Fire sales, pecuniary externality and inefficient banking

Axelle Arquié*

November 27, 2013

[Link to the latest version]

Abstract

To investigate the social cost of fire sales, I build a banking model *à la* Diamond and Dybvig in which an aggregate liquidity shock hits consumers preferences before asset returns are realized. Banks can sell assets to mutual funds when this shock is too high to pay the depositors hit by the shock. The wealth of the buyers of assets, the mutual funds, results from a portfolio allocation by households between bank deposits and mutual funds shares. I identify one combination of imperfections that gives rise to a pecuniary externality implying a welfare loss: an asymmetry of information between banks and depositors regarding depositors' liquidity needs on top of an incomplete market feature. Distortions lay both in the banks' choice, between assets and reserves, and in the households' choice, between bank deposits and mutual funds shares. High liquidity shock triggers bank default for low level of funds wealth. The price then falls below the fundamental value at a cash-in-the-market value. As banks do not realize that such a fire sale situation might occur, they do not optimally insure depositors against the idiosyncratic liquidity risk. It implies both an *ex ante* and an *ex post* welfare loss. Whereas efficiency would require banks to keep a liquidity buffer, they do not keep enough reserves and invest too much in assets. Imposing liquidity ratios allows to get closer to efficiency but not to reach the constrained efficient allocation. Indeed, liquidity ratios cannot help alleviate the distortion laying in the households' choice.

*PhD candidate in Paris 1 Sorbonne and Paris School of Economics

1 Introduction

In response to the liquidity crisis experienced by some financial institutions in 2008, the new Basel III regulatory framework introduces liquidity ratios aiming at reducing the distortions arising from fire sales. The premise is that banks do not manage correctly their liquidity and do not keep enough high quality liquid assets to be able to face the onset of a liquidity crisis. During such an aggregate liquidity crisis, a large number of banks all suddenly face an increase in their cash outflows, leading to a massive liquidation of assets and a collapse of price. Liquidity dries out precisely when many banks conjointly need it. Fire sales have been theoretically defined by Shleifer and Vishny (1992) as a "*forced sale at a dislocated price*". In a frictionless world, fire sales only have mere innocuous redistributive effects between sellers and buyers. They do not imply any welfare loss in a setting without frictions. However, in presence of microeconomic imperfections, these fire sales can imply a social cost, as explained in Greeenwald and Stiglitz (1986).

My model investigates how the mechanism at play in fire sales can result in inefficiencies focusing on banking. I build a banking model in which an aggregate liquidity shock hits the consumers preferences. If hit, consumers only care about present consumption and cannot postpone consumption; they are impatient. The unlucky outcome is to be hit by such shock and liquidity type is private information. Before the shock hits, *ex ante* identical households allocate their portfolio between bank deposits and shares of mutual funds. Bank deposits provide depositors with some degree of insurance against this liquidity risk, by pooling all households' endowment (Diamond and Dybvig 1983). Banks promise to serve impatient depositors a non contingent rate. Funds shares can provide a higher return but only benefit to patient consumers as their profits are realized in the last period. Before the shock, banks decide how much to invest in assets and how much to keep in liquid reserves. If the liquidity shock is such that banks do not hold enough reserves to pay impatient consumers the fixed promised rate, they can sell some of their assets to funds at an endogenous price.

In this setting, my contribution is twofold. First, I show that a pecuniary externality arises and reduces welfare because of both incomplete markets and an asymmetry of information between banks and depositors regarding depositors' liquidity needs. I show that this inefficiency lays both in the banks' choice between assets and reserves; and in the households' choice between bank deposits and mutual funds share. Second, I show that a binding liquidity ratio increases welfare but does not allow to reach the efficient constraint allocation. Indeed, liquidity ratios are only a partial remedy to the pecuniary externality: it helps alleviating the inefficiency laying in banks' choice but not in the households' choice. Formally, I contribute to the literature by providing a framework in which a social computation of welfare is immediate and does not require the definition of any weights to attribute arbitrarily to agents. Indeed, an *ex ante* analysis of welfare is straightforward as all households are *ex ante* identical and as all profits *in fine* accrue to them. Funds give back profits to patient consumers in the last period, not by assumption, but because patient consumers are the only ones to still care about consumption at that time. Besides, my model allows to endogenize the wealth of funds, i.e. the liquidity available to buy back assets sold by banks.

Funds are the buyers of the assets sold by banks and their wealth is fixed after the realization of the aggregate liquidity shock. Indeed, funds' wealth is endogenously determined by households before the shock. Funds have access to an additional technology of production, the *late* assets, whose returns are not shared with impatient consumers. While banks can only invest in *early* assets before the realization of the shock, funds can both benefit from the *early* asset technology if banks sell those assets, and invest in the *late* asset after the shock has hit. *Late* asset technology is not available to banks. In the first period, before the shock hits, households allocate their endowment between bank deposits - providing an insurance, and shares of mutual funds - offering returns which might be higher for high liquidity shock.

The externality arising in the model is a pecuniary and not a technological one. Indeed, funds buying back early assets from banks benefit from the exact same return as banks. There is no change of ownership cost so that any technological externality is ruled out. Any externality then goes through the price. For low levels of the funds wealth, the model features a cash-in-the-market pricing (Allen and Gale 2005) of the early assets sold by banks in case of default. Either the price is set at its fundamental value, determined by the respective productivity of early and late assets, either it falls below. When falling below, it is only determined by the ratio between funds wealth over the amount of early assets sold by banks to pay impatient depositors.

In the model, incomplete markets are necessary for the pecuniary externality to arise, in line with Allen and Gale (2004) findings. Markets are incomplete in that banks cannot insure against the aggregate liquidity shock - banks only provide an insurance against the *idiosyncratic* risk of being hit by the shock. There are no Arrow securities available at the beginning of the world, and the wealth available to buy back assets, the funds' wealth, is fixed whatever the realization of the liquidity shock. If markets were complete, the bank would never need to sell any assets whatever the size of the liquidity shock, ruling out any pecuniary externality.

On top of incomplete markets, an asymmetry of information creates inefficiency. Banks cannot distinguish the true liquidity needs of its depositors and do not know which consumer has been hit and which consumer is only pretending to be an impatient consumer. Because liquidity type is private information, the bank is subject to fundamental bank run. A fundamental bank run occurs when patient depositors know that they will receive less than impatient depositors and therefore withdraw early instead of waiting. The bank then needs to design an incentive compatible contract making sure that patient depositors do not misrepresent their type and pretend to be impatient. If banks cannot pay impatient consumers the promised rate *and* pay patient consumers at least the same rate, patient depositors run. The bank defaults when it can no longer satisfies this condition. Run and default are the two same notions in this setting: the possibility of bank run triggers default.

This no run condition depends on the price of assets sold by banks, which the bank incorrectly anticipates in the decentralized economy. Indeed, the bank is price taker on the secondary market of early assets. This price taking assumption is theoretically grounded on an atomicity argument but it also does make sense with regards to the international financial markets.

On the banks' side, I describe the social loss as inefficient banking, both *ex ante* and *ex post*. Banks wrongly anticipate both their probability of default and the rate they will be able to pay depositors in case of default. Indeed, both this probability and this rate depend on the price of sold assets, which is taken as given by banks. Banks fail to realize that the price can fall below its fundamental value to its cash-in-the-market value, for low level of funds' wealth and high liquidity shock triggering default. Therefore, banks make a non optimal choice between reserves and early asset and invest too much in assets and do not keep enough reserves. This is an *ex ante* cost, analyzed in terms of inefficient insurance: banks over estimate the expected rate they can serve impatient depositors.

