper 11/190

Purnanandam, Amiyatosh (2011) "Originate-to-Distribute Model and the Subprime Mortgage Crisis," *Review of Financial Studies*, v. 24, no. 6, pp. 1881-1915.

Santos, Tano (2017) “Antes del Diluvio: The Spanish banking system in the first decade of the euro,” in After the Flood: How the Great Recession Changed Economic Thought, Edward L. Glaeser, Tano Santos and E. Glen Weyl, eds. The University of Chicago Press.

Schuermann, Til (2014) “Stress Testing Banks,” International Journal of Forecasting, vol. 30, 717-728.

Shin, Hyun Song (2011) “Global Savings Glut or Global Banking Glut,” Vox, December 20th 2011.

Shin, Hyun Song (2012) “Global Banking Glut and Loan Risk Premium,” *IMF Economic Review*, 60, 155-192.

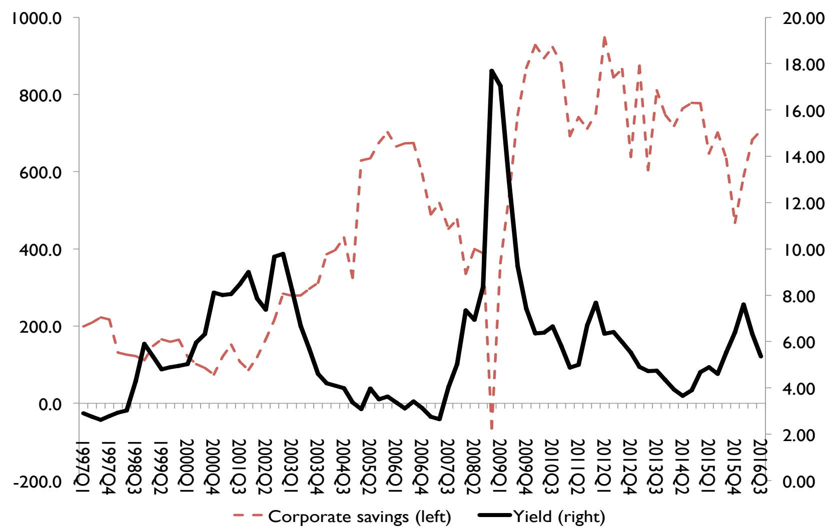


Figure 1: Corporate savings in billions of US$ (left axis) and BofA Merrill Lynch US High Yield Option-Adjusted Spread in % (right axis). Quarterly:1997Q1-2017Q1. Not seasonally adjusted. Data Source: Federal Reserve Bank of St. Louis

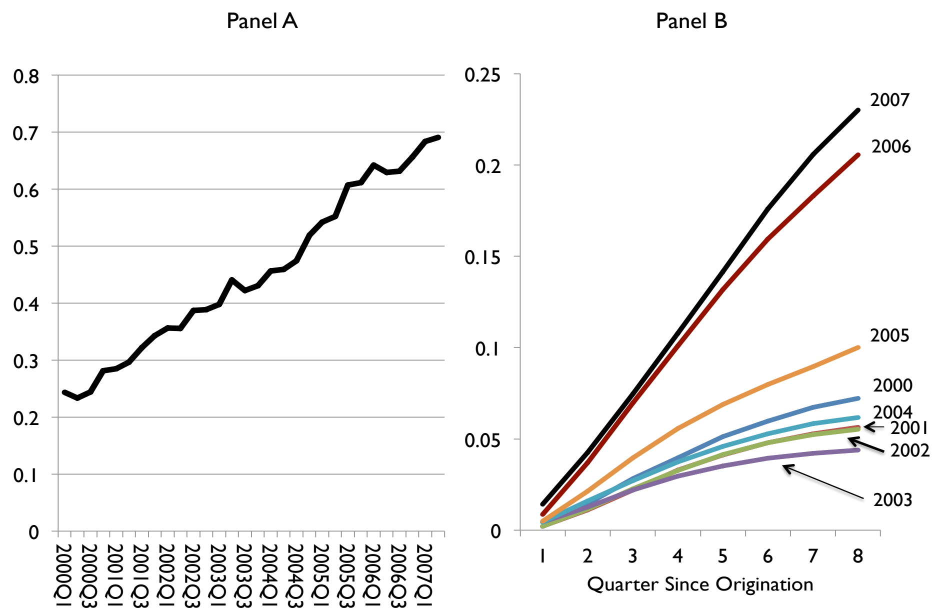


Figure 2: Panel A: Fraction of loans that were originated with no/low/reduced documentation among non-agency mortgages with known documentation status in the Loan Performance (LP) database. Panel B: Average cumulative default paths for the non-agency securitized loans by year of their origination. The cumulative delinquency rates are of all privately securitized residential mortgages calculated from loan level, monthly Corelogic-Blackbox data. A loan is defined as delinquent if it ever becomes 60 days past due and is considered delinquent thereafter. The graph is separated by year of mortgage origination, the y-axis plots cumulative delinquency rates in each quarter following origination which is depicted on the x-axis. Data source: Corelogic

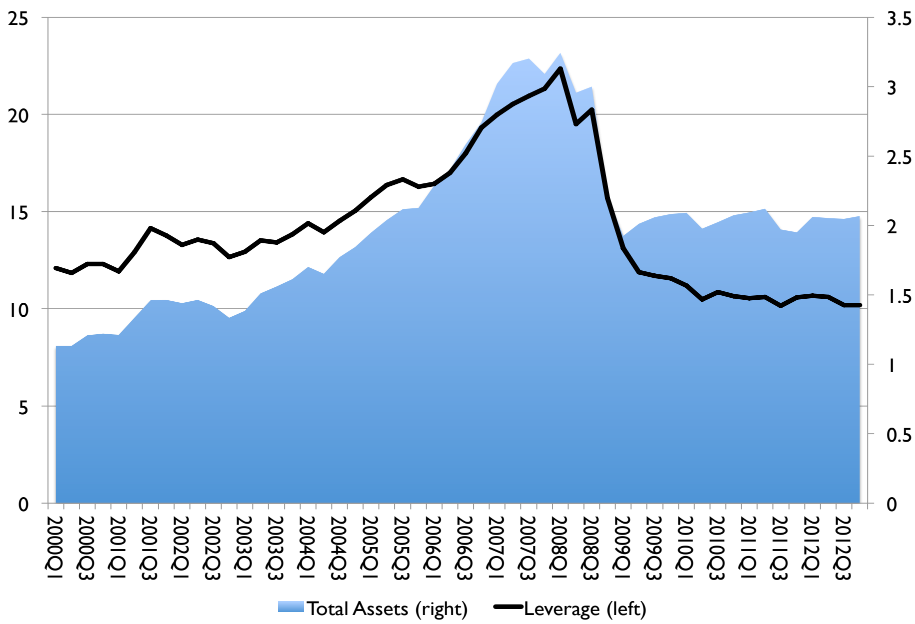


Figure 3: Broker dealers: Total assets in billions of dollars and leverage (total assets to book equity). Quarterly: 2000Q1-2007Q2 Data source: Federal Reserve Board

