Fire sales, inefficient banking and liquidity ratios

Axelle Arquié*

September 10, 2015

[Link to the latest version]

Abstract

In a Diamond and Dybvig setting, I introduce a choice by households between the liquid contracts of banks (from which households can withdraw before assets mature) and the illiquid contracts of "funds" (from which households cannot withdraw early), allowing to uncover a new externality of fire sales. Fire sales occur when banks need to sell assets at a depreciated price to face withdrawals by impatient depositors, i.e. households hit by the aggregate liquidity shock. Funds buy assets from banks as households cannot withdraw early from their illiquid contracts. The banking sector is both too risky (banks do not invest enough in reserves) and too large (households invest too much in the liquid contracts of banks) with respect to constrained efficiency. The insurance between patient and impatient households is not efficient because with incomplete markets the price effect of fire sales implies a redistribution between patient and impatient that is not internalized by agents. Liquidity ratios cannot restore constrained efficiency.

JEL-Classification: G010, G210
Keywords: Fire sales, Liquidity ratios

*European University Institute, Max Weber Fellow

1 Introduction

In response to the liquidity crisis experienced by some financial institutions in 2008, the new Basel III regulatory framework introduces a liquidity coverage ratio aimed at reducing the distortions presumably arising from fire sales. 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 innocuous redistributive effects between sellers and buyers and do not imply any welfare loss. However, Greeenwald and Stiglitz (1986) show that in the presence of microeconomic imperfections fire sales can imply a social cost.

A liquidity crisis is an episode during which most agents seek liquidity by selling assets while some others can afford to postpone consumption and then to buy assets. The analysis presented here relies on the idea that fire sales imply some wealth redistribution, at work through the price effect of fire sales, between the two *ex post* types of households: those hit by a liquidity shock and those preserved from it. A novelty is to show that even when households are *ex ante* identical, welfare costs arise when markets are incomplete.

To formalize those ideas, I build a banking model à la Diamond and Dybvig including: *i*) aggregate uncertainty, *ii*) a market for assets and *iii*) a choice by households between a liquid and an illiquid contract.

There are three periods. Assets mature in the last period. In the intermediary period, a liquidity shock hits some households who then only care about present consumption: they are impatient. Other households are patient and can postpone consumption until the last period. The proportion of impatient households is stochastic introducing aggregate uncertainty. There are three agents in the model: households, banks offering liquid contracts and another financial institution called *fund*, offering illiquid contracts. Before the shock hits, *ex ante* identical households allocate their endowment between *i*) liquid contracts offered by banks from which they can withdraw early, i.e. when the liquidity shock hits, and *ii*) illiquid contracts offered by funds from which no early withdraw is possible. Banks decide before the shock how much to invest in assets, how much to keep in liquid reserves and the fixed rate on deposits to serve impatient households. When banks do not hold enough reserves to pay impatient households, they sell assets to funds. Funds have liquidity available to buy assets as households cannot withdraw early from their contracts. If the shock is high enough, a fire sales episode occurs and banks default.

Crucially, both the liquid contract of banks and the illiquid contract of funds are in zero net supply. An important result is that efficiency requires an investment by households in the illiquid contracts of funds. The illiquidity of the contract offered by funds allows those institutions to commit liquidity to the purchase of assets, regardless of the state of the word (i.e. the size of the liquidity shock). Therefore, in the presence of micro imperfections, the illiquidity of the contract, i.e. its non contingency, has a social value[1].

My contribution is threefold. First, this paper provides a general equilibrium analysis of fire sales by introducing a micro-founded choice by households between liquid and illiquid contracts, the later allowing to understand how the supply of liquidity available to buy assets is determined. Second, I motivate a pecuniary externality of fire sales that is, to my knowledge, new in the literature, in a setting with no deadweight loss of assets sales, no collateral constraint and crucially with *ex ante* identical households. The insurance between impatient and patient households is not optimal because fire sales price effect implies a redistribution of wealth across the two *ex post* types of households that is not internalized by agents. Third, I show that imposing liquidity ratios on banks is not sufficient to restore first best as the inefficiency lies both in the banks choice and in the households choice. A tax on deposits can complement liquidity ratios. The paper also shows that *ex post* policies cannot deal with fire sales externalities.