An *ex post* cost also arises in case of default. Default occurs for high liquidity shock which means when there are a lot of impatient consumers. In case of default, funds buy back early assets at a price below the fundamental value, which generates a transfer of wealth from banks to funds. But this transfer of wealth implies a welfare loss: only patient consumers - who receive funds' profits - benefit from this transfer to the detriment of impatient consumers, precisely at a time when patient consumers are a few and impatient consumers are numerous, hence the *ex post* welfare loss.

On the households' side, I show that the portfolio allocation of households between bank deposits and funds shares is not optimal in the decentralized economy, leading to another efficiency than inefficient banking only. The same argument as for banks applies. Households do not correctly anticipate the probability of banks to default and the rate the bank can serve depositors when defaulting. They overestimate what the bank can pay them.

This paper belongs to the literature on banking. It is based on the seminal work by Diamond and Dybvig (1983) which defines the role of banks as provider of insurance against the idiosyncratic liquidity risk to risk adverse consumers. Whereas Diamond and Dybvig (1983) focus on banks runs defined as panics phenomenon, I will refer to fundamental banks runs as defined by Allen and Gale (1998).

In the corporate finance literature, many papers have focused on the social cost of fire sale arising mainly from a financial constraint (see Davila (2011) for a thorough review of literature) following the seminal work by Schleifer and Vishny (1992). See for instance Lorenzoni (2008), Hombert (2009).

In the banking literature, some papers have also focused on fire sales. Stein (2010) shows that private banks create too much money compared to the social optimum. Carletti and Agnolli (2013) investigate liquidity shortage in a model in which a mixed equilibrium can emerge with some banks defaulting and selling assets, and other banks buying back sold assets. They show that competition is beneficial to financial stability.

Allen and Gale (2004) show that an inefficiency arises because of an incomplete market feature and not because of incomplete contract alone. My paper does not aim at investigating the optimal form of financial arrangements but rather takes the institutional constraints as given. In this setting, the banking contract implies a non contingent promise by the bank to serve a fixed rate. As in Allen and Gale (1998), default allows to restore some contingency in the banking contract and is then optimal for some high realization of the liquidity shock. My model contributes to this literature by showing that if incomplete markets are necessary for the externality to arise, asymmetry of information between depositors and banks regarding the liquidity types of consumers makes it worse.

2 Model

I build a three-period (0, 1 and 2) banking model *à la* Diamond and Dybvig in which a stochastic liquidity shock hits consumers preferences in period 1. There are three types of agents: a mass 1 continuum of households, a representative bank and a representative fund. When the size of the shock is sufficiently large, some assets are sold by banks to funds at an endogenous price.

The timing is as follows. Period 0 is divided into two sub-periods. First, households allocate optimally their endowment between depositing in banks and buying profits' shares of funds. After the portfolio allocation of households has been made, banks receive deposits D. In the second sub-period of period 0, banks decide how much early assets S to invest and how much liquid reserves L to keep. They design an optimal incentive compatible banking contract which defines the promised rate $\overline{c}$ to serve in period 1 to withdrawers, before assets have matured. This fixed rate is non contingent, following the banking contract terms described further. Funds are passive during period 0, they only collect the wealth that households endow them with.

The households portfolio has a liquid and an illiquid component. Bank deposits can be withdrawn in period 1 whereas funds' profits share cannot. The latter are realized in period 2. Investing in funds is risky as households only receive the funds' profits shares if they are patient while bank deposits rather offer an insurance against liquidity risk.

In period 1, the liquidity shock hits the consumers. The liquidity shock is a source of both idiosyncratic and aggregate uncertainty. As in Diamond and Dybvig (1983), any consumers can be hit and runs the idiosyncratic risk of being an impatient consumer. The shock is also a source of aggregate uncertainty as the size of the shock, i.e. the fraction θ of consumers being hit, is stochastic. The shock distribution is drawn from a law that has a continuous probability distribution function, known by all agents, that we will choose to be uniform. After the realization of the shock, either the bank holds enough reserves to be able to pay its depositors the promised rate, or it needs to sell an amount $X(\theta)$ of total early projects S to funds. The latter buy back these early assets at a price $P(\theta)$ and invest $Y(\theta)$ in a new productive assets, the late assets.

There are three different technologies in the economy, with both different timings and different returns: early assets, late assets and storage. Storage provides 1 unit next period for 1 unit stored. Storage is available to both households, banks and funds[1]. Between period 1 and period 2, patient consumers can store until period 2 what they have withdrawn from the bank - a situation called a bank run.

Early assets are productive projects undertaken by banks in period 0 which mature in period 2. They provide a constant return to scale of R^E; whether sold to funds or kept until maturity[2]. A second productive technology called late assets is available to funds only in period 1. These late assets mature in period 2. They provide a constant return to scale of R^L where I assume:

Assumption 1

$$1 \leq R^L \leq R^E \tag{1}$$

Both early and late assets yield at least more than storage when held to maturity and late assets are less productive than early assets.

2.1 Households

There is a mass 1 of *ex ante* identical households on a continuum between 0 and 1. *Ex post*, households are not longer identical because in period 1, the liquidity shock hits a stochastic fraction θ of them. Consumers who are hit are called impatient and only care about period 1 consumption. Any consumption in period 2 would provide them a zero utility.

Consumers that are not hit are called patient consumers. They are of type 2 and only care about period 2 consumption. Patient consumers can either wait period 2 to withdraw when assets have matured or they withdraw early in period 1. The second option is called a bank run: patient depositors misrepresent their type and pretend to be impatient consumer in period 1. In that case, they do not consume immediately but store goods until period 2.

The utility function is twice continuously differentiable, increasing, strictly concave and satisfies Inada conditions $u'(0) = \infty$ and $u'(\infty) = 0$. Overall, the *ex ante* expected consumption of a given household is:

[1] Nevertheless, between period 0 and period 1, households will never store but rather deposit in banks: indeed, banks are maximizing their depositors' utility. So, if storage is optimal, bank will keep reserves for households. There is no loss in generality in assuming that households are not storing but rather depositing in a bank that will store if optimal between period 0 and period 1.

[2] So each asset provides a return to the bank of R^E if kept until maturity and of $P(\theta)$ if sold. As the technology is not impacted by this change of ownership (each early asset still yields R^E), funds get a return of $\frac{R^E}{P(\theta)}$.

$$U(c_1, c_2; \theta) = \begin{cases} u(c_1) & \text{if consumer is impatient in aggregate state } \theta \\ u(c_1 + c_2) & \text{if consumer is patient in aggregate state } \theta \end{cases}$$

where c_i is the amount withdrawn from bank in period i.

Households are endowed in period 0 with E units of goods. They optimally allocate between deposits D at banks or investment W in funds giving them right to profits shares in period 2 if they are not hit by the liquidity shock. Indeed, households do not get any direct payment from funds shares if they are impatient. Funds budget constraint is:

$$D + W = E$$

2.2 Banks

Banks receive the households' deposits D in period 0. As the banking sector is competitive, banks maximize the utility of their depositors, given the deposits D received. They invest S in early assets and keep L in reserves. Reserves L are put in storage. This is the component of the bank portfolio that is totally liquid and safe as it provides a certain return. Early assets mature in period 2 but can be sold to funds at an endogenous price $P(\theta)$. As will be made clearer later, this price can fall below its fundamental value. Therefore, early assets are risky and thus only partially liquid in the sense that they can yield a very low return in some states of the world.

In period 1, banks pay the promised rate $\bar{c}$ to consumers withdrawing and c_2 to patient consumers in period 2, if they can afford to do so without risking a bank run. Otherwise they default. The resource constraint of the bank is:

$$S + L = D$$

Banks provide an insurance against the idiosyncratic liquidity risk. The extent of this insurance depends on the risk aversion of the consumers. Nevertheless, banks cannot insure against the aggregate liquidity shock. There are no Arrow securities available in period 0 and the wealth available to buy back assets, funds wealth, is fixed after the realization of the shock. This is the incomplete market feature of the model. It is necessary for the externality to arise.