Figure 4: Flows and stocks in a strict CIM equilibrium

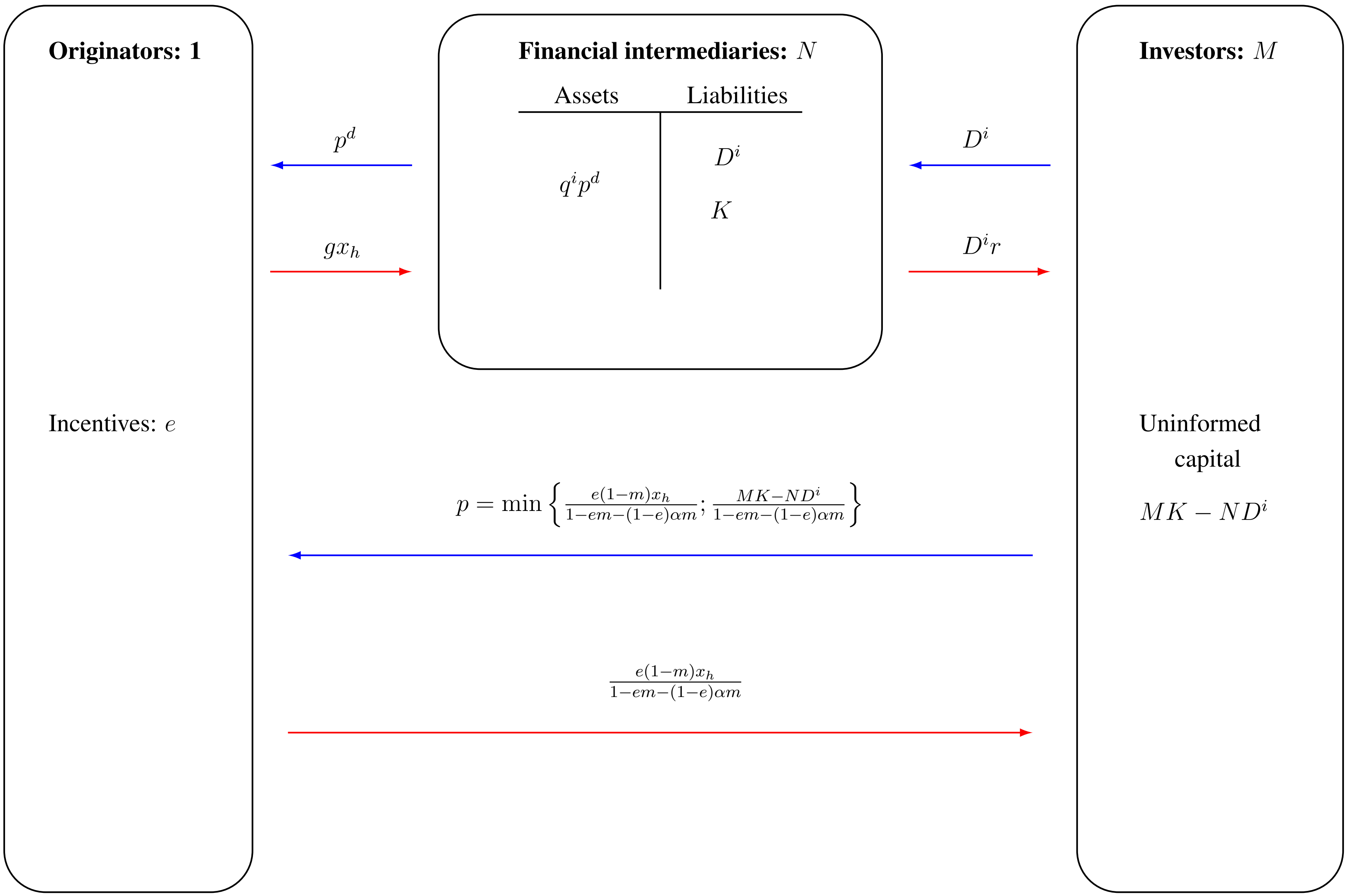


Figure 5: Panel A: Prices in the exchange p and expected payoff (dashed line) as a function of capital, K,. Panel B: Expected rate of return in the opaque market, R, expected rate of return in the exchange $r^x(= r)$; the horizontal line is set at one. In this example, $\psi(e) = \theta\frac{e^2}{2}$ with $\theta = .25$. In addition, $\kappa = .15$, $\eta = .5$, $M = .75$ and $x_h = 5$.

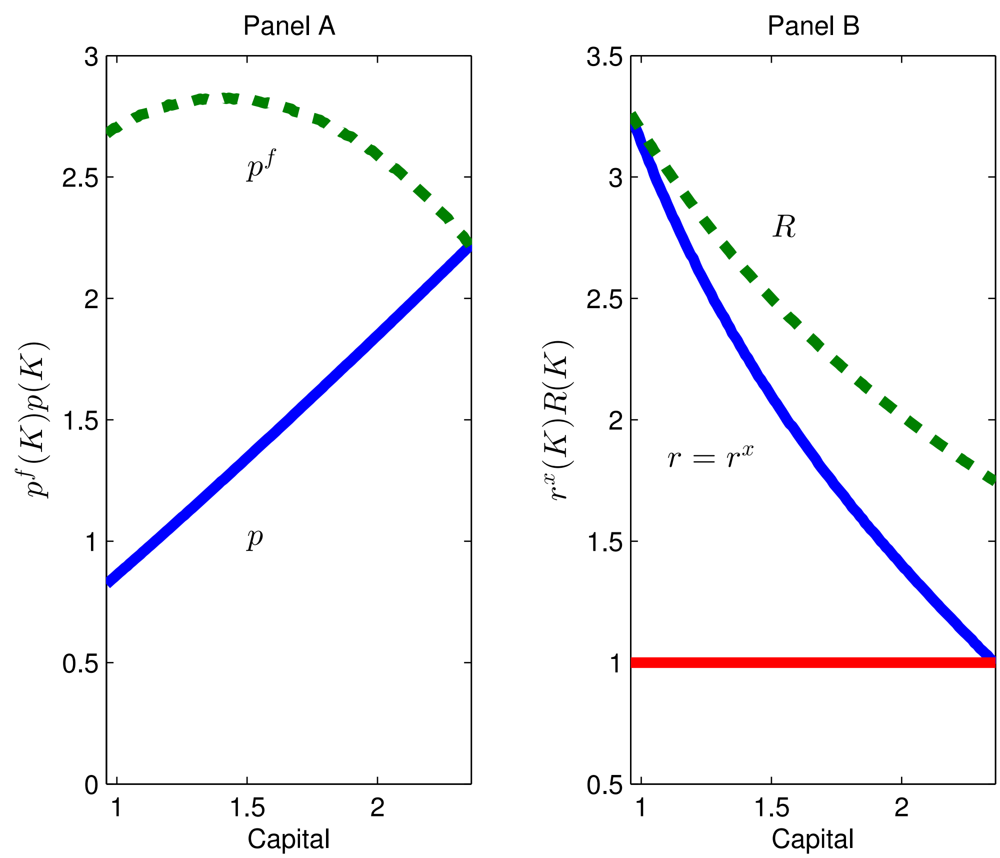


Figure 6: Panel A: Origination effort e as a function of capital, K, for the cases $\alpha = 0$ (dashed line, in green) and $\alpha = .2$ (continuous line, in blue). Panel B: $x_h p$, as a function of capital K (continuous line, in blue, left axis) and matching probability m as a function of capital, K, (dashed line, in green, right axis). In this example, $\psi(e) = \theta \frac{e^2}{2}$ with $\theta = .25$. In addition, $\kappa = .15$, $\eta = .5$, $M = .75$ and $x_h = 5$.

