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

March 12, 2016

Abstract

We investigate the effects of an increase in liquidity (a "savings glut") on the incentives to originate high quality assets and on the fragility of the financial sector. Originators incur private costs when originating high quality assets. Assets are subsequently distributed in two markets: A private market where informed intermediaries operate and an exchange where uninformed liquidity trades. Uninformed liquidity pays the same price irrespective of the quality of the assets, which discourages good origination. Informed liquidity instead creams skims the best assets paying a premium over the uninformed price, which encourages originators to supply good assets. We show that the positive origination effects of an increase in liquidity matter when the overall level of liquidity is low whereas the opposite is true when liquidity is abundant - an increase in liquidity has a non-monotone effect on origination incentives. Leverage increases monotonically with liquidity and is highest precisely when incentives for good asset origination are at their lowest. Thus plentiful liquidity leads to fragile balance sheets: On the asset side there are more low quality assets and on the liability side more of those assets are funded with debt. We relate our findings to some of the stylized facts observed in financial markets in the lead up to the Great Recession and draw policy conclusions from the model.

*We thank Hyun Shin for insightful comments on an earlier version of the paper.

Electronic copy available at: http://ssrn.com/abstract=2743700

"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, which sets 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, or 'savings gluts', are still poorly understood. This paper identifies a mechanism linked to spread compression caused by a savings glut in a model of origination and distribution of assets that broadens our framework in Bolton, Santos, and Scheinkman (2011 and 2016).

In our model it takes costly effort for an originator to produce an asset of high quality. Irrespective of the quality, the originated asset is distributed to investors in two different markets: An organized securities market, where the asset can be sold to uninformed investors, and an over-the-counter (OTC) market, where the asset may be bought by an informed intermediary, who has superior, but noisy, information about the quality of the asset. The central mechanism we study is the incentives financial markets provide originators and how these market incentives are affected by changes in capital flows or liquidity. There are no market incentives provided by the organized exchange, since investors in this market are unable to distinguish between high and low quality assets. Market incentives for origination of good assets only exist to the extent that informed investors 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–the difference in asset prices in the OTC and organized market–so much that origination incentives are undermined. We focus on the case where capital in the hands of informed investors is relatively scarce and thus in equilibrium informed traders have a higher rate of

Electronic copy available at: http://ssrn.com/abstract=2743700

return and leverage their superior information by borrowing from uninformed traders. Thus informed traders act as *financial intermediaries*. We show that intermediary leverage increases monotonically with liquidity and is highest when origination incentives are at their lowest. In other words, the savings glut gradually increases the fragility of financial intermediaries: on the liability side, they are increasingly leveraged, and on the asset side, their balance sheets increasingly contain non-performing assets.

This general result provides a unifying explanation, based on one simple economic mechanism, for why savings gluts are associated with both origination of non-performing assets and increasing leverage of financial intermediaries late in a credit cycle. An alternative explanation by Minsky (1992), which became prominent after the financial crisis of 2007-09, essentially points to investor psychology. According to Minsky financial fragility and mounting exposure to non-performing assets is simply due to the growing risk appetite of investors. But he establishes no clear link between greater risk taking and increasing origination of non-performing assets.

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. Many commentators attribute the crisis of 2007-09 to this real-estate and lending boom. But this explanation is too narrow and leaves out the heart of the crisis, namely the failure of over-leveraged financial institutions. Borio and Disyatat (2011) emphasize instead the critical role played by the financial sector in the crisis (see also Tokunaga and Epstein, 2014). 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. In effect, banks acted as a multiplier of the savings glut, pouring vast quantities of new fuel on the fire. There is plenty of evidence around of this 'excess elasticity. Adrian and Shin (2010), most notably, have shown how US brokers-dealers considerably increased both their balance sheets and leverage in the years leading up to the great recession. And, Keys, Piskorski, Seru and Vig (2013), among others, have shown how this expansion came at the cost of a severe deterioration in underwriting standards of mortgages (see, for example, their Figure 4.4).

Our model predicts all these facts in a simple framework. We show how an increase in aggregate savings gives rise to an increase in asset prices, a compression in spreads across financial markets and a deterioration of underwriting standards at origination. In addition, we show that the balance sheets of financial intermediaries display the high elasticity with respect to savings suggested by Borio and Disyatat (2011) and that this elasticity is largely driven by leverage, which displays the strong procyclical features described in Adrian and Shin (2010).

The economy is comprised of two sectors in our model: An asset originator sector and a financial sector in which these assets are distributed. Originators incur costly effort to produce high quality assets. These assets are then distributed to the financial sector. We take the volume of originated projects and the distribution decision as exogenous, but we briefly discuss the implications of endogenizing these decisions in Section 5. The heart of our analysis is a rich modeling of the financial sector, which features three markets as in Bolton, Santos and Scheinkman (2016). First, there is an OTC market where, informed investors cream-skim the best assets generated by originators. Their information is the only source of incentives for originators to produce good assets. Second, there is a public market, an exchange where the assets that are not cream skimmed by informed investors are distributed and where buyers cannot select among assets.[1] There is thus a single price for the assets sold in the exchange. Demand for these assets can come from both informed traders and a pool of uninformed investors. These uninformed investors though cannot access the OTC market, as they do not have an information about the quality of originated assets. Third, there is a collateralized debt, or repo market, where financial intermediaries can borrow from investors.

A key assumption that considerably simplifies our analysis is that in equilibrium *cash-in-the-market pricing* of assets (Allen and Gale, 1999) prevails - that is asset prices are determined by the ratio of aggregate liquidity and the volume of distributed assets. Cash-in-the-market pricing implies that even when investors are risk-neutral, aggregate liquidity affects asset prices and as a consequence affects origination incentives. A more conventional way to deliver the same effect would be to assume that investors are risk-averse, but this would

[1] One can think as the assets that are not cream skimmed as being pooled and the shares against that pool of collateral as the assets traded in the exchange.

substantially complicate the algebra. The Allen and Gale device may be thought as a particularly convenient way of obtaining the same effect as risk aversion.[2] We thus assume that the aggregate liquidity (savings) that investors deploy in the market is exogenously determined. We compare equilibria for different levels of aggregate liquidity, keeping constant the distribution of capital across informed and uninformed investors. Thus an increase in liquidity does not mean an increase in the capital in the hands of "dumb" investors. The starting point of our analysis is to identify conditions under which an increase in aggregate savings leads to higher asset prices on the exchange. When that is the case, a key first observation is that asset prices in the OTC market rise less than proportionately, 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 the asset price on the exchange. By arbitrage, the repo rate is then also reduced.