On top of this incomplete markets feature, an asymmetry of information between banks and depositors generates the pecuniary externality at play in the model. Banks cannot distinguish between patient and impatient consumers and are then subject to bank run. They are forced to design an incentive compatible contract to avoid bank runs. The contract must satisfy a no run condition applying the revelation principle: patient depositors must have an incentive not to misrepresent their type. When the liquidity shock is so high that the bank can no longer satisfy this condition, it defaults. Default and bank run are the same notion in this setting.

2.3 Funds

In period 0, funds receive the wealth W that households choose to invest in their profits' share. They are not active until period 1. At that time, they invest $Y(\theta)$ in late assets that mature in period 2 yielding R^L. If banks sell early assets, they also buy back an amount $X(\theta)$ for a price $P(\theta)$. Crucially, funds get the same return on these sold early assets as the banks would have. There is no change of ownership cost, ruling out any technological externality. The funds' budget constraint writes:

$$Y(\theta) + P(\theta)X(\theta) = W$$

Crucially, the wealth W households invest in funds is chosen in period 0 before the realization of the shock. Funds cannot go back on the market to raise more funds in case of a high liquidity shock. There is a missing market here.

3 Banking contract

I now turn to the description of the financial arrangement between banks and depositors. A crucial feature of this contract is that it is not complete. The paper does not aim at defining the optimal contract. Nevertheless, Allen and Gale (2004) have shown that incomplete contracts alone should not be a source of inefficiency.

Banking contract's terms are taken as exogenous here. Those terms stipulate that the bank has to promise a non-contingent payments $\bar{c}$ to *any* depositor willing to withdraw in period 1 as the bank cannot distinguish between types. Patient consumers not withdrawing in period 1 get whatever is remaining in period 2, after early assets have matured. When the bank does not default, it pays the promised rate on deposits:

$$\overline{c_1} = \overline{c}D$$

The amount promised cannot be made contingent on the realization of the liquidity shock. Even when the liquidity shock is large, banks need to pay the fixed promised rate. If it is not able to do so, using its own reserves or selling early assets, it needs to declare bankruptcy. Allowing for bankruptcy introduces some contingency in the banking contract that restore the inefficiency arising from its non contingent feature, as shown in Allen and Gale (1998).

The bankruptcy is a situation in which every depositors, whatever their type, receive an equal share of the liquidation value of portfolio. Let us define these bankruptcy consumptions as c_B. The liquidation value of the portfolio is equal to all liquid reserves put in storage plus early assets sold at a price P^*, where P^* is the price of sold early assets when everything is sold ($X = S$). The consumption in case of default is then:

$$c_B = L + SP^*$$

The existence of this asymmetry of information makes banks vulnerable to bank runs and forces bank to design an *incentive compatible contract*. A bank run is defined as follows.

Definition 1 *A bank run happens when patient depositors, i.e. consumers that have not been hit by the liquidity shock, withdraw from the bank in period 1, instead of waiting period 2.*

The analysis is restricted to fundamental bank runs, in line with Allen and Gale (1998, 2004) literature. I will not focus on sunspot bank runs, where banks runs are described as panics phenomenon in which, when deciding whether to run or not, consumers take their decision *conditional on other patient consumers running*. I apply the following rule of equilibrium selection: whenever a no run equilibrium exists, I select the no run equilibrium, meaning that I rule out sunspot bank run. The objective of this paper is not to study such multiple equilibria situations. Alternatively, assuming that banks benefit from deposit insurance also rules out sunspot bank runs.

Patient consumers run when they know they will receive more if they misrepresent their type, withdraw c_1 in period 1, store until period 2 rather than waiting period 2 to withdraw. Crucially, the no run condition for a given patient consumer must not include its share of funds' profits. Indeed, patient consumers receive their share of profits whether they have run or not.

Applying the revelation principle, the solvency condition (or no run condition) states that for the bank to be solvent, it must be able to pay patient consumers *at least* the same amount in present value as the fixed rate $\overline{c}$ served to impatient consumers. If not, a bank run could happen, the bank defaults instead. It is the possibility of bank run that triggers default. The solvency condition (or no run condition) ensures that the banking contract is incentive compatible. It writes:

$$\text{Solvency or no run condition} \qquad \theta\overline{c}D + \theta\overline{c}D\frac{P(\theta)}{R^E} \leq L + X(\theta)P(\theta) \tag{2}$$

where $(\overline{c}D)$ is the minimum level of consumption the bank must be able to pay patient consumers for them not to run, $\frac{P(\theta)}{R}$ is the present value of one unit of good at date 2, θ is the realized value of the size of the liquidity shock (i.e. the number of impatient consumers) and where the price and the number of early asset sold by the bank depend on the realization of this shock.

Bankruptcy happens above a certain threshold of the liquidity shock called θ^*. This threshold is defined as the size of the liquidity shock for which:

$$\theta^*\overline{c}D + \theta^*\overline{c}D\frac{P^*}{R^E} = L + SP^* \tag{3}$$

Let us define P^* as the price of early assets when all early assets are sold ($X = S$).

$$\forall\theta \geq \theta^* \text{ , } P = P^*$$

4 Equilibrium

In this section, I study two types of economies. In the efficient constrained economy, the price of sold early project $P(\theta)$ is not taken as given whereas in the decentralized economy, this price is taken as given by agents. I then show how a pecuniary externality arises in the decentralized economy because the effect on price is not internalized by agents and because the model presents two imperfections combined: incomplete market and asymmetry of information on depositors' liquidity type.

In both cases, decentralized or efficient constrained equilibrium, the model is solved backward. I solve for each type of equilibrium separately, following the same steps in each case. I first focus on period 1 decisions made after the realization of the liquidity shock: I solve for funds' problem and then solve for the period 1 consumptions served by banks to patient and impatient consumers. The period 1 decisions do not differ in the two economies. Second, I study the period 0 decision, i.e. I solve for households' and banks' problems. They differ in each equilibrium, giving rise to an externality that reduces welfare.

4.1 Decentralized economy

Definition 2 *A decentralized perfectly competitive equilibrium is defined as the equilibrium of an economy in which:*
i) depositors' type (impatient / patient) is private information;
ii) banks design an incentive compatible contract with depositors whose terms are defined above;
iii) banks maximize their depositors' utility as a result of perfect competition in the banking sector;
iv) banks are price takers on the secondary market of sold assets as a result of atomicity;
v) households optimally allocate their endowment between bank deposits and funds' profits shares and take the asset price as given; they only consider individual variables;
vi) funds maximize their profits.

The households' choice variables (W, D) and the banks' choice variables (S, L) do not depend on the realization of the liquidity shock θ as the decisions are made before the shock hits. The funds' choice variables $(Y,\, X)$ depend on θ because the choice is made after the realization of the shock.

To solve for period 1 decisions, I first solve for the funds' problem and then solve for the consumptions served by the bank to both types of depositors.

Funds' program

The funds' program is identical in both economies studied: decentralized and efficient constrained. Solving for this problem yields the price of early assets on the secondary market. Funds receive the wealth W from households in period 0. They remain inactive in period 0. In period 1, they divide their wealth between buying back $X(\theta)$ early assets sold at a price $P(\theta)$ by the banks or investing $Y(\theta)$ in late assets. Funds give back their profits to patient consumers in period 2 who are the only consumers left to still care about consumption at that time.