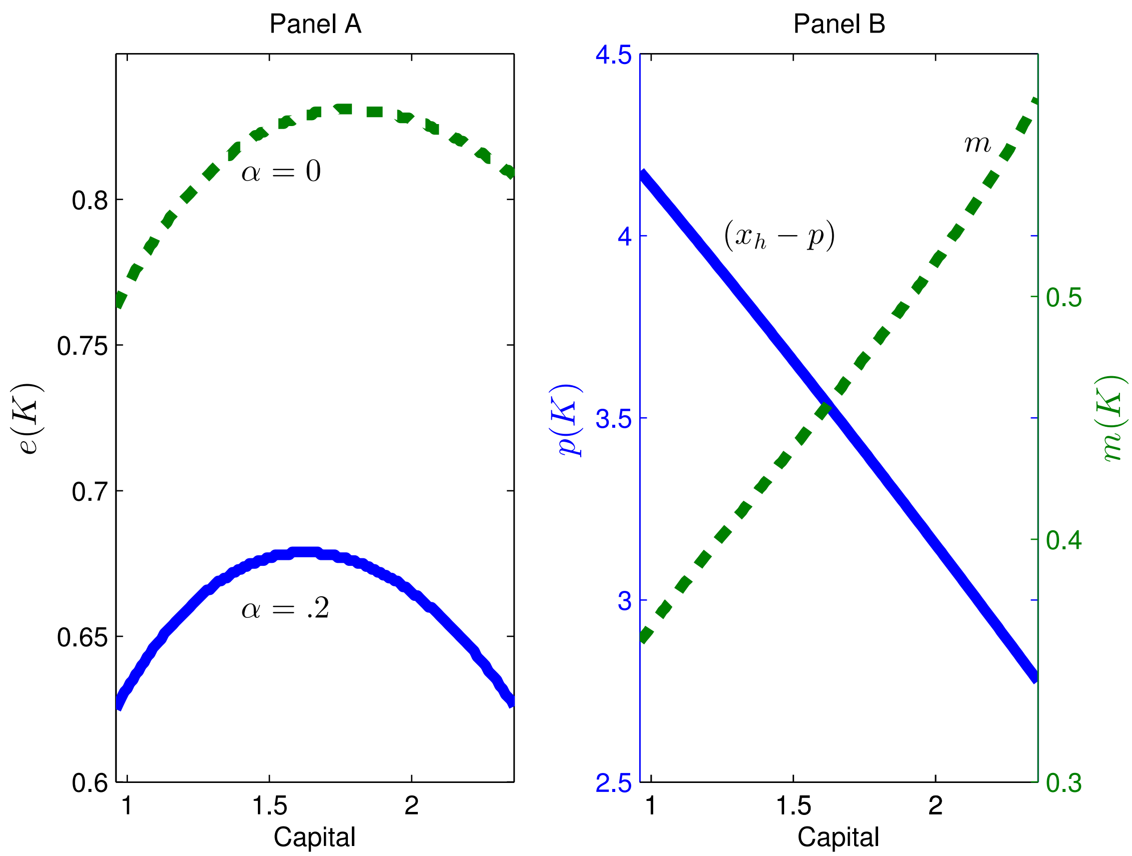


Figure 7: Fragility. Panel A: Fraction of high payoff assets in the balance sheet of financial intermediaries, g, as a function of capital, K. Panel B: Leverage, $\ell := D/K$, as a function of capital K for the cases $\alpha = 0$ (dashed line, in green) and $\alpha = .2$ (continuous line, in blue). In this example, $\psi(e) = \theta \frac{e^2}{2}$ with $\theta = .25$. In addition, $\kappa = .15$, $\eta = .5$, $M = .75$ and $x_h = 5$.