Informed investors lever their knowledge by using their balance sheets in repo markets. We show that a consequence of the increase in liquidity is that the balance sheet of the financial intermediaries grows with it. Moreover the growth in the balance sheet is mostly driven by debt though there is also some growth in book equity. The end result is that leverage amongst financial intermediaries grows as aggregate liquidity 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 assets distributed, and are willing to pay more for a high quality asset, the increase of their balance sheets ought to result in better incentives for originators. However, there is a countervailing effect through the narrowing of spreads, which undoes incentives. We show that when the size of the funds flowing into financial markets is relatively low, the first effect dominates, improving origination standards, but when the flow becomes so large that it turns into a savings glut, the price difference narrows and origination standards drop.

In a savings glut intermediaries become more fragile not just as a result of their increased leverage, but also because of the deterioration of the asset side of their balance sheet. This is due to the fact that, although intermediaries are informed, they on occasion mistake a

[2] We elaborate on this interpretation in Section 5.

non-performing asset for a high quality asset. When underwriting standards deteriorate, intermediaries are more exposed to making such a mistake because the deal flow they see is of lesser quality. All in all, we show that even if intermediaries information does not change they end up accumulating a greater fraction of non-performing assets in a savings glut. In sum, our model predicts several of the main stylized patterns observed during the years leading up to the crisis of 2007-09. We do so by integrating Bernanke's savings glut hypothesis and the elasticity of balance sheets view of Borio and Disyatat (2011). The model's predictions do not rely on extrapolation biases in asset prices or any other behavioral bias. We do not dispute that these biases may have played a role and in the last section of the paper we explore what differences these biases introduce in the model.

Other related literature. We provide microfoundations for one of the leading hypotheses on the origins of the crisis of 2007-09 in the US, Spain, and Ireland. - a savings glut combined with 'balance-sheet elasticity' of financial intermediaries. Other commentators, most notably Shin (2012), Gourinchas (2012) and Borio and Disyatat (2011), have argued that it is the rise in global liquidity more than global imbalances that is the major cause. The advent of the euro, in particular, argues Shin (2011), meant that: "money (i.e. bank liabilities) was free-flowing across borders, but the asset side remained stubbornly local and immobile." The growth of the real estate sector in Spain is a case in point. Spain experienced an upsurge in liquidity largely intermediated by Spanish banks, which played the role of "informed" intermediaries, as in our model, while the securitization machine distributed real-estate-backed assets directly to other Eurozone investors.[3] More systematic evidence that financial crises are preceded by elevated credit growth and low interest rates, and that credit growth is the best predictor of financial instability is provided by Jordà, Schularick and Taylor (2011).

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 (2015) propose a model of intermedi-

[3]See Santos (2015) for an account of flows in the Spanish banking sector in the years leading up to the Eurozone crisis.

ation 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. Boissay, Collard and Smets (2016) offer a dynamic 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 greatly increased the size of their balance sheets (and their leverage), increasingly relying on repo financing, while repo rates kept falling.

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.

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 the collateral requirements. Therefore, even after exhausting their borrowing capacity, informed traders may not have sufficient capital to deploy 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 a high quality asset is traded on 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 and that

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

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 $g x_h \geq p$, which is a condition satisfied by the equilibrium price on the exchange p, as we show below.[4]

[4]See Bolton, Santos and Scheinkman (2016) for a detailed derivation of the Nash bargaining solution.

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 risk-free 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 intermediary can borrow against purchased assets at the exchange market value p of these assets with a haircut $\eta \geq 0$. 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. There are two reasons for selecting this price: First, the price p is the best valuation estimate of uninformed investors, who in equilibrium are the lenders. Second, in the event of default the lender wants to sell the collateral and we assume that a lender can only realize the price p by selling immediately.[5] Constraint (10) also features a haircut η, which, in practice,

[5]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,

reflects lenders' beliefs about the risk with respect to being able to use the underlying collateral to make up for any deficiency in repayment. Haircuts do not immediately respond to changes in this underlying risk (see Gorton and Metrick, 2009 and Copeland, Duffie, Martin and McLaughlin, 2012). Accordingly, we take this haircut to be a constant for simplicity.

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:

$$-\psi\left(e-\underline{e}\right)+e\left(mp^d+(1-m)p\right)+(1-e)\left[\alpha\left(mp^d+(1-m)p\right)+(1-\alpha)p\right]. \tag{11}$$

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 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:

$$V^u\left(q^u\right) := q^u p r^x - D^u r. \tag{12}$$

Feasible choices for D^u satisfy the leverage constraint:

$$\eta p q^u \leq K. \tag{13}$$

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.

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:

$$\begin{aligned} V^i\left(q^i, y\right) &= q^i p^d R + \left(y - q^i\right) p r^x - D^i r \\ &= q^i p^d (R - r) + p(y - q^i)(r^x - r) + Kr \end{aligned} \tag{14}$$

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

$$\eta p y \leq K - (p^d - p) q^i. \tag{15}$$

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:

$$D^i = p^d q^i - K \leq (1 - \eta) p q^i. \tag{16}$$

Note that the constraint ((16)) 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: quantities

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

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

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:

$$Mq^u + N(q^i + y) = 1$$

and

$$MD^u + ND^i = 0.$$

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 $r = r^x$. To be sure, 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 (strict) 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 with maximal leverage such that $R > r > 1$ and $p > 0$.

Our analysis focuses mostly on strict CIM equilibria. In a strict CIM equilibrium there is market segmentation, with informed investors trading in the OTC market and uninformed investors trading on the exchange. In addition, informed investors act as intermediaries and borrow from uninformed investors, as $R > r$ and $r^x = r$. We illustrate the financial flows that obtain in a strict CIM equilibrium in Figure 1.

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 $p > 0$ and $m < 1$.

In the next section we establish the existence and uniqueness of a strict CIM equilibrium 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 $R > r^x > 1$. Moreover, 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 $\tilde{\alpha} > 0$ and suppose that*

$$\frac{\underline{e}\tilde{g}\kappa x_h}{\tilde{g} - \underline{e} + \underline{e}\kappa} < \tilde{K} < \underline{e}x_h. \tag{17}$$

Then there exists a neighborhood $\mathcal{N}$ of $(\tilde{K}, 0, \tilde{\alpha})$ such that for $(K, N, \alpha) \in \mathcal{N}$, $N > 0$ and $\alpha \geq 0$ there is a unique equilibrium, which is a strict CIM equilibrium.

Proof: This follows from Propositions A.8 and A.9 in the Appendix. □

The inequality (17) puts an upper and lower bound on the amount of capital in the economy for a strict CIM equilibrium to exist. 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 prefer to deploy their capital on the exchange. Second, K cannot be too high for then there could be so much liquidity in the economy that assets trade at their fair value p^f.