The inefficiency lies *i*) in the banks' choice between assets and reserves - banks over invest in assets - and *ii*) in the households' portfolio allocation - households deposit too much into the liquid contracts of banks and not enough into the illiquid contracts of funds. As a result, in case of fire sales, not only are banks needing too much liquidity with respect to efficiency but also funds do not have the efficient amount of resources to buy back assets. Liquidity ratios can increase welfare by addressing the banks' inefficient choice, but cannot address the inefficient households portfolio allocation.

The pecuniary externality arises from the cash-in-the-market feature of the model, which is due to market incompleteness. When the shock hits, as markets are incomplete, the supply of liquidity available to buy assets is fixed and is determined

[1] More specifically, the micro imperfections are the following: banks cannot insure against the aggregate liquidity shock or, which is equivalent, funds cannot raise additional money when the shock hits.

exactly by the amount households invested in the illiquid contracts of funds in the previous period. During a fire sales episode, the price of assets falls below the fundamental value to a cash-in-the-market value (Allen and Gale 2005). The impossibility to withdraw from the illiquid contract means that only patient households collect its returns: profits are realized in the last period when asset mature and when only patient still care about consumption. Therefore, fire sales imply a redistribution from impatient towards patient because, as the price falls below the fundamental value, profits of funds, i.e. of patient households, increase. A pecuniary externality arises because agents do not take into account the impact of their decisions on this fire-sale price, whereas it implies redistributive effects between patient and impatient that matter for welfare.

The fact that agents do not internalize this price effect of fire sale explains that insurance between patient and impatient is inefficient *whether a fire sales episode occurs or not*. Therefore any *ex post* policy is unable to mitigate the inefficiency. During a fire sales episode, the payment banks make to depositors after defaulting is too low for insuring an efficient risk sharing towards impatient. But even when fire sales do not occur, the insurance is not efficient either, because *i*) the fixed rate paid on deposits by banks to impatient is too high (similar to a price effect) and *ii*) the amount deposited in banks is too high (similar to a quantity effect). Therefore, at the equilibrium, the redistribution towards impatient in the case in which fire sales do not happen can be too high.

The illiquid contract can be interpreted as contractual saving i.e. contracts offered by pension funds or insurance companies: withdrawals is authorized only upon the occurrence of some specific events defined in the contract (retirement, death or disease / disability). Therefore, the simplified financial structure of the model (liquid or illiquid contracts) captures a real-world feature: households choose between bank deposits or contractual savings. As in the model, pension funds and insurance companies cannot be subject to an uncertain aggregate liquidity crisis as members cannot withdraw[2].

Empirical evidence suggest that contractual savings - similar to the illiquid contracts of funds in the model - may have contributed to increasing the depth and the liquidity of financial markets (see Impavido *et al.* 2003 or Impavido and Musalem 2000). One possible explanation for the correlation between the development of contractual

[2] In the real world, members cannot withdraw before the occurrence of the event which is either purely idiosyncratic (death or disease) or expected (retirement). Given that the retirement age is generally not subject to a sudden regulatory change, pension funds anticipate precisely their streams of outflows and cannot be subject to a sudden aggregate liquidity crisis.

savings and the increased depth and liquidity of financial markets is that contractual saving are not subject to unexpected withdrawals[3]. Furthermore, during an aggregate liquidity crisis, pension funds and insurance companies are still receiving the contractual payments from members and thus have some liquidity available. Therefore, the fact that in the model funds holding illiquid contracts are on the buying side of the market seems consistent with reality.

This paper belongs to the literature on banking building on the seminal work of Diamond and Dybvig (1983) and Allen and Gale's series of papers (1998, 2004, 2005). I rely on the concept of cash-in-the-market pricing as defined in Allen and Gale (2005).

Some papers focus on fire sales in models featuring a mixed equilibrium in which identical banks are taking different decisions[4]: on an interbank market, some banks sell assets while some other banks buy them. The difference with my paper is that the buyers of assets here are not banks but rather financial institutions (funds) holding illiquid contracts whose returns can only benefit patient households[5]. Therefore, fire sales imply a redistribution of wealth between patient and impatient households. This redistributive effect of fire sales across *ex post* types of households (patient / impatient) is not internalized by agent, implying an inefficient insurance which is absent from these models. Furthermore, in this model, a mixed equilibrium is not necessary for an equilibrium to exist contrary to Allen and Gale (2004).