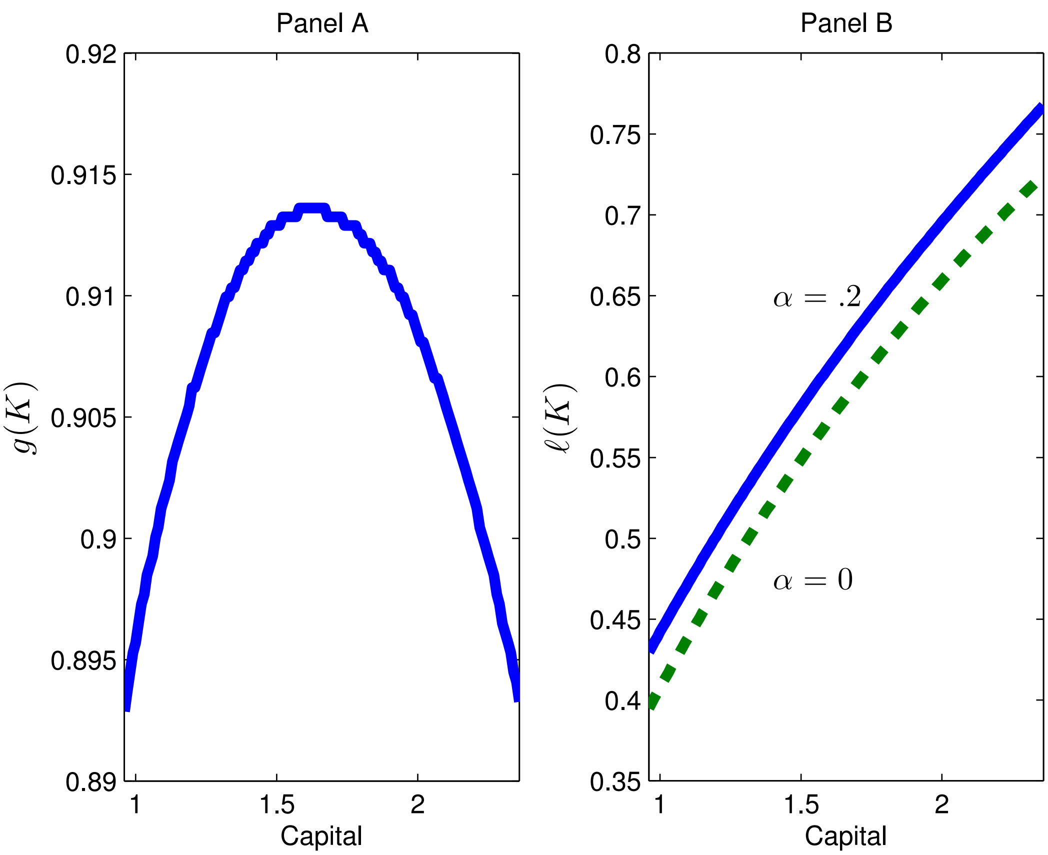
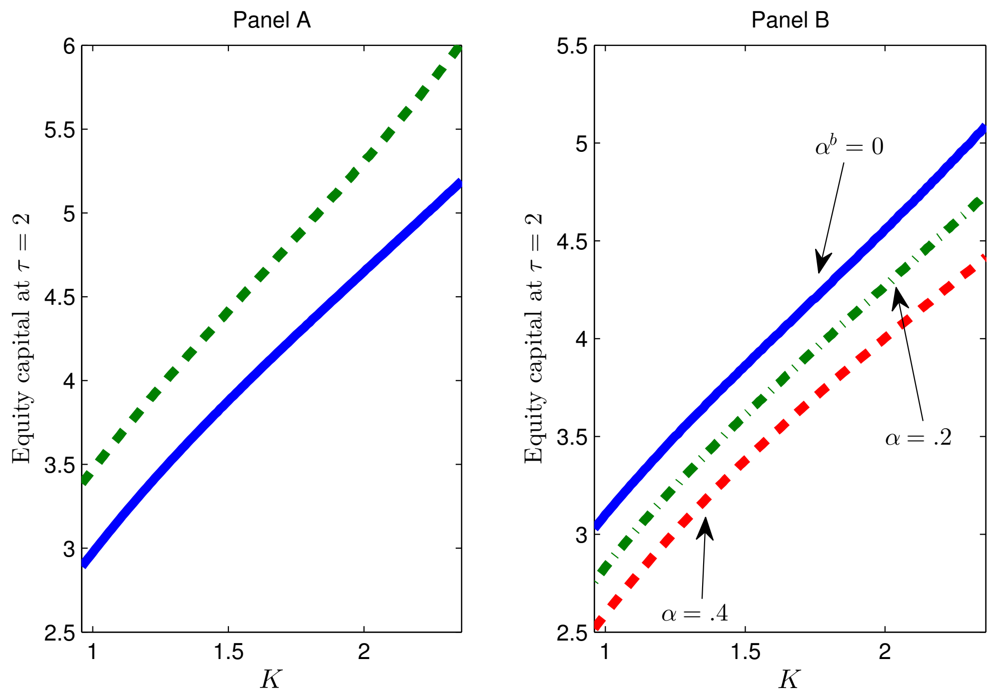


Figure 8: Panel A: Equity capital under the wrong beliefs, E_o^2 (dashed line) and the true equity capital at date 2, E_{true}^2, of the single optimistic intermediary who believes $\alpha = 0$ when $\alpha = .2$. Panel B: Equity capital when the entire intermediary sector believes that $\alpha_o = 0$ when the true value of α is $\alpha = .4$ (dashed line) and true capital of financial intermediaries (continuous line). In this example, $\psi(e) = \theta\frac{e^2}{2}$ with $\theta = .25$. In addition, $\kappa = .15$, $\eta = .5$, $M = .75$ and $x_h = 5$.



For Online Publication

A Appendix

A.1 Equations defining equilibria

In this section we write down necessary and sufficient conditions for a vector describing prices (p, p^d, r, R, r^x) and quantities $(e, g, q^u, D^u, q^i, y, D^i, m)$ in which $R > r > 1$ to be an equilibrium. Recall that in this case we necessarily have $r^x = r$, and $y = q^i$. In fact, we can parametrize the equilibrium which much fewer variables. Given (p, r) and candidate choices (e, q^u, q^r), we may define

$$\begin{aligned}
p^d &:= \kappa g x_h + (1-\kappa)\, p && \text{(A.1)}\\
m &:= \min\left\{\frac{Nq^i}{e+\alpha\,(1-e)}, 1\right\} && \text{(A.2)}\\
g &:= \frac{e}{e+\alpha\,(1-e)} && \text{(A.3)}\\
r^x &:= \frac{e(1-m)x_h}{p\,(1-m\,(e+\alpha\,(1-e)))} \qquad \text{and} \qquad R := \frac{gx_h}{p^d} && \text{(A.4)}\\
D^u &:= pq^u - K && \text{(A.5)}\\
D^i &:= p^d q^i - K && \text{(A.6)}
\end{aligned}$$

(p, r) and candidate choices (e, q^u, q^i) form an equilibrium with $R > r$ if and only if,

$$\begin{aligned}
D^u &\leq 0 && \text{(A.7)}\\
D^i &= (1-\eta)\, pq^i && \text{(A.8)}\\
p &= \min\left\{\frac{MK - ND^i}{1-m\,(e+\alpha\,(1-e))}, \frac{e(1-m)x_h}{1-m\,(e+\alpha\,(1-e))}\right\} && \text{(A.9)}\\
\psi'\,(e-\underline{e}) &= (1-\alpha)\, m\kappa\,(gx_h - p) && \text{(A.10)}\\
r &= r^x < R && \text{(A.11)}\\
Mq^u &+ Nq^i = 1 && \text{(A.12)}\\
MD^u &+ ND^i = 0 && \text{(A.13)}
\end{aligned}$$

Equation (A.8) is the leverage constraint and (A.6) is the budget constraint of informed investors. Equation (A.9) is the price in the exchange, which is the minimum of the cash-in-the-market price and the fair value one. (A.10) is the first order condition of originators. Equations (A.12) and (A.13) are the market clearing conditions.

It follows from (A.6) and (A.8) that in an equilibrium with $r > R$ then

$$q^i = \frac{K}{\kappa\left(gx_h - p\right) + \eta p} \tag{A.14}$$

$$D^i = \frac{(1-\eta)pK}{\kappa\left(gx_h - p\right) + \eta p}. \tag{A.15}$$

Furthermore since $D^i > 0$, $D^u \leq 0$. Thus (A.7) can be ignored.

It is easy to verify that in a strict CIM equilibrium, the pair (p, e) completely determines all the prices and quantities in equilibrium. Given p, equations (A.1), (A.5), (A.6), (A.12) and (A.13) uniquely determine (q^u, D^u, q^i, D^i). Thus given (p, e) the RHS of equations (A.1)-(A.6) are uniquely defined. As a consequence, we may parameterize the set of equilibria by the pair (p, e).