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 ((16) 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 1.

4 Comparative Statics with respect to K

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. To this end we derive key comparative statics properties of CIM equilibria as a function of K. In particular, we characterize how asset prices in the exchange, p, origination incentives, e, and leverage of financial intermediaries, D^i, vary with K.

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 the net effect of an increase in K is an increase in the price p.

Proposition 2 *There exists an $\bar{\epsilon} > 0$ such that if $NK_2 < \bar{\epsilon}$ and if there are continuous functions $p(K)$ and $e(K)$ defined in an interval (K_1, K_2) such that $(p(K), e(K))$ is a strict CIM equilibrium for parameters (K, N, α), then the price in the exchange $p(K)$ is an increasing function of K.*

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

Note that a stricter bound than in Proposition 1 is required to guarantee that prices in the exchange are an increasing function of K, specifically that the total amount of capital in the hands of *informed intermediaries*, NK, is sufficiently small. 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.

Propositions 1 and 2 are illustrated in Figure 2. 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.[6] As K rises beyond this low value informed intermediaries returns R grow larger and larger relative to uninformed investors' returns r^f. But, note that both returns decline as more savings get channeled into asset markets. In sum, Figure 2 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

As we have already hinted at above, the asset price changes induced by changes in K, dp/dK, 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. More precisely, since: i) $p \leq gx_h$; ii) $\psi(.)$ is convex, with $\psi(0) = \psi'(0) = 0$; and, iii) $\psi'(1 - \underline{e}) > x_h > \kappa g x_h$, the uniquely optimal origination effort $e < 1$ satisfies the first-order condition:

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

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

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

1. The precision of intermediaries' information about asset quality as captured by the term $(1-\alpha)$. The higher the precision $(1-\alpha)$ with which an asset yielding x_h is identified the higher are origination incentives.

2. Originators' market incentives are given by the prospect of selling a high quality asset at price p^d to an informed intermediary rather than at price p. But to be able to sell an asset at price p^d it is not sufficient to originate a high quality asset. Conditional on producing such an asset, the originator must also get an offer from an informed intermediary. This occurs with probability m, so that the higher is the matching probability m the stronger are origination incentives.

3. Finally, the size of the spread $\left(p^d - p\right) = \kappa(gx_h - p)$ naturally affects origination incentives. Intuitively, 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.

Given that asset prices p increase with aggregate savings K in a strict CIM equilibrium, the spread $\left(p^d - p\right)$ is decreasing with aggregate incentives.[7] This is the main reason why savings gluts undermine origination incentives. Savings gluts are associated with spread compression, which reduces incentives to bring higher quality assets to the market.

But there is an important countervailing effect through the matching probability m. As we show in Figure 3 Panel B, the matching probability m may well be increasing in K. In the example plotted in Figure 3 we assume that $\psi(e) = \theta\frac{e^2}{2}$, with $\theta = 0.25$; $\kappa = 0.15$; $\eta = 0.5$; $M = 0.75$, and $x_h = 5$. The underlying economic reason why m may be increasing in K is that the capital that informed intermediaries can deploy increases with K both directly–as intermediaries' own capital NK is increasing–and indirectly, as intermediaries' borrowing capacity increases, and their borrowing costs decrease with p. As a result, intermediaries may be able to purchase more high quality assets, thereby increasing m and origination incentives.

When m increases with K, as in the example, the net effect of an increase in savings K on origination incentives e is such that the effect through m dominates at low levels of K,

[7] There is a second order effect through the dependence of the spread on g, as defined in (3). The lower is α the smaller is this second order effect.

and the effect through the spread $\left(p^d - p\right) = \kappa(gx_h - p)$ dominates at high levels of K, so that on net e at first rises with K and then decreases with K, as shown in Panel A of Figure 3. In other words, $e(K)$ is a non-monotonic function of K. This non-monotonicity of $e(K)$ is a robust outcome of the model, and we have identified sufficient conditions under which it holds in the proposition below. Essentially, as long as informed intermediaries' capital is not too large or, alternatively, 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) Let* $K_1 < K_2$ *and suppose that there are continuous functions* $p(K)$ *and* $e(K)$ *defined in* (K_1, K_2) *such that* $(p(K); e(K))$ *is a strict CIM equilibrium for parameters* (K, N, α) *with* $K_2 N < \bar{\epsilon}$*. Then if* $\eta < \kappa$ *then (a) If* K_1 *is small enough,* $e'(K_1) > 0$*; (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, suppose that $\kappa = 0$; this would imply that $\left(p^d - p\right) = 0$, so that there would be no origination incentives at all. Observe also that when the haircut η is larger, financial intermediaries can borrow less, so that m is lower other things equal. To make up for the lower m the spread $\left(p^d - p\right)$ must be larger to preserve incentives, which explains why κ must also be larger when η is larger.

An additional economic effect that complicates the analysis is that when K increases, so do asset prices. This means that, although intermediaries' debt capacity always increases with K, m may not necessarily increase, as asset prices could rise more than intermediaries' debt capacity. Still, the robust result is that under the sufficient condition in 3 $e(K)$ is non-monotonic.

Proposition 3 is a central result of our analysis. It provides a powerful explanation based on economic fundamentals for why late in a lending boom (or cycle) origination incentives deteriorate. This phenomenon is commonly observed in episodes shortly preceding the onset of a financial crisis and it has puzzled economic historians (see Kindleberger and Aliber, 2005).

A popular explanation, most notably by Minsky (1992), is based on investor psychology, and emphasizes the growing risk appetite, to the point of recklessness, of investors over the lending boom. We do not need to appeal to such 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.

The non-monotonicity of $e(K)$, in turn, explains the non-monotonicity of $p^f(K)$ (see Panel A of Figure 2). There are two effects influencing the average quality of assets traded in the exchange: Origination effort and cream skimming.[8] The latter effect always reduces the quality of assets traded in the exchange. But when $e(K)$ increases with K the overall increase in asset quality at first outweighs the effects of cream-skimming, so that the non-monotonicity of $e(K)$ also translates into a non-monotonicity of $p^f(K)$. Under cash-in-the-market pricing, however, the expected payoff of assets traded on the exchange is delinked from asset pricing: The price of the average asset traded in the exchange $p(K)$ rises with K even though the expected payoff $p^f(K)$ decreases. As a result, the deterioration of origination standards is masked by the monotone narrowing of spreads. This prediction of our model is consistent with the evidence of Krishnamurthy and Muir (2015).