Funds' decisions are made after the realization of the liquidity shock, so for a given θ. The funds' program is:

$$\max_{X(\theta),Y(\theta)} R^L Y(\theta) + R^E X(\theta)$$

subject to:

$$W = Y(\theta) + P(\theta)X(\theta)$$

The first order condition gives the price of the sold early projects as the ratio between the marginal return of investing in new late assets and the marginal return of buying back early assets. It makes funds indifferent between holding sold early assets or new late assets.

$$P(\theta) = \frac{R^E}{R^L} \equiv P^F$$

This price is the fundamental price, defined only by the relative productivity of early and late assets. As funds cannot take on debt or short sell assets:

$$Y(\theta) \geq 0$$

Therefore, for the secondary market of early assets to clear, it is necessary that:

$$PX(\theta) \leq W$$

where $PX(\theta) = \theta\bar{c}D - L$. This quantity is exactly equal to what the bank is missing to be able to pay impatient consumer the promised rate. As θ increases, X increases since the fundamental price is fixed, and $Y(\theta)$ decreases to zero, up to the point where the market cannot clear anymore.

The wealth of funds available to buy back early assets is fixed in period 1 as it was determined by households' decision in period 0. Then, in some states of the world, for high realization of the liquidity shock, the wealth W can be too low to buy all early assets sold by banks at the fundamental price $P^F = \frac{R^E}{R^L}$. The fundamental price would in these cases imply that the market cannot clear. Then, the price has to fall below its fundamental value for the market to clear. A *cash-in-the-market* pricing situation arises, as defined by Allen and Gale (2005). The price is no longer determined by productivity but by the amount of cash available to buy assets and the amount of assets sold.

Definition 3 *Fire sales are situations in which the early assets are sold by banks to funds at a price strictly smaller than their fundamental value P^F.*

I refer to Schleifer and Vishny (1992) definition of fire sales: fire sales are situations in which an agent is *forced* to sell an asset at a *dislocated price*. Here, the bank is forced to sell all its assets to pay back impatient consumers, and the price brutally falls below the fundamental value.

Theorem 4.1 *When $W < X\frac{R^E}{R^L}$, then $X = S$ and $P^* = \frac{W}{S}$*

This theorem states that the cash-in-the-market pricing arises in case of high liquidity shock triggering default and for low levels of funds' wealth. In this case, the amount of early assets sold jumps to S, and the price falls to its cash-in-the-market price, implying a discontinuity in the price at P^*. It is straightforward from this theorem that keeping more reserves reduces the probability of defaulting at a cash-in-the-market pricing. In other words, keeping more reserves reduces the probability of experiencing fire sales.

Liquidity shock thresholds

In period 1, three cases can arise, depending on the realization of the liquidity shock: *i*) the liquidity shock is so low that the bank does not need to sell any early assets to pay impatient consumers; *ii*) the liquidity shock is such that the bank needs to sell early assets, and the funds wealth is sufficient for the price to be at its fundamental value P^F; *iii*) the liquidity shock is so high that the funds wealth is not sufficient for the price to remain at the fundamental value: it falls at its cash-in-the-market value, P^*, and the bank needs to default.

c_1 is the consumption of type 1 consumers (impatient). c_2 is the consumption of type 2 consumers (patient) given by banks. C_2 is patient consumers' total consumption, including both the payment by banks and their share of funds' profits, $\pi(\theta)$:

$$C_2 = c_2 + \frac{\pi(\theta)}{1-\theta}$$

In each three cases, the consumptions c_1 and C_2 (including funds profits share) vary. Let us define these consumptions which depend on the realized level of liquidity shock.

$\overline{\theta}$ is the threshold below which the bank does not need to sell. It is defined as:

$$\overline{\theta}\overline{c}D = L$$

When $\theta < \overline{\theta}$, no early assets are sold. Impatient consumers get the promised rate on their deposits. Patient consumers get the remaining reserves after payment of patient consumers $(L - \theta\overline{c}D)$, plus returns on early asset (R^E) divided among them, plus their share of funds profits. In this case, as no early assets are sold, funds invest their whole wealth W in new late assets. The consumptions are then:

$$\text{When } \theta < \overline{\theta}\ , \begin{cases} c_1 = \overline{c}D \equiv \overline{c_1} \\ C_2 = \frac{L-\theta\overline{c}D+SR^E+WR^L}{1-\theta} \end{cases}$$

θ^* is the threshold above which the bank has to default. This threshold is interpreted as the probability of default: the larger it is, the smaller is the probability of default. From equation (3), it is defined as:

$$\theta^* = \frac{R^E L + R^E S P^* - \overline{c}P^* D}{\overline{c}D(R^E - P^*)}$$

This threshold depends on P^*, the price in case of default. I will show that in the decentralized economy banks do not correctly anticipate their probability of default because they take the early asset price on the secondary market as given.

For $\overline{\theta} \leq \theta < \theta^*$, the price is at its fundamental value and the bank is solvent. Early assets are sold at the fundamental price because the wealth of funds is sufficient. Impatient consumers get the promised rate on their deposits. Patient consumers get their share of funds' profits and the returns on the early assets not sold to funds $(R^E(S-X))$ divided among them. Indeed, a fraction of early assets' return XR^E is allocated to impatient consumers. The consumptions are then:

$$\text{When } \overline{\theta} \leq \theta < \theta^* \ , \ \begin{cases} c_1 = \overline{c}D \equiv \overline{c_1} \\ C_2 = \frac{R^E S + R^L(W+L) - \theta \overline{c} R^L D}{1-\theta} = \frac{R^E(S-X)+R^L W}{1-\theta} \end{cases}$$

When $\theta \geq \theta^*$, the early asset price falls at the cash-in-the-market value and the bank defaults. The last case is the case for which the bank defaults because it cannot both pay the promised rate and still satisfy the no run condition. The price falls at the cash-in-the-market price $P^* = W/S$. Both patient and impatient consumers receive their share of the liquidation value. Patient consumers get their share of funds' profits in addition.

$$\text{When } \theta^* \leq \theta \ , \ \begin{cases} c_1 = L + SP^* \equiv c^B = L + W \\ C_2^B = c^B + \frac{R^E S}{1-\theta} \end{cases}$$

Theorem 4.2 $\lim_{\theta \to \theta^* -} C_2 \ \leq \ \lim_{\theta \to \theta^* +} C_2$. *Therefore, patient consumers benefit from a bankruptcy.*

Proof 1 *As θ tends to θ^* to the left, $R^E X$ tends to $R^L W$ as this is the condition for having a bankruptcy and P^* falling at the cash-in-the-market value. Therefore, the consumption of patient consumers before bankruptcy $C_2^B = \frac{R^E(S-X)+R^L W}{1-\theta}$ becomes increasingly smaller than the consumption of patient after bankruptcy $C_2^B = L+W+\frac{R^E S}{1-\theta}$ as θ approaches θ^*.* $\square$

This theorem states that, as the number of early assets sold by the bank X increases, the consumption of patient consumers before the bankruptcy decreases because the price is fixed at the fundamental value. Bankruptcy allows to increase the consumption of patient consumers: they get the whole return on early assets SR^E. The funds get richer thanks to the bankruptcy that increases the funds profits from WR^L to SR^E. Indeed, $WR^L < R^E S$ is the condition for having a bankruptcy at a cash-in-the-market price. In the meantime, banks get poorer due to the bankruptcy and the collapse of price. This wealth redistribution from banks to funds that goes through the price is in fact a redistribution from impatient consumers to patient consumers, to the detriment of impatient consumers. And this redistribution occurs precisely when the impatient consumers are the more numerous, for high realization of the liquidity shock. Therefore, it implies an *ex post* welfare loss.

Banks' problem

I first solve for the banks' problem, taking deposits D and funds' profits shares W as given and then for the households' problem.