My approach is also different from papers in which the social loss of fire sales arises from the fact that natural buyers - i.e. efficient users of the asset who are the only agents that can get its full return - are liquidity constrained, as in Acharya and Yorulmaze (2008)[6]. In the present model, both funds and banks are natural buyers.

The paper proceeds as follows. Section 2 describes the model. Section 3 derives the equilibrium in the decentralized economy and in the centralized economy (defined as a constrained efficient one). Section 4 describes the pecuniary externality lying in the bank's and in the households' choice by comparing the two economies, examines the effect of liquidity ratios on welfare and provides a numerical example. A tax on deposits is also discussed. Section A concludes.

[3]This is particularly true for life insurance, closed-end pension plans and closed-end mutual plans funds from which withdrawals are not allowed before the defined event. It is possible to withdraw from open-end pension plans but these withdrawals imply strong tax penalties.

[4]See Carletti and Leonello (2013) or Allen and Gale (2004).

[5]The fact that only patient households receive the returns of the illiquid contracts is not an additional assumption: it is a mere consequence of the illiquidity of the contract and of the fact that assets mature in the last period.

[6]See Schleifer and Vishny 2011 on the concept of natural buyers.

2 Model

I build a three-period (0, 1 and 2) banking model *à la* Diamond and Dybvig in which an aggregate liquidity shock hits consumers preferences in period 1. *Ex ante*, there are three types of agents: a mass 1 continuum of households, a representative bank offering liquid contracts to households and a financial institution offering illiquid contracts that I call fund.

2.1 Technology

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 households, banks and funds[7]. There are two productive assets : early and late assets. Early assets are productive projects undertaken in period 0 which mature in period 2. For one unit invested in period 0, early assets yield R^E units at maturity i.e. in period 2 whether sold to funds or kept until maturity: there are no change of ownership costs. A second productive technology called late assets can be undertaken in period 1 and also mature in period 2. For one unit invested in period 1, late assets yield R^L in period 2. Both early and late assets yield at least more than storage when held to maturity and early assets are weakly more productive than late assets:

Assumption 1.

$$1 \leq R^L \leq R^E.$$

2.2 Liquidity shock

There is a mass 1 of *ex ante* identical households on the unit interval $[0, 1]$. *Ex post*, households are no longer identical because in period 1, a liquidity shock hits a fraction θ of them. As in Diamond and Dybvig (1983), any consumer i can be hit by the idiosyncratic shock χ_i. The structure of the shock is as follows:

[7]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.

$$\chi_i = \begin{cases} 1 & \text{if households } i \text{ is impatient} \\ 0 & \text{if households } i \text{ is patient} \end{cases}$$

where χ_i follows a binomial $\mathcal{B}(1, \theta)$ with parameter θ.

Consumers who are hit are called impatient and only care about period 1 consumption. Consumption in period 2 provide them no utility. Consumers that are not hit are called patient consumers. They only care about period 2 consumption. The size of the shock, i.e. the fraction θ of consumers being hit, is stochastic - introducing aggregate uncertainty. The shock distribution is drawn from a law that has a continuous probability distribution function, known by all agents. It is assumed to be uniform over the interval $[0, 1]$: $\theta \hookrightarrow \mathcal{U}([0, 1])$.

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

$$U(c_1, c_2; \theta) = \begin{cases} u(c_1) & \text{if } \chi = 1 \\ u(C_2) & \text{if } \chi = 0 \end{cases}$$

where c_1 is consumption by impatient households and C_2 consumption by patient households.

2.3 Households portfolio choice

Households are endowed in period 0 with E units of goods. They allocate their wealth between the liquid contracts of banks, in which they invest D, and the illiquid contracts of funds, in which they invest K. The budget constraint in period 0 writes:

$$D + K = E$$

The contracts offered by banks is liquid as households can withdraw early in period 1, i.e. before asset maturation in period 2. On the other hand, contracts offered by funds are illiquid as households cannot withdraw early in period 1. Therefore, households hit by the liquidity shock, i.e. impatient households, loose their investment in the illiquid contracts of funds. Only patient households receive returns from illiquid contracts, not by assumption but as a consequence of the structure of the model and of the Diamond and Dybvig assumption of extreme preference: returns are realized only when asset mature in period 2, when impatient households no longer care about consumptions.