We show next that savings gluts are also accompanied by increasing financial fragility.

4.3 Intermediary Financial Fragility

In the strict CIM equilibrium informed intermediaries exhaust their borrowing capacity, so that (16) is met with equality. When aggregate savings K increase so do asset prices $p(K)$, which relaxes the constraint (16), thereby increasing intermediary leverage. More formally, we define leverage as follows:

$$\mathcal{L}_{BE} = \frac{\text{Total assets}}{\text{Book equity}} = \frac{K + D^i}{K} = 1 + \ell^i \qquad \text{with} \qquad \ell^i := \frac{D^i}{K}.^9$$

[8] We studied the effect of cream skimming in Bolton, Santos and Scheinkman (2016).

[9] 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 have to immediately mark down any assets they acquire,

The next proposition shows that an increase in K induces an increase in leverage $\ell^i(K)$.

Proposition 4 *(Leverage) There exists an $\hat{N}$ such that if ($\widetilde{p}$; $\widetilde{e}$) is a strict CIM equilibrium for parameter values $(\tilde{K}, N, \alpha)$, with N sufficiently small, then $\mathcal{L}_{BE}$ 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 emphasize their finding that there is: "a strongly positive relationship between changes in leverage and changes in the balance sheet size." [Adrian and Shin, page 419, 2010]

What is more, to the greater fragility on the liability side, induced by the higher leverage, also corresponds a greater fragility on the asset side of intermediaries' balance sheets. This is due to the deterioration in origination standards, which also affects the quality of assets distributed to informed intermediaries. This effect is far from obvious. After all, intermediaries are able to identify high quality assets through their informational advantage. Yet, the fraction g =prob$(x_h|\sigma_h)$ of assets paying off x_h acquired by intermediaries as given in (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 distributed in the OTC market increases. This is shown in Figure 4. Panel A illustrates how the asset side of intermediaries' balance sheet mirrors the non-monotonicity of origination effort in K. Panel B illustrates Proposition 4: As aggregate savings K increase so does intermediaries' leverage. Thus, for a while there is a positive correlation between leverage and asset quality on intermediaries' balance sheets, but once aggregate savings reach such a high point that there is a glut compressing spreads and eroding origination incentives, the correlation turns negative.

which is counterfactual.

Moreover, 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, $\ell^i = D^i/K$.

4.4 Implications

The savings glut pinpointed by Bernanke (2005) and commonly proposed as a fundamental cause of the financial crisis conjures the image of a financial system that is not equipped to absorb vast pools of new savings. But beyond this image there is limited appreciation of the precise mechanisms that lead an economy awash with liquidity to a financial crisis. Our model and formal analysis is an attempt to uncover these mechanism and thereby gain sharper policy implications.

The central mechanism in our model is the effect of the savings glut in compressing spreads, and thereby undermining origination incentives. The growing demand for assets is met with greater supply of lower quality assets. Our key observation is that cash-in-the-market pricing masks the effect of spread compression and the deterioration of origination incentives. Informed intermediaries believe that their cream-skimming of high quality assets is unaffected by the savings glut, although the underlying pool from which they are selecting assets is worse as a result of poorer origination. And uninformed investors are affected by changes in the cash-in-the-market price p, which does not reflect the deteriorating fair-value price of the assets on the exchange p^f. In other words, as a result of the savings glut, asset prices on the exchange rise even when the fair value of assets falls.

A second, reinforcing, mechanism is through the rise of leverage of financial intermediaries caused by the savings glut and the higher collateral values it induces. This expansion of intermediaries balance sheets through increased leverage occurs just as underwriting standards of originated assets deteriorate, resulting in greater financial fragility of the financial intermediary sector.

These main predictions of our model are broadly in line with developments in the financial sector in the run-up to the crisis of 2007-09: Inflows of greater emerging market savings into global asset markets produced the rise in asset prices, a compression of yields, an expansion of repo markets and bank wholesale funding along with a deterioration of mortgage origination standards and greater fragility of financial intermediaries.

Of course, some financial institutions, such as Lehman Brothers, Merrill Lynch and Bear Stearns, were more aggressive in expanding their balance sheets, to the point where they ultimately failed. One intriguing possibility in terms of our model is that the 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.

4.5 Other 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.

4.5.1 Distribution of capital and knowledge

How are asset prices and origination incentives affected when the increase in liquidity is concentrated within the financial intermediary sector? We can address this question by looking at the comparative statics with respect to N, the measure of informed traders. Indeed, by increasing N we increase the relative liquidity of the intermediary sector, in equilibrium.

Proposition 5 *Let N_2 and K be such that $N_2K < \bar{\epsilon}$ and $N_1 < N_2$. Suppose there exists continuous functions $p(N)$ and $e(N)$ for $N \in [N_1, N_2)$ such that ($p(N); e(N)$) is a strict CIM equilibrium for parameters (K, N, α). Then $p(N)$ is decreasing and $e(N)$ is increasing in N.*

Proof: Follows from Lemma A.2. □

In words, asset prices on the exchange p are a decreasing function of the proportion of capital held by intermediaries (and therefore, since $N = 1 - M$, an increasing function of the proportion of savings held by "dumb" investors). Moreover, origination incentives, e, are an increasing function of the proportion of capital held by intermediaries, N. Intuitively, an increase in intermediary capital results in a higher probability m of selling a high quality asset to an intermediary and also an increase in the spread $(p^d - p)$, so that origination incentives are improved.

4.5.2 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 need to characterize the comparative statics with respect to η. As the proposition below establishes, a lower haircut allows informed intermediaries to borrow more, resulting in higher asset prices on the exchange and lower origination incentives.

Proposition 6 *Let N and K be such that $NK < \bar{\epsilon}$, and suppose that there exist continuous functions $p(\eta)$ and $e(\eta)$ that correspond to strict CIM equilibria for parameter values (K, N, α, η), $\eta \in [\eta_1, \eta_2]$. 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. Instead, 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.

5 Robustness

5.1 Endogenous origination volume and distribution

A simplifying assumption in our analysis so far has been that the total volume of originated assets is fixed and completely 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. As this increased real origination volume was not sufficient to quell the market's thirst for new mortgage-backed securities, it was further augmented at the peak of the cycle by a larger and larger volume of synthetically created assets, CDOs and CDO^2s.

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 thus 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[10].

5.2 Overconfidence and 'bad apples'

Our model can also be extended to introduce overconfident investors. Consider, for example, the equilibrium situation where informed investors have a screening technology with $\alpha = .2$, but a single, atomistic, informed investor believes that her $\alpha = 0$. That is, this *optimistic* investor believes that the risk management systems in place guarantee that only good assets enter the balance sheet. The 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).