The banking sector is perfectly competitive so that banks maximize their depositors' utility. Banks design the optimal contract after households have made their deposits: given the deposits D received, the bank chooses L, S, and $\bar{c}$ to maximize its depositors' utility. The problem then writes:

$$\max_{\bar{c},L,S} \left[E_\theta u[\theta c_1(\theta) + (1-\theta)C_2(\theta)]\right]$$

subject to the budget constraint:

$$D = L + S$$

The Lagrange multiplier is μ_1. The contract needs to be incentive compatible: when the bank can no longer satisfy the no run condition, it defaults. The solvency or no run condition is included in the problem through the solvency threshold θ^*. Banks take the price of early assets as given and so when choosing the amount of early assets S and of reserves L, they ignore their impact on the price of sold assets $P(\theta)$[3]. They incorrectly anticipate their probability of default.

[3] In a setting with micro imperfection and missing markets, this price-taking feature will be the source of the pecuniary externality as will be made clear later with thorough details.

The three first order conditions (with respect to S, L and $\overline{c}$) combined writes:

$$\begin{aligned}
&\frac{\overline{c}}{L+SP^*}(1-P^*)\int_0^{\theta^*}\theta D u'(\overline{c_1})d\theta + \int_{\theta^*}^{1}\theta[1-P^*]u'(c_1^B)d\theta = \\
&\int_0^{\overline{\theta}}[R^E - 1 + \frac{\theta D\overline{c}}{L+SP^*}(1-P^*)]u'(C_2)d\theta \\
&+\int_{\overline{\theta}}^{\theta^*}[(R^E - R^L) + \frac{R^L\theta D\overline{c}}{L+SP^*}(1-P^*)]u'(C_2)d\theta \\
&+\int_{\theta^*}^{1}[R^E - (1-P^*)(1-\theta)]u'(C_2^B)d\theta
\end{aligned} \tag{4}$$

The bank designs the contract to ensure that the expected marginal utility of an impatient consumers equals marginal utility of a patient agent. The returns of early and late assets and the risk aversion of households are the deep parameters that determine the bank's choice.

Households' problem

A given household i chooses deposits D^i and funds' share W^i knowing that his choice does not have any impact on aggregate variables D, L, $Y(\theta)$, $X(\theta)$. All aggregate variables are denoted thereafter without any subscript. Individual variables are denoted with the subscript i.

At this time, households do not know whether they will be an impatient or a patient consumer when the liquidity shock hits in period 1. Crucially, they know that banks maximize their utility with respect to $\overline{c}, L, S$: they maximize with respect to D^i and W^i their utility that is maximized by the bank with respect to $\overline{c}$, L and S.

$$\max_{D^i,W^i}\left[\max_{\overline{c},L,S} E_\theta u[\theta c_1^i(\theta) + (1-\theta)C_2^i(\theta)]\right]$$

subject to its budget constraint $E^i = D^i + W^i$.

Using the envelop theorem, the first order condition then writes, where subscripts i have been omitted as households are identical and of mass 1:

$$\int_0^{\theta^*} \theta \overline{c} u'(\overline{c_1}) d\theta + \int_\theta^1 \theta \frac{L + SP^*}{D} u'(c_1^B) + \mu_1 = \int_0^{\overline{\theta}} [R^L - \frac{L - \theta \overline{c} D + SR^E}{D}] u'(C_2) d\theta$$
$$+ \int_{\overline{\theta}}^{\theta^*} [R^L - \frac{(S - X)R^E}{D}] u'(C_2) + \int_{\theta^*}^1 [\frac{R^E S}{W} - \frac{(1-\theta)(L + SP^*)}{D}] u'(C_2^B) d\theta$$

where μ_1 is the Lagrange multiplier associated with the bank's budget constraint $D = L + S$.

It means that households choose D and W such that the expected marginal utility gain if they are impatient is exactly compensated by the utility gain if they are patient. Indeed, they are optimizing across two states of the world: the case where they are hit by the liquidity shock and the case where they are not.

4.2 Constrained efficient economy

In this section, I now turn to the constrained efficient problem, where the main difference is that the price is no longer taken as given.

Definition 4 *An efficient constrained equilibrium is defined as the equilibrium of an economy in which:*
i) depositors' type (impatient / patient) is private information;
ii) banks design a contract with depositors whose terms are defined above;
iii) banks maximize their depositors' utility;
iv) banks take into account the effect of their period 1 choice of S and L on period 1 price $P(\theta)$ of sold assets;
v) households optimally allocate their endowment between deposits and profits shares of funds taking into account the effect of their action on price and on aggregate variables;
vi) funds maximize their profits.

To solve for the constrained efficient equilibrium, let us first recognize that all period 1 decisions are identical to the decentralized economy. The funds' problem and the different levels of consumptions, given θ and period 0 decisions, are identical in the decentralized and in the constrained efficient economy so that I do not reproduce the whole section here. On the contrary, period 0 decisions, i.e. households' and banks' problems, differ from the decentralized equilibrium. Let us solve for these decisions now.

Banks' problem

The banks maximize its depositors' utility with respect to S and L, taking S as given. The bank's problem is modified with respect to the decentralized economy because the bank now internalizes the effect of its choice of S, $\overline{c}$, and L on the price of early assets on the secondary market. We will get into further details regarding the differences arising in the next section. Let us first write the three first order conditions (with respect to S, L and $\overline{c}$) combined in one equation. The differences with the decentralized economy are highlighted in red.

$$\frac{\overline{c}}{L+SP^*}(1-P^*\frac{\boldsymbol{\overline{c}D-W-L}}{\boldsymbol{R^ES-W}})\int_0^{\theta^*}\theta Du'(\overline{c_1})d\theta+\int_{\theta^*}^1\boldsymbol{\theta}u'(c_1^B)d\theta=$$
$$\int_0^{\overline{\theta}}[R^E-1+\frac{\theta D\overline{c}}{L+SP^*}(1-P^*\frac{\boldsymbol{\overline{c}D-W-L}}{\boldsymbol{R^ES-W}})]u'(C_2)d\theta$$
$$+\int_{\overline{\theta}}^{\theta^*}\frac{R^L\theta D\overline{c}}{L+SP^*}(1-P^*\frac{\boldsymbol{\overline{c}D-W-L}}{\boldsymbol{R^ES-W}})]u'(C_2)d\theta+\int_{\theta^*}^1[R^E-\boldsymbol{(1-\theta)}]u'(C_2^B)d\theta \tag{5}$$

They are of two types: in the decentralized economy, the bank incorrectly anticipates first, its probability of default and, second, the levels of consumption it is able to serve depositors in case of default.

Households' problem

The problem is identical to the decentralized one except for two major differences. First, households now choose directly the aggregate variables D and W rather than the individual ones D^i and W^i, as a social planner would do. Second, households now recognize the impact of their choice on the price of early asset and no longer take it as given. They recognize that their decisions have an impact on the insolvency threshold θ^*.

Let us first define $E[U(ND)]$ the expected utility when θ tends to θ^* to the left so that the bank remains solvent:

$$E[U(ND)]=\theta^*u(\overline{c_1})+(1-\theta^*)u[\frac{R^ES+R^L(W+L)-\theta\overline{c}R^LD}{1-\theta}]$$

where $U(ND)$ stands for consumption when there is no default.

$E[U(D)]$ the expected utility when θ tends to θ^* to the right so that the bank is defaulting:

$$E[U(D)] = \theta^* u(L+W) + (1-\theta^*) u \left[L + W + \frac{R^E S}{1-\theta^*} \right]$$

where $U(D)$ stands for consumption in case of default.