A.2 A system of equations in p and e

We will consider the model as parameterized by (K, N, α) and to save notation we will often set $\theta = (N, \alpha)$. The system of equations (A.9)-(A.15) can be simplified to yield a tractable system in two equations with two unknowns, the price in the exchange p and the originators' effort e. Indeed straightforward algebra shows that in a strict CIM equilibrium with $R > r$ and $r^x > 1$ then

$$f^1\left(p, e, K, \theta\right) := p - K\left(M + N\gamma\right) = 0 \tag{A.16}$$

$$f^2\left(p, e, K, \theta\right) := \left(e + \alpha\left(1-e\right)\right)\psi'(e - \underline{e}) - \left(1-\alpha\right)NK\beta = 0 \tag{A.17}$$

where

$$\gamma := \frac{\eta p}{\kappa\left(gx_h - p\right) + \eta p} \tag{A.18}$$

$$\beta := \frac{\kappa\left(gx_h - p\right)}{\kappa\left(gx_h - p\right) + \eta p} = 1 - \gamma, \tag{A.19}$$

Notice that

$$
\begin{aligned}
-\beta_p = \gamma_p &= \frac{\eta\kappa g x_h}{\left(\kappa\left(g x^h - p\right) + \eta p\right)^2} && \text{(A.20)}\\
-\beta_e = \gamma_e &= -\frac{\eta p \kappa x_h g_e}{\left(\kappa\left(g x^h - p\right) + \eta p\right)^2} \quad \text{where} \quad g_e = \frac{\alpha}{\left(e + \alpha\left(1-e\right)\right)^2} > 0 && \text{(A.21)}
\end{aligned}
$$

The function f is (mathematically) well defined for $0 \le N \le 1$, for $0 \le \alpha \le 1$ and for $0 \le p \le x_h$, $\underline{e} \le e \le \bar{e}$ and $K > 0$.

A "converse" also holds: Under additional conditions, the zeros of the system (A.16)-(A.17) correspond to CIM equilibria with $R > r = r^x > 1$. To show this let

$$
f\left(p, e, K, \theta\right) := \begin{pmatrix} f^1\left(p, e, K, \theta\right) \\ f^2\left(p, e, K, \theta\right) \end{pmatrix}. \tag{A.22}
$$

Suppose that (p, e) with $e \ge \underline{e}$ solves $f\left(p, e, K, \theta\right) = 0$, with $N \ge 0$ and $0 \le \alpha \le 1$. Use equation (A.3) to calculate the implied g and then equation (A.14) to calculate the implied q^i. Then equations (A.2), (A.1) and (A.4) can be used to yield the implied $m, p^d, r^x$and R. If $R > r^x$, choose $r = r^x$, set D^i using (A.15) and D^u using (A.13). It is then easy to check that (A.10) is satisfied. Set q^u using (A.6). Walras Law insures that the goods market is in equilibrium, that is (A.12) holds. Thus to obtain a strict CIM equilibrium from a zero of f it suffices to apply this algorithm and verify that $R > r := r^x > 1$. This algorithm also shows that there is a unique strict CIM equilibrium that would generate the same (p, e). For this reason, with a slight abuse of language, we call (p, e) ***a strict CIM equilibrium.*** The algorithm also shows that the implied prices (p^d, r^x, r, R) and quantities $(g, q^u, D^u, q^i, y, D^i, m)$ vary continuously with (p, e, K, θ).

We make use of the implicit function theorem below, both to show the existence of an equilibrium where $R > r = r^x > 1$ and to characterize its basic properties. We will use as a starting point $\alpha = N = 0,\ e = \underline{e}$ and thus, to guarantee that this starting point is interior, we need to extend the domain of f. Notice first that f is mathematically defined for $N < 0$, and $\alpha > \frac{-\underline{e}}{1-\underline{e}}$. We extend the definition of f for $e \in (\underline{e} - \epsilon, 1)$, ϵ small. We do this by defining $\psi(z)$ for $z > -\epsilon$ in a C^2 manner, while preserving convexity. Notice that the convexity of the extended ψ guarantees that $\psi_z(z) \le 0$ for $z < 0$.

The calculation of the matrix of the partial derivatives of f at (p, e, K, N, α) is straightforward.

$$f_p^1 := \frac{\partial f^1}{\partial p} = 1 - NK\gamma_p \tag{A.23}$$

$$f_e^1 := \frac{\partial f^1}{\partial e} = -NK\gamma_e \tag{A.24}$$

$$f_p^2 := \frac{\partial f^2}{\partial p} = (1-\alpha)NK\gamma_p \tag{A.25}$$

$$f_e^2 := \frac{\partial f^2}{\partial e} = (1-\alpha)\psi' + (e + \alpha(1-e))\,\psi'' + (1-\alpha)NK\gamma_e \tag{A.26}$$

Write

$$\partial_{p,e} f\,(p, e, K, \theta) = \begin{pmatrix} f_p^1 & f_e^1 \\ f_p^2 & f_e^2 \end{pmatrix} \qquad \text{with} \qquad |\partial_{p,e} f| = f_p^1 f_e^2 - f_p^2 f_e^1, \tag{A.27}$$

For later use, notice that

$$f_K^1 := \frac{\partial f^1}{\partial K} = -\left(M + N\gamma\right) < 0 \tag{A.28}$$

$$f_K^2 := \frac{\partial f^2}{\partial K} = -\left(1-\alpha\right) N\beta \leq 0 \tag{A.29}$$

Lemma A.1 *There exists $\bar{\epsilon} > 0$ such that if $N < \bar{N}(K) := \min\{1, \frac{\bar{\epsilon}}{K}\}$ and (p, e) is a strict CIM equilibrium for parameter values (K, N, α) and if e is the associated effort in this equilibrium then:*

$$f(p, e, K, N, \alpha) = 0$$

$$f_e^2(p, e, K, N, \alpha) > 0 \tag{A.30}$$

$$|\partial_{p,e} f(p, e, K, N, \alpha)| > 0. \tag{A.31}$$

Proof: We have already shown that in a strict CIM equilibrium for parameter values (K, N, α), $f(p, e, K, N, \alpha) = 0$. To prove the remaining claims, first notice that (A.9) guarantees that $gx_h < p$. Thus $\gamma < 1$ and $\beta > 0$. Furthermore

$$\begin{aligned} NK\gamma_p = NK\frac{\eta\kappa g x_h}{\left(\kappa\left(gx_h - p\right) + \eta p\right)^2} &= Nq\frac{\eta\kappa g x_h}{\kappa\left(gx_h - p\right) + \eta p} \\ &= \left(e + \alpha\left(1-e\right)\right) m \frac{\eta\kappa g x_h}{\kappa\left(gx_h - p\right) + \eta p}. \end{aligned}$$

Here we used (A.20) for the first equality, (A.14) for the second equality and (A.2) for the third equality. If $\eta \geq \kappa$ then,

$$\kappa < \frac{\eta\kappa g x_h}{\kappa\left(g x_h - p\right) + \eta p} \leq \eta, \tag{A.32}$$

and if $\kappa > \eta$

$$\eta < \frac{\eta\kappa g x_h}{\kappa\left(g x_h - p\right) + \eta p} < \frac{\eta\kappa g x_h}{\eta\left(g x_h - p\right) + \eta p} \leq \kappa.$$