Note first 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 $\mathcal{L}^*_{BE} > \widetilde{\mathcal{L}}_{BE}$, as the optimistic investor would have less collateral to post in the repo market. In other words, 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

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

[10] 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).

whereas the true level of capital is

$$E^2_{true} = \widetilde{q} g^* x_h - r^* \widetilde{D} < E^2_o.$$

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

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

In sum, optimistic investors have smaller balance sheets and take on less leverage, but 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 5 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.

There is evidence that during the credit bubble many financial intermediaries, although fully cognizant of the link between home price appreciation (HPA) and MBS values, did not consider possible the sharp nationwide negative scenarios of HPA 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 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 expected payoff should warrant. 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 5 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 $\alpha = .4$.

5.3 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 phenomenon. But, an alternative account 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 attitudes of uninformed investors. When the discount rate of uninformed investors fluctuates 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 Conclusion

We have developed an extremely simple model 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 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.

Asset originators distribute assets across the two markets. 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 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 the level of aggregate liquidity 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 sheet of financial intermediaries becomes 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 more of the balance sheet is financed with leverage. Indeed, we show that leverage is increasing in aggregate liquidity conditions. Thus as liquidity becomes abundant, there are worse assets on the intermediaries balance sheets with more leverage. Our model thus offers a particularly simple account, based on a straightforward economic mechanism, of the years leading up to the Great Recession. In essence, we argue that the savings glut emphasized by Bernanke (2005) in his account of low interest rates in the run-up to the crisis was amplified by the financial sector, as suggested by Borio and Disyatat (2011), and that this resulted in weaker origination standards and a fragile financial system.

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.

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.*

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*, forthcoming.

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

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

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.

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

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

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.

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

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.

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

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 (2014) "Antes del Diluvio: The Spanish banking system in the first decade of the euro," manuscript, Columbia University.

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.

Tokunaga, Junji and Gerald Epstein (2014) "The Endogenous Finance of Global Dollar Based Financial Fragility in the 2000s: A Minskian Approach," University of Massachusetts Amherst, Political Economy Research Institute Working Paper Series, no. 340.

Figure 1: A graphical representation of the model

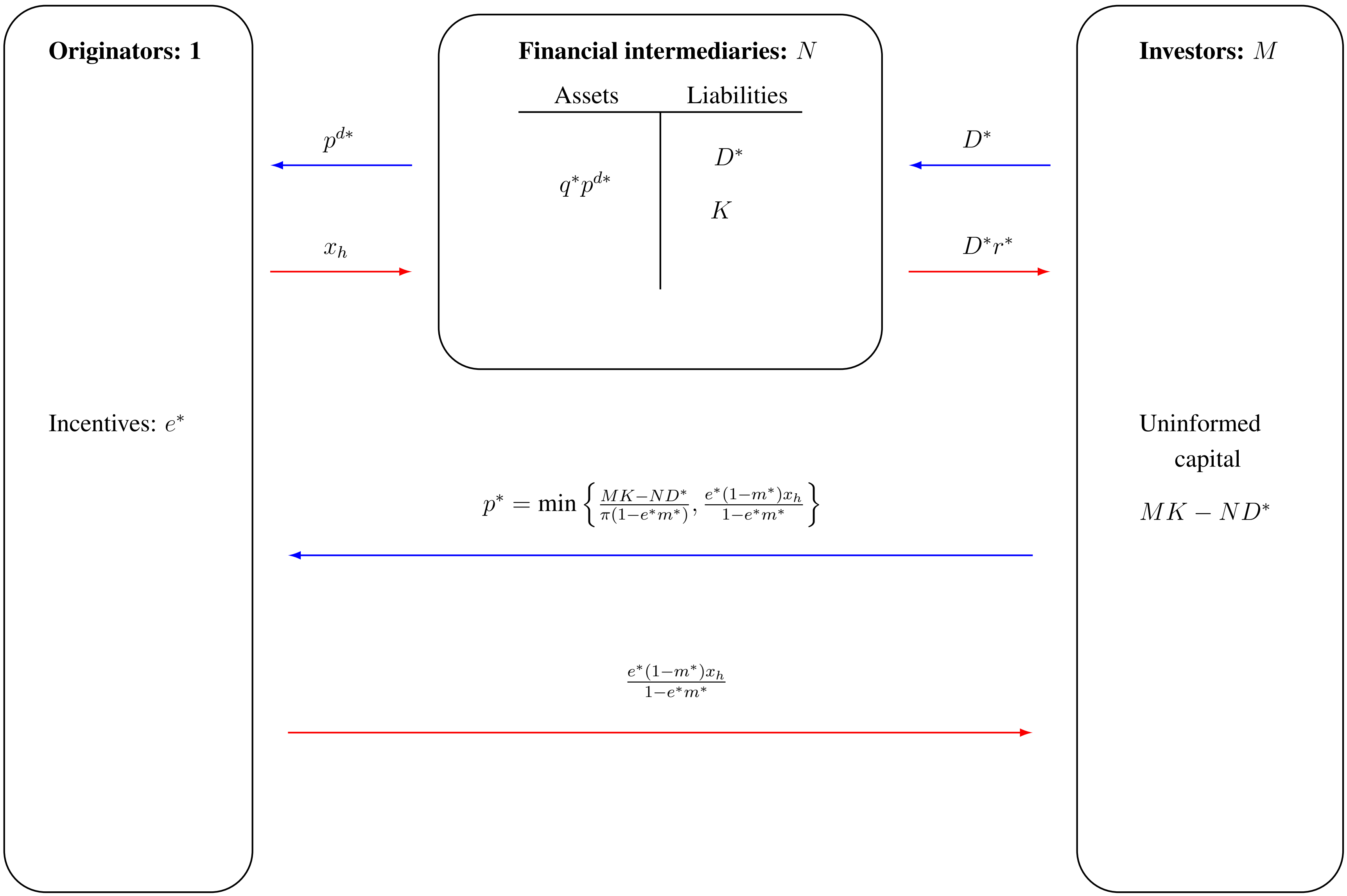


Figure 2: Panel A: Prices in the exchange p and expected payoff (dashed line) as a function of capital, K,. Panel B: Expected rate of return in the opaque market, R, expected rate of return in the exchange $r^x(= r)$; the horizontal line is set at one. In this example, $\psi(e) = \theta\frac{e^2}{2}$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