The first order condition of households in the efficient constrained economy is then:

$$\begin{aligned}
&\int_{\overline{\theta}}^{\theta^*} \theta \overline{c} u'(\overline{c_1}) d\theta - \int_{\theta^*}^{1} \theta u'(c_1^B) d\theta \\
&+ \Big[E[U(ND)] - E[U(D)] \Big] \frac{R^E (P^* + 1)(\overline{c} D - L) - R^E (W P^* + R^E S)}{\overline{c} D (R^E S - W)(R^E - P^*)} \\
&= \int_0^{\overline{\theta}} [R^L + \theta \overline{c} - R^E] u'(C_2) d\theta + \int_{\overline{\theta}}^{\theta^*} [R^L (1 + \theta \overline{c})] u'(C_2) d\theta \\
&- \int_{\theta^*}^{1} [R^E - (1-\theta)] u'(C_2^B) d\theta
\end{aligned}$$

A comparison between the first order condition in the two economies makes it clear that an inefficiency arises in the households' choice. They do not choose the optimal levels of W and D.

5 The pecuniary externality and liquidity ratios

In this section, I focus on the externality arising in the bank's choice, rather than on the externality laying in the households' choice. Let us compare equation (4) and equation (5). These two equations combine the three first order conditions with respect to L, S and $\overline{c}$, respectively in the decentralized economy and in the constrained efficient economy. This comparison allows to identify clearly the pecuniary externality.

The first order conditions with respect to $\overline{c}$ and L are identical in both economies. Only the first order condition with respect to S differs in the decentralized and in the efficient constrained economy. The allocation would be optimal if the bank was not choosing an inefficient level of S. Therefore, all differences in the bank choice arise from the choice of S by banks. The externality only lays in the choice of S because the bank incorrectly anticipates the asset price at which it will be able to sell its assets on the secondary market. It does not realize that in case of default the price falls to the cash-in-the-market price W/S when the funds wealth is too low. Therefore, choosing more S in period 0 implies a lower bankruptcy price in period 1. The bank ignores this fact when making its choice of S.

I now make use of two notions defined above. $E[U(ND)]$ is the expected utility of a given depositor when the realized liquidity shock is θ^* and the bank remains solvent: it is the limit value to the left of the expected consumption of a given depositor when θ tends to θ^*. $E[U(D)]$ is the expected utility of a given depositor when the realized liquidity shock is θ^* and the bank is defaulting: it is the limit value to the left of the expected consumption of a given depositor when θ tends to θ^*. Given the assumption on the utility, i.e. because of the concavity of utility, let us state a preliminary result:

Lemma 5.1 *When θ tends to θ^*, $E[U(D)] < E[U(ND)]$.*

Proof 2 *The result follows directly from the strict concavity of the utility function.*
□

For a known size of the liquidity shock close to θ^*, the expected utility of a given depositor who *ex ante* does not know if he will be impatient or patient is higher before bankruptcy than after bankruptcy. This result allows to state a second theorem.

Theorem 5.2 *The value of the partial derivative of the utility with respect to S is greater in the decentralized economy (denoted by a subscript "dec") than in the constrained efficient economy (denoted by a subscript "soc") for a given value of S, $S = \overline{S}$.*

$$\frac{\partial U_{dec}}{\partial S}(S = \overline{S}) > \frac{\partial U_{soc}}{\partial S}(S = \overline{S})$$

Proof 3 *In the decentralized economy, it writes:*

$$\frac{\partial U_{dec}}{\partial S} = \int_0^{\theta^*} R^E u'(C_2) d\theta + \int_{\theta^*}^1 \theta P^* u'(c_1^B) d\theta + \int_{\theta^*}^1 [(1-\theta)P^* + R^E] u'(C_2^B) d\theta$$
$$+ [E[U(ND)] - E[U(D)] \frac{R^E P^*}{\overline{c} D(R^E - P^*)}$$

And in the efficient constrained economy, it writes:

$$\frac{\partial U_{soc}}{\partial S} = \int_0^{\theta^*} R^E u'(C_2) d\theta + \int_{\theta^*}^1 [(1-\theta)P^* + R^E] u'(C_2^B) d\theta$$
$$+ [E[U(ND)] - E[U(D)] \frac{R^E P^*}{\overline{c} D(R^E - P^*)} \frac{\overline{c} - W - L}{R^E S - W}$$

Using the no run condition with $P = P^F$, and the facts that $\overline{c}D \leq \theta\overline{c}D + \theta\overline{c}D\frac{P(\theta)}{R^E}$ and $L + SR^E/R^L \leq L + R^E S$ (as $R^L \geq 1$ and $R^E \geq R^L$), we get that $\frac{\overline{c}-W-L}{R^E S - W} \leq 1$. The partial derivative with respect to L and $\overline{c}$ are the same in the two economies, the allocations $(L, \overline{c}, \overline{S})$ are the same, i.e. for a given value of S[4]. *Hence the result for a given value of S, $\overline{S}$.* $\square$

This theorem shows that the decentralized bank believes that increasing S marginally increases the utility more than it does in reality, once the pecuniary externality is taken into account. It explains why the bank chooses to invest more in early asset S than the efficient constrained allocation.

The externality arises in the first place because of the incomplete market feature of the model as developed by Allen and Gale (2004). In my model, there is a missing market because banks cannot insure against the aggregate liquidity shock. But the pecuniary externality arises because on top of incomplete markets, there is an asymmetry of information between banks and depositors regarding liquidity types. This informational imperfection forces the bank to design an incentive compatible contract. The externality created by this asymmetry of information has two dimensions. First, the bank incorrectly estimates its probability of default and second it

[4]Obviously, the whole allocations $(L, \overline{c}, S)$ are different, hence the externality.

incorrectly anticipates the payments it will be able to make to depositors in case of default.

The welfare loss arises from an incorrect evaluation of the probability of default as expressed by the missing term $\frac{\bar{c}D-L-W}{R^E S-W}$ in the decentralized first order condition. Indeed, the bank takes into account the impact of its choice on the thresholds θ^* so on its probability of default. An inefficiency arises in the decentralized economy when the bank chooses S as it does not fully understand how its choice of S will impact the early asset price and thus its probability of default. It fails to recognize that this price might fall below the fundamental value in case of default for low funds wealth. Let us compare the partial derivative of θ^* with respect to S in the decentralized and in the constrained efficient equilibrium.

$$\frac{\partial\theta^*}{\partial S_{dec}} = \frac{R^E P^*_{dec}}{\bar{c}_{dec} D_{dec}(R^E - P^*_{dec})}$$

$$\frac{\partial\theta^*}{\partial S_{soc}} = \frac{R^E P^*_{soc}}{\bar{c}_{soc} D_{soc}(R^E - P^*_{soc})} \frac{\overline{c_{soc}} D_{soc} - L_{soc} - W_{soc}}{R^E S_{soc} - W_{soc}}$$

where the subscript "soc" denotes an efficient constrained equilibrium variable and the subscript "dec" an decentralized equilibrium variable. They only differ by one term, $\frac{\bar{c}D-L-W}{R^E S-W}$.

In the decentralized equilibrium, this partial derivative is always positive so that the bank believes that the probability of default is monotone in S. Indeed, $R^E - P^* > 0$ as $P \leq R^E/R^L$ with $R^L \geq 1$. The bank incorrectly thinks that increasing the number of early projects always decreases its probability of default. It does not take into account the effect of its choice on the bankruptcy price $P^* = W/S$ of sold assets and believes that the bankruptcy price does not depend on its choice of S. Indeed, the more S the bank chooses, the more the assets price falls in case of default when the price is at the cash-in-the-market value. The bank does not understand that above a certain threshold of S, increasing S can increase its probability of default.