Since $\underline{e} < (e + \alpha(1-e)m) < 1$, for any α,

$$\begin{aligned} f_p^1 &\geq 1 - \max\{\eta, \kappa\} > 0, & \text{(A.33)} \\ f_p^2 &\leq (1-\alpha)\max\{\eta, \kappa\}. & \text{(A.34)} \end{aligned}$$

Notice that if $\eta \leq \kappa$,

$$\frac{p}{\left(k(g x_h - p) + \eta p\right)^2} \leq \frac{p}{(\eta g x_h)^2} \leq \frac{1}{\eta^2 g x_h}$$

and if $\kappa > \eta$,

$$\frac{p}{\left(k(g x_h - p) + \eta p\right)^2} \leq \frac{1}{\kappa^2 g x_h}$$

Thus

$$(1-\alpha)|\gamma_e| \leq \frac{(1-\alpha)g_e}{g\min\{\eta,\kappa\}} \leq \frac{(1-\alpha)\alpha}{\underline{e}^2\min\{\eta,\kappa\}} \leq \frac{1}{4\underline{e}^2\min\{\eta,\kappa\}}.$$

Since $\psi'' >> 0$, there exists ϵ_1 such that if $NK < \epsilon_1$ then $f_e^2 \geq \frac{\underline{e}\inf\psi''}{2} >> 0$. Hence (A.30) holds. In addition, there exists $\bar{\epsilon} \leq \epsilon_1$ such that if $NK < \bar{\epsilon}$ then

$$\left(1 - \max\{\eta, \kappa\}\right)\frac{\underline{e}\inf\psi''}{2} - \frac{\max\{\eta,\kappa\}NK}{4\underline{e}^2\min\{\eta,\kappa\}} > 0,$$

which insures (A.31). □

Lemma A.2 *If $N < \bar{N}(K) := \min\{1, \frac{\bar{\epsilon}}{K}\}$ and (p, e) is a strict CIM equilibrium for parameter values (K, N, α) then $p_N < 0$, $e_N > 0$ and $\mathcal{L}_N < 0$.*

Proof: This follows from the previous Lemma, since

$$f_N^1 = K\beta > 0 \qquad \text{and} \qquad f_N^2 = -(1-\alpha)K\beta < 0.$$

In addition, since $e_N > 0$ and $g_e > 0$, p/p^d decreases with N and thus $\mathcal{L}_N < 0$. □

Lemma A.3 *If $N < \bar{N}(K) := \min\{1, \frac{\bar{\epsilon}}{K}\}$ and (p, e) is a strict CIM equilibrium for parameter values (K, N, α, η) then $p_\eta > 0$ and $e_\eta < 0$.*

Proof: This follows since $\gamma_\eta > 0$, $f^1_\eta = -KN\gamma_\eta < 0 > 0$ and $f^2_\eta = (1-\alpha)KN\gamma_\eta > 0$. $\square$

The following Lemma gives bounds for the price in the exchange p.

Lemma A.4 *In a strict CIM equilibrium prices in the exchange satisfy $K > p > \frac{MK}{2-\eta} > 0$.*

Proof: Equations (A.5),(A.6), (A.12), (A.13) and $p^d \geq p$ imply $p \leq K$. Furthermore, in a strict CIM equilibrium

$$
\begin{aligned}
p &= \frac{MK - ND^i}{1 - m\left(e + \alpha\left(1-e\right)\right)} \\
&> MK - ND^i \\
&= MK - N\left(1-\eta\right)pq^i \\
&> MK - \left(1-\eta\right)p, \qquad \text{since} \qquad Nq^i < 1.
\end{aligned}
$$

$\square$

Lemma A.5 *Suppose $(\tilde{p}, \tilde{e})$ is a strict CIM equilibrium for parameter values $(\tilde{K}, \tilde{N}, \tilde{\alpha})$, with $\tilde{N}\tilde{K} < \bar{\epsilon}$. Then (i) There exists a $\delta > 0$ and $\epsilon > 0$ such that for $|K - \tilde{K}| < \delta$ there exists a unique $(p(K), e(K))$ that is a strict CIM equilibrium for parameter values $(K, \tilde{N}, \tilde{\alpha})$ and within ϵ from $(\tilde{p}, \tilde{e})$. Furthermore (ii) $p'(\tilde{K}) > 0$, and (iii) When $\eta < \kappa$, then (a) If $\tilde{K} < \frac{\kappa\tilde{g}x_h}{\kappa+\eta}$ then $e'(\tilde{K}) > 0$; (b) if $\tilde{K} > \frac{(2-\eta)x_h}{2-\eta+M}$ then $e'(\tilde{K}) < 0$; and (c) if $e'(\tilde{K}) = 0$ then $\tilde{K}$ is a local maximum of e.*

Proof: To simplify notation we omit the arguments $\tilde{N}$ and α in what follows. Since $(\tilde{p}, \tilde{e})$ is a strict CIM equilibrium, by Lemma A.1 there exists $\delta > 0$ and δ' such for $|K - \tilde{K}| < \delta$ there exists unique $(p(K), e(K))$ with $|(p(K), e(K)) - (\tilde{p}, \tilde{e})| < \delta'$ and $f(p(K), e(K), K) = 0$. Furthermore the functions p and e are at least C^2. The continuity of the candidate equilibrium constructed from a zero of $f(p, e, K)$ guarantees that $(p(K), e(K))$ is indeed a strict CIM equilibrium and the uniqueness of the zeros of $f(p, e, K)$ in a neighborhood of $(\tilde{p}, \tilde{e})$ establishes (i).

(ii) is immediate since $f_K^1 < 0$ and $f_K^2 \leq 0$.

Proof of (iii): For $|K - \tilde{K}| < \delta$ let

$$\mathcal{A}(K) := \frac{f_p^2(p(K), e(K)) f_K^1(p(K), e(K)) - f_p^1(p(K), e(K)) f_K^2(p(K), e(K))}{1 - \alpha}$$

and thus

$$e'(K) = \frac{\mathcal{A}(K)(1-\alpha)}{|\partial_{p,e} f(p(K), e(p(K))|}$$

where, omitting the argument $(p(K), e(K))$ to simplify the expressions,

$$\begin{aligned}
\mathcal{A}(K) &= NK\beta_p (M + N\gamma) + (1 - NK\gamma_p)) N\beta \\
&= N\beta + NMK\beta_p + N^2 K [\gamma\beta_p - \gamma_p\beta] \\
&= N\beta + NMK\beta_p + N^2 K [\gamma\beta_p + \beta_p\beta] \\
&= N\beta + \beta_p NK \qquad \text{using (A.19) and the fact that } M + N = 1.
\end{aligned}$$

Thus, writing again the argument $(p(K), e(K))$

$$\mathcal{A}(K) = N\beta(p(K), e(K)) + \beta_p(p(K), e(K)) NK.$$

If $\eta < \kappa$ then

$$(gx_h - p) < \frac{\beta}{|\beta_p|} < \frac{\kappa}{\eta}(gx_h - p). \tag{A.35}$$