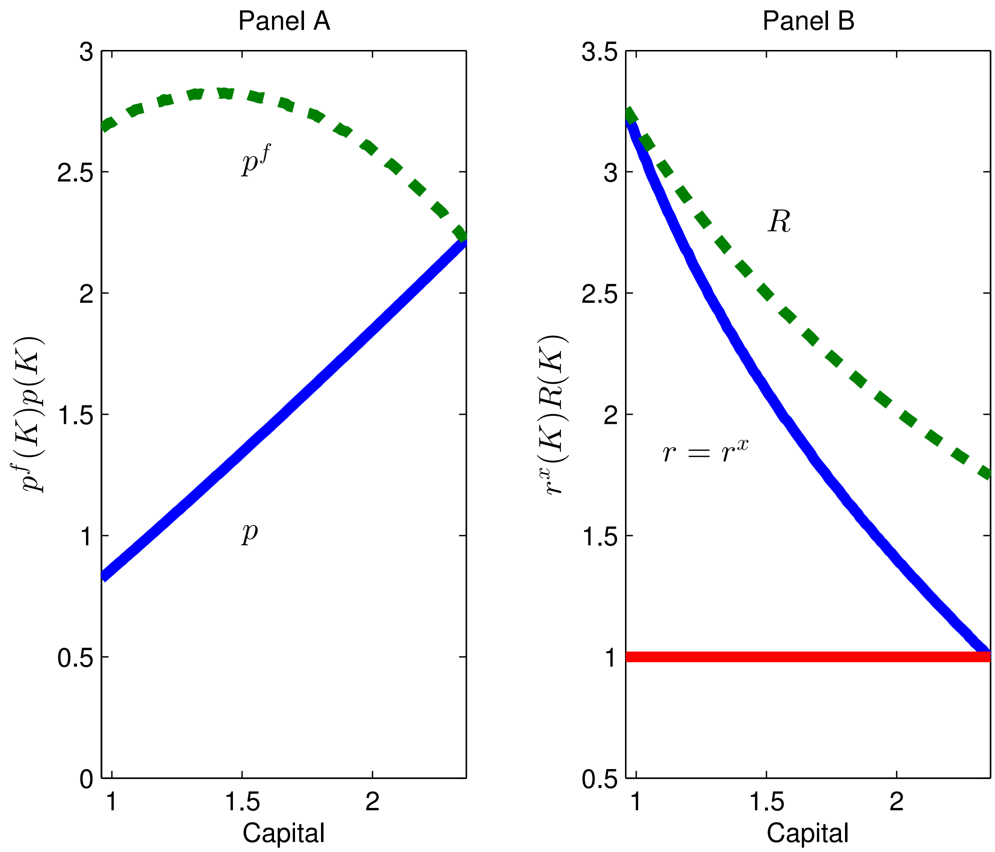


Figure 3: Panel A: Origination effort e, as a function of capital, K, for the cases $\alpha = 0$ (dashed line, in green) and $\alpha = .2$ (continuous line, in blue). Panel B: $x_h p$, as a function of capital K (continuous line, in blue, left axis) and matching probability m as a function of capital, 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 $x_h = 5$.

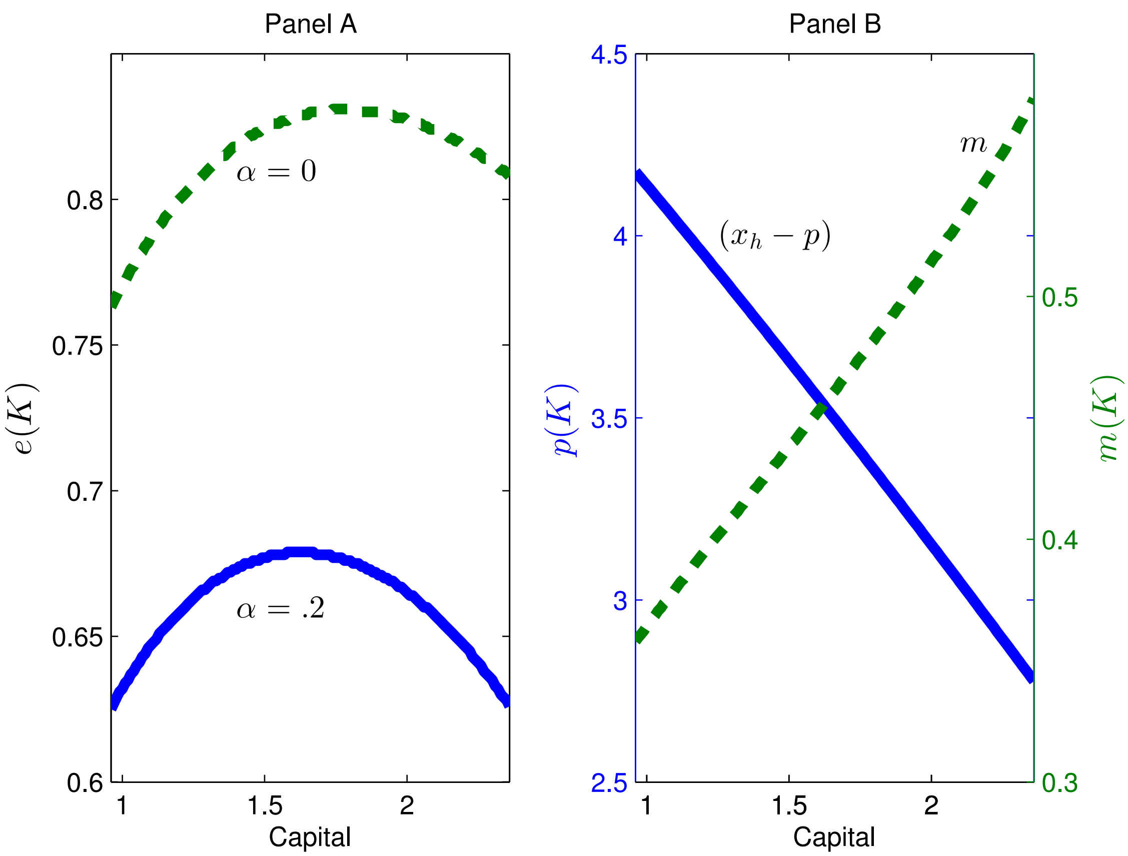


Figure 4: Fragility. Panel A: Fraction of high payoff assets in the balance sheet of financial intermediaries, g, as a function of capital, K. Panel B: Leverage, $\ell := D/K$, as a function of capital K for the cases $\alpha = 0$ (dashed line, in green) and $\alpha = .2$ (continuous line, in blue). In this example, $\psi(e) = \theta\frac{e^2}{2}$ with $\theta = .25$. In addition, $\kappa = .15$, $\eta = .5$, $M = .75$ and $x_h = 5$.