The partial derivative of the probability of default with respect to S in the efficient equilibrium is on the contrary not always of the same sign. The efficient bank understands that the probability of default is not monotone in S. At a certain point increasing S increases its probability of default as the bankruptcy price $P^* = W/S$ is decreasing in S. When at the equilibrium allocation this derivative is negative in the constrained efficient economy, the efficient bank keeps a sufficient liquidity buffer:

$$\frac{\partial\theta^*}{\partial S_{soc}} \leq 0 \Leftrightarrow L_{soc} \geq \overline{c}_{soc}D_{soc} - W_{soc}$$

In this case, the efficient bank keeps a sufficient amount of liquid reserves L, to be able to cover its obligations that depends on the level of D (and so on W) and on the fixed promised rate $\overline{c}$.

Whenever the bank is defaulting, assets are sold at a price below the fundamental value. Then, the bank gets poorer. When choosing S *ex ante*, the bank does not anticipate correctly the amount $L+SP^*$ it will be able to pay depositors. Therefore, the degree of insurance provided is not efficient. A transfer of wealth from impatient consumers to patient consumers happens through the collapse of price P^* that makes funds richer to the detriment of the bank. And the bank does not anticipate this transfer of wealth, whereas this redistribution matters for social welfare.

There is also an *ex post* cost as this redistribution of wealth from banks to funds and so from impatient to patient depositors happen precisely when impatient consumers are the most numerous, i.e. in case of default so for high realization of the liquidity shock.

Liquidity ratios suggested in the Basel III framework aim at forcing banks to hold more liquid assets than they spontaneously do as they ignore the potential effects of fire sales. In my model, the proposed liquidity coverage ratio would be equivalent to forcing bank to hold more reserves L and invest less in early assets S. This makes sense in the model in which the bank invests too much in S in the decentralized economy with respect to the constrained efficient economy.

Formally, the bank problem with liquidity ratios is similar to the decentralized economy problem with an additional constraint:

$$\max_{\overline{c},L,S} E_\theta u[\theta c_1(\theta) + (1-\theta)C_2(\theta)]$$

subject to the same former budget constraint whose Lagrange multiplier is μ_1:

$$D = L + S$$

and to the new liquidity ratio constraint whose Lagrange multiplier is μ_2:

$$S \leq \alpha D \text{ with } \alpha \leq 1$$

Then, we now have one instrument to try reach the efficient allocation of the bank. To examine the efficiency of ratio, I only focus on the banks problem. When comparing the decentralized bank allocation with the constrained efficient bank allocation, I assume that the households are making the same choice in both economies so that the bank problem in both economies are directly comparable with a same level of D and W. I do not hold their choice constant but rather assume that in both cases, households are making the same choice. It allows to abstract from the inefficiency arising from the households choice. Indeed, I want to study how constraining the choice of S can help alleviate the inefficiency in the bank choice.

The partial derivatives of the Lagrangian with respect to L and $\overline{c}$ are the same in the decentralized economy and in the constrained efficient economy. As stated above, the only inefficiency lays in the choice of S.

In the decentralized economy with ratio, we have:

$$\frac{\partial \mathcal{L}_{dec}}{\partial S} = \frac{\partial U_{dec}}{\partial S} - \mu_1 - \mu_2$$

In the efficient constraint economy, we have:

$$\frac{\partial \mathcal{L}_{soc}}{\partial S} = \frac{\partial U_{soc}}{\partial S} - \mu_1$$

Theorem 5.3 *Imposing $\mu_2 > 0$ allows to get the decentralized allocation closer to the constrained efficient allocation. Therefore, binding liquidity ratios increase welfare.*

Proof 4 *I assume that households are making the same choice in the decentralized economy and in the efficient constrained economy (but I do not hold this choice constant: households fully take into account the existence of ratios). As demonstrated above, $\frac{\partial U_{dec}}{\partial S}(\overline{S}) > \frac{\partial U_{soc}}{\partial S}(\overline{S})$ for a given S. Besides, the allocation is the same except for the derivative with respect to S. Indeed, we know that $\frac{\partial \mathcal{L}_{soc}}{\partial L} = \frac{\partial \mathcal{L}_{dec}}{\partial L} = \mu_1$ and $\frac{\partial \mathcal{L}_{soc}}{\partial \overline{c}} = \frac{\partial \mathcal{L}_{dec}}{\partial \overline{c}}$.*

Overall, L, $\overline{c}$, W and D are the same in both economies, decentralized and constrained efficient.

Therefore, to have $\frac{\partial \mathcal{L}_{dec}}{\partial S}(L, \overline{c}, W, D)$ closer to $\frac{\partial \mathcal{L}_{soc}}{\partial S}(L, \overline{c}, W, D)$ for a given S, the Lagrange multiplier on the ratio constraint must be strictly positive: $\mu_2 > 0$. $\square$

This theorem states that any binding constraint that lowers S to make it closer to its efficient constrained level increases welfare. In particular, liquidity ratios allow to constrain S in such a way and are therefore welfare improving.

Let us examine a numerical example. The utility is given by a constant relative risk aversion functional form: $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$. Early asset return is $R^E = 1.6$, late asset return is $R^L = 1.2$, households' endowment is $E = 50$, and the risk aversion coefficient is $\gamma = 2$.

First, for this set of parameters, the condition $W - \frac{R^E}{R^L} S < 0$ is satisfied so that the default occurs at a cash-in-the-market price $P^c = W/S$.

The first result is that the portfolio allocation of households is inefficient. In the decentralized economy, banks deposits equal $D = 49.85$, 99% of the endowment, and investment in funds equals $W = 0.15$, 1% of the endowment, whereas in the constrained efficient economy, we have $D = 47.57$, 95% of the endowment, and $W = 2.43$, 5%. With these given parameters, households invest too much in deposits in the decentralized economy.

The second result is that the decentralized economy chooses too much early assets S and too few liquid reserves L with respect to the constrained efficient optimum. The amount invested in S is scaled by the amount of deposits available. In the decentralized economy, $S = 8.48$ or $S/D = 0.1702$, $17,02\%$ of deposits, whereas $S = 7.94$ or $S/D = 0.1670$, 16.7% of deposits, in the efficient constraint equilibrium.

The partial derivative of the utility with respect to S in the decentralized economy is superior to the partial derivative in the efficient constrained economy as shown on figure 1. The bank does not correctly anticipate the marginal effect of increasing S on the expected utility. It believes that it will increase it more than it does in reality.

Let us now examine the impact of a liquidity regulation. As observed in figure 2, liquidity ratios allow to increase utility but do not allow to reach the constrained efficient allocation. The utility is maximized for ratios that force the choice of assets S to be exactly at the efficient constrained level.