If $K < \frac{\beta}{|\beta_p|}$ then $\mathcal{A}(K) > 0$. Since $g(\tilde{\alpha}e) = \frac{e}{e + \tilde{\alpha}(1-e)}$ is monotone increasing in e, $g \geq \tilde{g}$. The upper bound on p established in Lemma A.4 implies (a). Furthermore, if $K > \frac{(2-\eta)x_h}{2-\eta+M}$ then the lower bound in (A.35) and the upper bound on p established in Lemma A.4 imply (b). Suppose now that that $e'(\tilde{K}) = 0$, and hence $\mathcal{A}\left(\tilde{K}\right) = 0$. Since $\eta < \kappa$, $\beta_{pp}(p(\tilde{K}), e(\tilde{K})) < 0$, and thus

$$\mathcal{A}'\left(\tilde{K}\right) = N\left[\beta_p(p(\tilde{K}), e(\tilde{K}))p'(\tilde{K}) + \beta_{pp}(p(\tilde{K}), e(\tilde{K}))p'(\tilde{K})\tilde{K} + \beta_p(p(\tilde{K}), e(\tilde{K}))\right] < 0.$$

This proves (c). □

Remark A.6 *Suppose the assumptions of Lemma A.5 hold at $(\tilde{K}, \tilde{N}, \tilde{\alpha})$. Then while*

$$|\partial_{p,e} f(p, e, K, \tilde{N}, \alpha)| > 0$$

one can prolong the domain of the functions $p(K)$ and $e(K)$. The Jacobian stays positive at least while $\bar{N}(K) \geq N$. Since the function $\bar{N}$ is non-increasing, a decrease in K is always possible, but an increase in K may lead to a violation of the bound on N. Furthermore, except for a K close enough to $\tilde{K}$, there is no guarantee that the solution $(p(K), e(K))$ would lead to a candidate equilibrium with $R(K) > r(K) := r^x(K) > 1$. However if for $K \in [K_1, K_2]$ $(p(K), e(K))$ leads to a CIM equilibrium with $R(K) > r(K) := r^x(K) > 1$ and $N \leq \bar{N}(K_2)$ we can use Lemma A.5 to compare the equilibria $(p(K), e(K))$. In particular the exchange price p increases with K. If $K_1$is small enough and K_2 large enough, the level of effort has a single global maximum - it increases if $K < \bar{K}$ and decreases for $K > \bar{K}$ for some $K_1 \leq \bar{K} \leq K_2$.

Proposition A.7 *(i) If $(\tilde{p}, \tilde{e})$ is a strict CIM equilibrium for parameter values $(\tilde{K}, N, \alpha)$ with $0 < N < \bar{N}(\tilde{K})$ and $e'(\tilde{K}) \leq 0$ then $\mathcal{L}'\left(\tilde{K}\right) > 0$. (ii) There exists a $\hat{N}$ such that if $(\tilde{p}, \tilde{e})$ is a strict CIM equilibrium for parameter values $(\tilde{K}, N, \alpha)$, with $N < \min\{\hat{N}, \bar{N}(\tilde{K})\}$ then $\mathcal{L}'\left(\tilde{K}\right) > 0$*

Proof. In a strict CIM equilibrium

$$D^i = \frac{(1-\eta)pK}{\kappa\left(gx_h - p\right) + \eta p}.$$

Omitting the argument $\tilde{K}$ to lighten up notation:

$$\begin{aligned}
(D^i)' &= \left[\left(p + p'\tilde{K}\right)\left(\kappa\left(gx_h - p\right) + \eta\right) - p\tilde{K}\left(\eta - \kappa\right)p' - p\tilde{K}\kappa g_e e' x_h\right] \\
&\times \frac{1-\eta}{\left(\kappa\left(gx_h - p\right) + \eta p\right)^2} \\
&= \left[p\left(\kappa\left(gx_h - p\right) + \eta p\right) + \tilde{K}\kappa x_h\left(p' - p g_e e'\right)\right] \times \frac{1-\eta}{\left(\kappa\left(gx_h - p\right) + \eta p\right)^2}
\end{aligned}$$

Thus,

$$\mathcal{L}'\left(\tilde{K}\right) = \frac{(D^i)'\tilde{K} - D^i}{\tilde{K}^2} = \frac{(1-\eta)\kappa x_h\left(p' - p g_e e'\right)}{\tilde{K}\left(\kappa\left(gx_h - p\right) + \eta p\right)^2}.$$

Thus (i) follows from Lemma A.5. Furthermore if $N < \bar{N}(K)$,

$$\begin{aligned} e' &< \frac{1}{|\partial_{p,e} f(p,e,K,N,\alpha)|} f_p^1 (1-\alpha) N \beta \\ &\leq \frac{(1-\max\{\eta,\kappa\})N}{|\partial_{p,e} f(p,e,K,N,\alpha)|} \\ &\leq \frac{N}{|\partial_{p,e} f(p,e,K,N,\alpha)|} \end{aligned}$$

and if $N < \bar{N}(K)$

$$\begin{aligned} p' &> \frac{1}{|\partial_{p,e} f(p,e,K,N,\alpha)|} \underline{e} \frac{\inf \psi''}{2} (M + N\gamma) \\ &\geq \frac{\underline{e}}{|\partial_{p,e} f(p,e,K,N,\alpha)|} \frac{\inf \psi''}{2} \gamma \end{aligned}$$

In the last equation we used $M + N\gamma = 1 - N + N\gamma > \gamma$. Furthermore,

$$\gamma \geq \frac{\min\{\kappa,\eta\} p}{\max\{\kappa,\eta\} x_h}$$

Thus by choosing

$$\hat{N} = \underline{e} \frac{\inf \psi''}{2} \frac{\min\{\kappa,\eta\}}{\max\{\kappa,\eta\} x_h},$$

we have that for $N < \min\{\hat{N}, \bar{N}(\tilde{K})$, since $g_e \leq 1$

$$p' - p g_e e' \geq \frac{p}{|\partial_{p,e} f(p,e,K,N,\alpha)|} \left[\underline{e} \frac{\inf \psi''}{2} \frac{\min\{\kappa,\eta\}}{\max\{\kappa,\eta\} x_h} - N \right] > 0 \tag{A.36}$$

establishing the second claim. □

Until now, we have assumed the existence of a CIM equilibrium. The next Lemma shows that for a set of parameter values there exists a CIM equilibrium with $R > r^x$. Recall that we defined for each $\tilde{\alpha} \geq 0$,

$$\tilde{g} = g(\tilde{\alpha}, \underline{e}) = \frac{\underline{e}}{\underline{e} + \tilde{\alpha}(1 - \underline{e})}$$

Proposition A.8 *(Existence) Suppose that*

$$\frac{\underline{e} \tilde{g} \kappa x_h}{\tilde{g} - \underline{e} + \underline{e}\kappa} < \tilde{K} < \underline{e} x_h. \tag{A.37}$$

Then (i) there exists a neighborhood $\mathcal{N}$ of $(\tilde{K}, 0, \tilde{\alpha})$ and $\epsilon > 0$, such that for every $(K, N, \alpha) \in \mathcal{N}$ there is a unique $p(K, N, \alpha) > 0$ and $e(K, N, \alpha) > 0$ that is solution of $f(p, e, K, N, \alpha) = 0$ with $|(p(K, N, \alpha), e(K, N, \alpha)) - (\tilde{K}, \underline{e})| < \epsilon$. The functions $p(K, N, \alpha)$ and $e(K, N, \alpha)$ are C^2. (ii) If $N > 0$ and $\alpha \geq 0$ then $e > \underline{e}$. (iii) One may choose $\mathcal{N}$ such that $(p(K, N, \alpha), e(K, N, \alpha))$ is a strict CIM equilibrium for the parameter values $(K, N, \alpha) \in \mathcal{N}$.