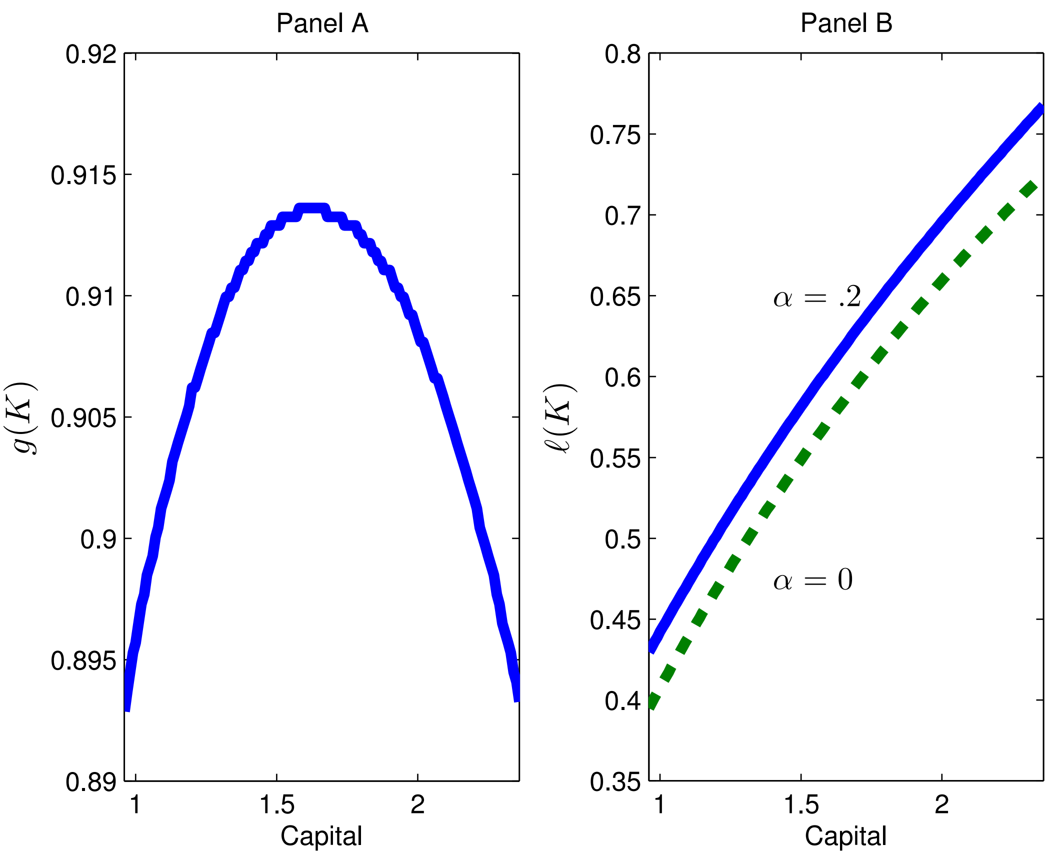
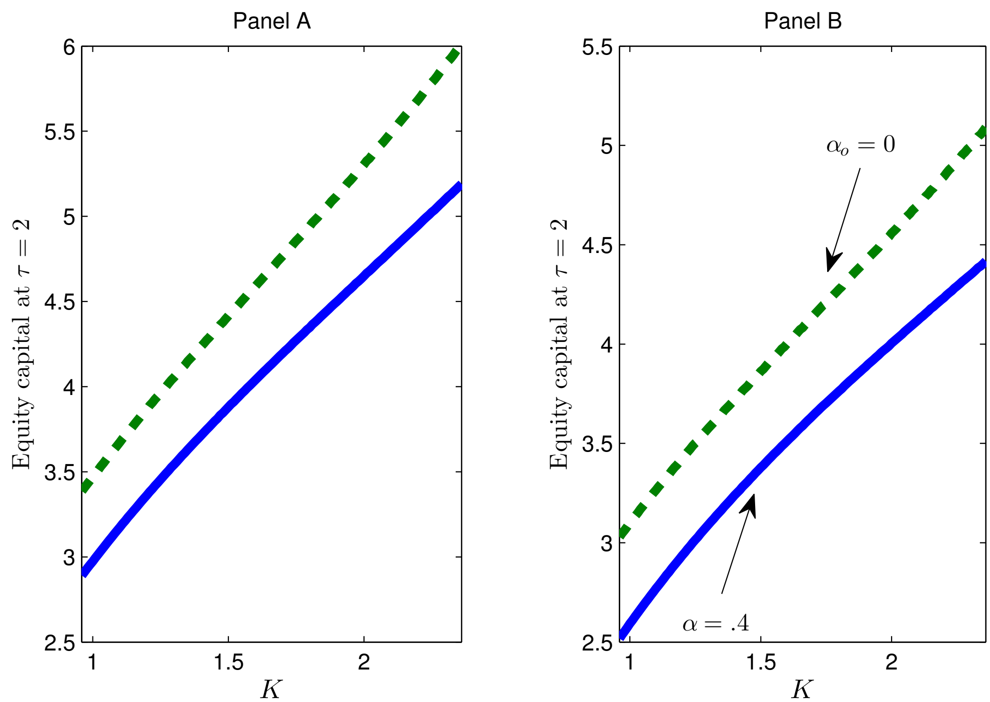


Figure 5: Panel A: Equity capital under the wrong beliefs, E_o^2 (dashed line) and the true equity capital at date 2, E_{true}^2, of the single optimistic intermediary who believes $\alpha = 0$ when $\alpha = .2$. Panel B: Equity capital when the entire intermediary sector believes that $\alpha_o = 0$ 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 $x_h = 5$.