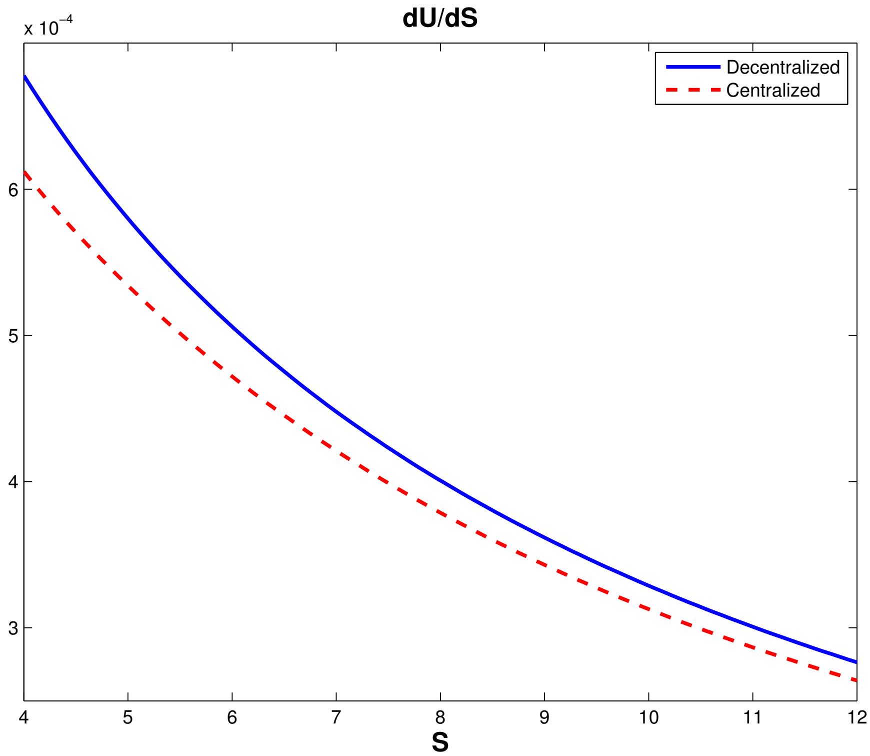

Figure 1: Derivative of the utility with respect to S

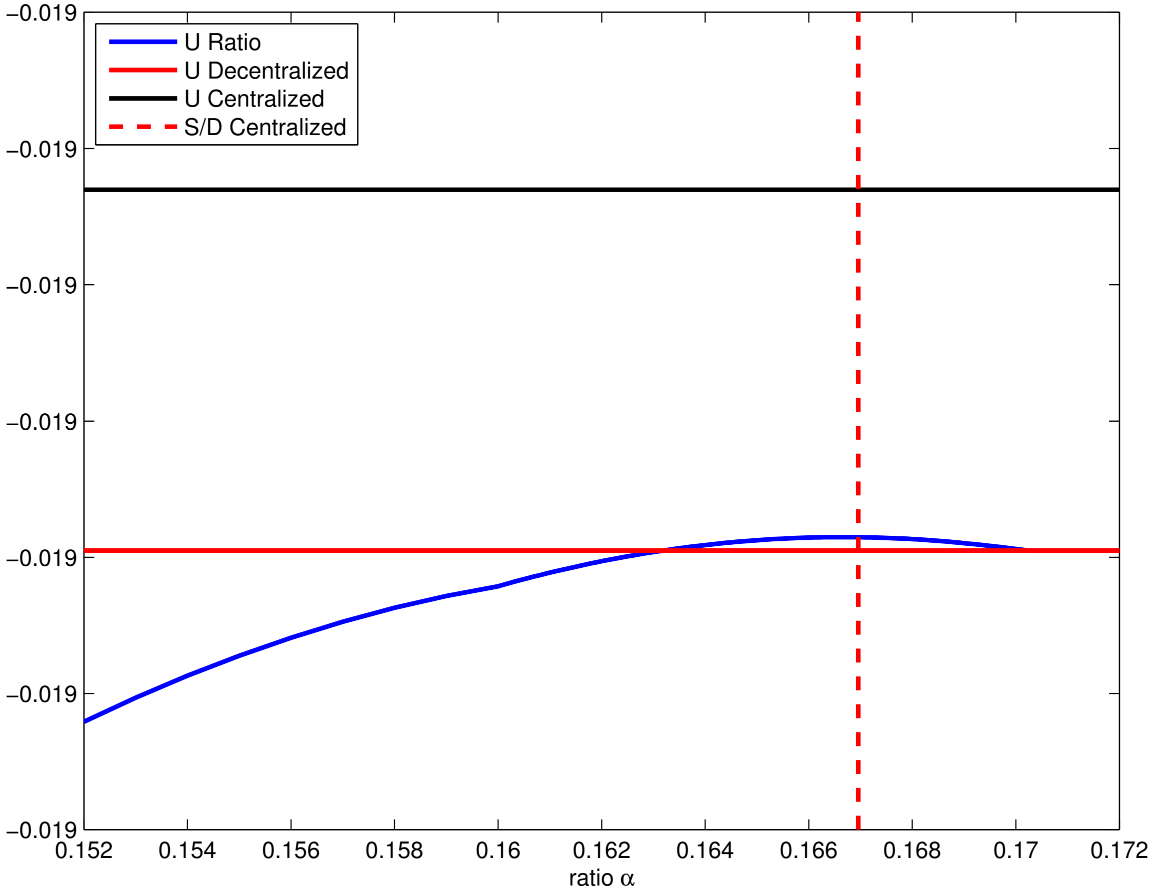

Figure 2: Utility as a function of ratios α

6 Conclusion

My model explains why fire sales generate a pecuniary externality that reduces welfare. This externality arises in the first place because of a combination of incomplete markets and asymmetric information between banks and depositors. Banks incorrectly anticipate the impact of their choice of assets on their probability of default and on the payments they will be able to make to depositors in case of default.

In this setting, I show that efficiency requires to keep a liquidity buffer, i.e. to keep more reserves and invest less in assets than what the decentralized bank spontaneously chooses to do. In the decentralized economy, the bank does not realize that increasing assets too much leads to an increase in the probability of default, as the price falls below its fundamental value to a cash-in-the-market value, and starts decreasing with the amount of assets sold. It thus chooses too much assets and too few reserves.

Imposing liquidity ratios allows to restore some efficiency in the choice of banks by forcing them to invest less in assets and to hold more reserves. Nevertheless, this regulation does not help alleviating the inefficiency on the households side. Under many calibrations, households tend to invest too much in deposits so that the wealth of funds available to buy back assets is too low.

References

[1] Viral Acharya, Hyun Song Shin, and Tanju Yorulmazer. Endogenous choice of bank liquidity: the role of fire sales. Bank of England working papers 376, Bank of England, November 2009.

[2] Viral V Acharya, Hyun Song Shin, and Tanju Yorulmazer. Fire-sale fdi. CEPR Discussion Papers 6319, C.E.P.R. Discussion Papers, May 2007.

[3] Viral V Acharya, Hyun Song Shin, and Tanju Yorulmazer. Fire sales, foreign entry and bank liquidity. CEPR Discussion Papers 6309, C.E.P.R. Discussion Papers, May 2007.

[4] Franklin Allen and Douglas Gale. Optimal financial crises. *Journal of Finance*, 53(4):1245–1284, 08 1998.

[5] Franklin Allen and Douglas Gale. Financial intermediaries and markets. *Econometrica*, 72(4):1023–1061, 07 2004.

[6] Franklin Allen and Douglas Gale. From cash-in-the-market pricing to financial fragility. *Journal of the European Economic Association*, 3(2-3):535–546, 04/05 2005.

[7] Elena Carletti and Agnese Leonello. Credit market competition and liquidity crises. CEPR Discussion Papers 9311, C.E.P.R. Discussion Papers, January 2013.

[8] Eduardo Dávila. Dissecting fire sales externalities. 2011.

[9] Douglas W Diamond and Philip H Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–19, June 1983.

[10] Bruce C Greenwald and Joseph E Stiglitz. Externalities in economies with imperfect information and incomplete markets. *The Quarterly Journal of Economics*, 101(2):229–64, May 1986.

[11] Johan Hombert. Solvency Shocks, Liquidity Shocks, and Fire Sales. Technical report, 2009. Mimeo.

[12] Guido Lorenzoni. Inefficient credit booms. *Review of Economic Studies*, 75(3):809–833, 2008.

[13] Andrei Shleifer and Robert W Vishny. Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance*, 47(4):1343–66, September 1992.

[14] Jeremy C. Stein. Monetary policy as financial-stability regulation. NBER Working Papers 16883, National Bureau of Economic Research, Inc, March 2011.