Proof: It is easy to check that $f(\tilde{K}, \underline{e}, \bar{K}, 0, \tilde{\alpha}) = 0$. Since $|\partial_{p,e} f(p, e, \tilde{K}, 0, \tilde{\alpha})| > 0$, the implicit function theorem guarantees that there exist a neighborhood $\mathcal{N}$ of $(\tilde{K}, 0, N, \tilde{\alpha})$ and $\epsilon > 0$ such that for each $(K, N, \alpha) \in \mathcal{N}$ there exists a unique $(p, e) > 0$ with $|(p, e) - (\tilde{K}, \underline{e})| < \epsilon$ such that $f(p, e, K, N, \alpha) = 0$, and that the functions $p(K, N, \alpha)$ and $e(K, N, \alpha)$ are C^2. Thus (i) holds. To establish (ii) note that if $N > 0$ and $\alpha \geq 0$ then from equation (A.17). $\psi_e(e(K, N, \alpha) - \underline{e}) > 0$ and hence $e(K, N, \alpha) > \underline{e}$.

To show (iii), we first show that $(\tilde{K}, \underline{e})$ is a strict CIM equilibrium for parameters $(\tilde{K}, 0, \alpha)$. In fact by setting $g = \tilde{g}$ and using equation (A.14) to calculate the implied q^i, it is easy to check that $m = 0$ solves (A.2), and p^d, r^x and R can be calculated using equations (A.1) and (A.4). The first inequality in (A.37) guarantees that $R > r^x$ an thus we may choose $r = r^x$, set D^i using (A.15) and $D^u = 0$ to satisfy (A.13). The second inequality in (A.37) insures that $r^x > 1$, and thus we may set $q^u = 1$. The remaining of the proof is as in the proof of Lemma A.5. Using the continuity of the implied strict CIM equilibrium prices and quantities with respect to the solution of $f(p, e, K, N, \alpha) = 0$ we may choose $\mathcal{N}$ such that the conditions $m < 1$, $R > r := r^x > 1$ are always satisfied whenever $N > 0$ and $\alpha \geq 0$. Hence $(p(K, N, \alpha), e(K, N, \alpha))$ is a strict CIM equilibrium when the parameters are given by $(K, N, \alpha) \in \mathcal{N}$ and $N > 0, \alpha \geq 0$. $\square$

Proposition A.9 *In addition, one can choose the neighborhood $\mathcal{N}$ such that there are no other equilibria other than the (unique) strict CIM equilibrium.*

Proof: Suppose there is a sequence of equilibria (p_n, e_n) associated with the sequence of parameter values $(K_n, N_n, \alpha_n) \to (\tilde{K}, 0, \tilde{\alpha})$. Using equations (A.8) and (A.6), since $g_n < 1$, we obtain:

$$(\kappa g_n x_h + (\eta - \kappa) p_n)\, q_n^i = K_n$$

Since $p_n < g_n x_h$

$$\eta p_n q_n^i < K_n.$$

Since $M_n \to 1$ and $K_n \to \tilde{K}$, Lemma A.4 implies that q_n^i is bounded, and thus $m_n \to 0$, and $e_n \to \underline{e}$. Since $p_n q_n$ is bounded, the leverage constraint implies D_n^i bounded and thus, in equilibrium, $D_n^u \to 0$. Given any $\delta > 0$ (A.9) implies that for n large, $p_n < K_n + \delta$. If δ is small enough, p_n cannot correspond to a Fair Value equilibrium if n is large, since $K_n < \underline{e} x_h$. If the equilibrium associated with N_n, α_n is a CIM equilibrium, then necessarily $p_n \to \tilde{K}$ and since $e_n \to \underline{e}$ and (A.37) holds, for n large $R_n > r_n^x \geq r_n$ and (p_n, e_n) is a strict CIM equilibrium. Thus $f(p_n, e_n, K_n, N_n, \alpha_n) = 0$. and since for n large, $|(p_n, e_n) - (\tilde{K}, \underline{e})| < \epsilon$, by the previous Proposition, $(p_n, e_n) = (p(K_n, N_n, \alpha_n), e(K_n, N_n, \alpha_n))$. □

Proof of Proposition 1. Notice that if $\tilde{\alpha} \leq \hat{\alpha} < 1$ then $\tilde{g} \in [\frac{\underline{e}}{\underline{e}+\hat{\alpha}(1-\underline{e})}, 1]$ and that $\frac{\underline{e}\tilde{g}\kappa x_h}{\tilde{g}-\underline{e}+\underline{e}\kappa}$ decreases with $\tilde{g}$. Hence $\frac{\underline{e}\tilde{g}\kappa x_h}{\tilde{g}-\underline{e}+\underline{e}\kappa} < \underline{e} x_h$ and is maximized when $\tilde{g} = \hat{g} := g(\hat{\alpha}, \underline{e})$. Set

$$K_a = \frac{\underline{e}\hat{g}\kappa x_h}{\hat{g} - \underline{e} + \underline{e}\kappa} \text{ and } K_b = \underline{e} x_h.$$

The Proposition is now a consequence of Propositions A.8 and A.9 □

Proof of Corollary 6. It follows from (20) that

$$\underline{\epsilon}\,(K) = 1 - \frac{(1-\eta)\,p}{p^d}.$$

Since $\alpha = 0$

$$\frac{\partial p^d}{\partial K} = (1-\kappa)\,\frac{\partial p}{\partial K} \qquad \Rightarrow \qquad \frac{\partial \underline{\epsilon}}{\partial K} \propto -\kappa x_h \frac{\partial p}{\partial K} < 0,$$

by Proposition 2. □

Proof of Corollary 7. It follows from (21) that

$$\frac{\partial \epsilon\,(K, \mathsf{L})}{\partial \mathsf{L}} = \frac{1}{p^d q}.$$

Since by Proposition 4, $\mathcal{L}$ increases with K so does the value of total assets in the hand of intermediaries $p^d q$ and thus

$$\frac{\partial^2 \epsilon (K, \mathsf{L})}{\partial \mathsf{L} \partial \mathsf{K}} < 0.$$

$\square$