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

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

$$m := \min\left\{\frac{Nq^i}{e+\alpha\,(1-e)}, 1\right\} \tag{A.2}$$

$$g := \frac{e}{e+\alpha\,(1-e)} \tag{A.3}$$

$$r^x := \frac{e(1-m)x_h}{p\,(1-m\,(e+\alpha\,(1-e)))} \quad \text{and} \quad R := \frac{gx_h}{p^d} \tag{A.4}$$

$$D^u := pq^u - K \tag{A.5}$$

$$D^i := p^d q^i - K \tag{A.6}$$

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

$$D^u \leq 0 \tag{A.7}$$

$$D^i = (1-\eta)\, pq^i \tag{A.8}$$

$$p = \min\left\{\frac{MK - ND^i}{1-m\,(e+\alpha\,(1-e))}, \frac{e(1-m)x_h}{1-m\,(e+\alpha\,(1-e))}\right\} \tag{A.9}$$

$$\psi'\,(e-\underline{e}) = (1-\alpha)\, m\kappa\,(gx_h - p) \tag{A.10}$$

$$r = r^x < R \tag{A.11}$$

$$Mq^u + Nq^i = 1 \tag{A.12}$$

$$MD^u + ND^i = 0 \tag{A.13}$$

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

$$\begin{aligned} q^i &= \frac{K}{\kappa\left(gx_h - p\right) + \eta p} && \text{(A.14)} \\ D^i &= \frac{(1-\eta)pK}{\kappa\left(gx_h - p\right) + \eta p}. && \text{(A.15)} \end{aligned}$$

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 $\boldsymbol{\theta} = (N, \alpha)$. 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

$$\begin{aligned} f^1\left(p, e, K, \boldsymbol{\theta}\right) &:= p - K\left(M + N\gamma\right) = 0 && \text{(A.16)} \\ f^2\left(p, e, K, \boldsymbol{\theta}\right) &:= \left(e + \alpha\left(1-e\right)\right)\psi'(e - \underline{e}) - (1-\alpha)\, NK\beta = 0 && \text{(A.17)} \end{aligned}$$

where

$$\begin{aligned} \gamma &:= \frac{\eta p}{\kappa\left(gx_h - p\right) + \eta p} && \text{(A.18)} \\ \beta &:= \frac{\kappa\left(gx_h - p\right)}{\kappa\left(gx_h - p\right) + \eta p} = 1 - \gamma, && \text{(A.19)} \end{aligned}$$

Notice that

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

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

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

Suppose that (p, e) with $e \geq \underline{e}$ solves $f\left(p, e, K, \boldsymbol{\theta}\right) = 0$, with $N \geq 0$ and $0 \leq \alpha \leq 1$. 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 with maximum leverage from a zero of f it suffices to apply this construction and verify that $R > r := r^x > 1$. In this case, we will call (p, r) *the strict CIM equilibrium associated with* (p, e). This construction also shows that there is a unique strict CIM equilibrium that would generate the same (p, e) and furthermore, 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 $(p, e, K, \boldsymbol{\theta})$.

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 $\psi_z(z) \leq 0$ for $z < 0$.

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

$$\begin{aligned}
f_p^1 := \frac{\partial f^1}{\partial p} &= 1 - NK\gamma_p && \text{(A.23)}\\
f_e^1 := \frac{\partial f^1}{\partial e} &= -NK\gamma_e && \text{(A.24)}\\
f_p^2 := \frac{\partial f^2}{\partial p} &= (1-\alpha)NK\gamma_p && \text{(A.25)}\\
f_e^2 := \frac{\partial f^2}{\partial e} &= (1-\alpha)\psi' + (e + \alpha(1-e))\,\psi'' + (1-\alpha)NK\gamma_e && \text{(A.26)}
\end{aligned}$$

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

$$\begin{aligned}
f_K^1 := \frac{\partial f^1}{\partial K} &= -\left(M + N\gamma\right) < 0 && \text{(A.28)}\\
f_K^2 := \frac{\partial f^2}{\partial K} &= -\left(1-\alpha\right)N\beta \le 0 && \text{(A.29)}
\end{aligned}$$

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:*

$$\begin{aligned}
f(p, e, K, N, \alpha) &= 0 \\
f_e^2(p, e, K, N, \alpha) &> 0 && \text{(A.30)}\\
|\partial_{p,e} f(p, e, K, N, \alpha)| &> 0. && \text{(A.31)}
\end{aligned}$$

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,

$$f_p^1 \quad \geq \quad 1 - \max\{\eta, \kappa\} > 0, \tag{A.33}$$

$$f_p^2 \quad \leq \quad (1-\alpha)\max\{\eta, \kappa\}. \tag{A.34}$$

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

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

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 $p_N < 0$ and $e_N > 0$.*

Proof: This follows from the previous Lemma, since

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

□

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 $f^1_K < 0$ and $f^2_K \leq 0$.

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

$$\mathcal{A}(K) := \frac{f^2_p(p(K),e(K))f^1_K(p(K),e(K)) - f^1_p(p(K),e(K))f^2_K(p(K),e(K))}{1-\tilde{\alpha}}$$

and thus

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

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

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

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 $\mathcal{L}'_{BE}\left(\tilde{K}\right) > 0$. (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 $\mathcal{L}'_{BE}\left(\tilde{K}\right) > 0$*

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,

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

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

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

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

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

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

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

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 $g_e \le 1$

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

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 $\tilde{\alpha} \ge 0$,

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

Proposition A.8 *(Existence) Suppose that*

$$\frac{\underline{e}\tilde{g}\kappa x_h}{\tilde{g} - \underline{e} + \underline{e}\kappa} < \tilde{K} < \underline{e} x_h. \qquad \text{(A.37)}$$

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 the maximum leverage CIM equilibrium (p, r) associated with $(p(K, N, \alpha), e(K, N, \alpha))$ is well defined, and hence (p, e) is a CIM equilibrium for the parameter values $(K, N, \alpha) \in \mathcal{N}$.

Proof: It is easy to check that $f(\bar{K}, \underline{e}, \bar{K}, 0, \tilde{\alpha}) = 0$. Since $|\partial_{p,e} f(p, e, \bar{K}, 0, \tilde{\alpha})| > 0$, the implicit function theorem guarantees that there exist a neighborhood $\mathcal{N}$ of $(\bar{K}, 0, N, \tilde{\alpha})$ and $\epsilon > 0$ such that for each $(K, \boldsymbol{\theta}) \in \mathcal{N}$ there exists a unique $(p, e) > 0$ with $|(p, e) - (\bar{K}, \underline{e})| < \epsilon$ 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), construct a CIM equilibrium with maximum leverage from the solution $(\bar{K}, \underline{e})$ at $(\bar{K}, 0, \alpha)$ by setting $g = \tilde{g}$ and using equation (A.14) to calculate the implied q. Then it is easy to check that $m = 0$ solves (A.2), and (A.1) and (A.4) can be used to yield the implied $p^d, r^x$and R. 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 remaining of the proof is as in the proof of Lemma A.5. The second inequality in (A.37) insures that $r^x > 1$. Thus using the continuity of the implied CIM with maximum leverage equilibrium 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 by the CIM equilibrium with maximum leverage associated with $p(K, N, \alpha), e(K, N, \alpha)$ whenever $N > 0$ and $\alpha \geq 0$. Hence (p, e) is a 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) CIM equilibrium with maximum leverage*

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 maximum leverage must hold. 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, $(p_n, e_n) = (p(K_n, N_n, \alpha_n), e(K_n, N_n, \alpha_n))$. $\square$