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

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

Contracts offered to households are incomplete: they can only invest in two types of contracts, a liquid contract - bank deposits - or an illiquid contract offered by funds. Bank deposits can be withdrawn in period 1 after the shock, whereas the illiquid contract does not allow for early withdrawals. This incomplete set of contracts capture in a simple way the empirical allocation by households between bank deposits and contractual saving offered by pension funds and insurance companies from which members are not allowed to withdraw before the occurrence of the event defined in the contract (retirement, death, disease).

As banks offer liquid contracts, they are in a position of demand for liquidity when the liquidity shock is too high. Hence, banks sell assets. Funds instead offer an illiquid contract. They have some liquidity available as they receive the contractual payments. Therefore, they are thus in a position of supply of liquidity: they buy assets. Therefore, the financial structure studied here which distinguish between a liquid and an illiquid contracts allows understanding the supply and the demand of liquidity and the possible inefficiency in their *ex ante* level as chosen by households: this is why the micro funded choice by households is meaningful.

The profits of funds are realized in period 2 when only patient households still care about consumption so that those profits accrue to them only. Therefore, funds operate a redistribution towards patient households whereas banks operate risk sharing towards impatient agents. In this regard, this financial structure of the model is an acknowledgment of the Jacklin critique: households know that when they invest in the illiquid contract, funds can *ex post* partially undo the risk sharing operated by banks by redistributing toward patient households.

Markets are also incomplete. As we shall see the incomplete market features of the model is the cause of the pecuniary externality at work in the model. First, there are no Arrow securities to ensure against the aggregate liquidity shock. Second, funds cannot raise funds, short sell or borrow in period 1: the liquidity available to buy assets is fixed after the shock hits. Third, there are trading restrictions that impede impatient households to sell to patient households i) their bank deposits and ii) their illiquid contracts.

2.4 Banks

Banks receive the households' deposits D in period 0. As the banking sector is perfectly competitive, banks maximize the utility of their depositors. This decision is made *taking as given the deposits D received.* Hence, the bank offers a non-linear contract: the rate offered depends on the amount of deposits received.

They invest S in early assets and keep L in reserves. Reserves L are put in storage. 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 price of early assets depends on the realization of the liquidity shock θ and can fall below its fundamental value. Therefore, early assets are risky and thus only partially liquid in the sense that they can yield a return below the fundamental value when sold in some states of the world. The budget constraint of the bank in period 0 writes:

$$S + L = D$$

I now turn to the description of the financial arrangement between banks and depositors. A crucial feature of this contract is its incompleteness. The paper does not aim at defining the optimal contract but rather takes the banking contract's terms as exogenous. However, note that the contract incompleteness alone is not *per se* a source of inefficiency because default allows restoring some contingency[8]. The banking contract's terms stipulate that the bank has to promise a fixed payment $\overline{c}$ to *any* depositor willing to withdraw in period 1. The amount promised cannot be made contingent on the realization of the liquidity shock.

Each consumer knows if she is impatient or patient in period 1 but this is a private information. Therefore a patient depositor can pretend to be impatient and withdraw early, i.e. in period 1 rather than wait for period 2. The existence of this asymmetry of information between banks and depositors forces banks to design an *incentive compatible contract.* The bank is solvent if it is able to pay patient consumers *at least*

[8]See Allen and Gale (2004): Allowing for default introduces some contingency in the banking contract that restores the inefficiency arising from its non contingent feature.

the same amount in present value as the fixed rate $\overline{c}$ served to impatient consumers. Otherwise, patient consumers would withdraw early, forcing the bank to default. The solvency condition states that to be solvent, the liquidation value of the bank portfolio must be sufficient to pay impatient the fixed rate and at least the same rate to patient depositors expressed in present value. It writes:

$$\theta\overline{c}D + (1-\theta)\overline{c}D\frac{P(\theta)}{R^E} \leq L + SP(\theta) \tag{1}$$

where $(\overline{c}D)$ is the minimum level of consumption the bank must be able to pay patient consumers for them not to withdraw early, $\frac{P(\theta)}{R^E}$ 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 $P(\theta)$ and the number of early assets $X(\theta)$ sold by the bank depend on the realization of the liquidity shock. Then, the bank's default happens above a certain threshold of the liquidity shock called $\overline{\theta}$. The default threshold is defined as the size of the liquidity shock for which:

$$\overline{\theta}\overline{c}D + (1-\overline{\theta})\overline{c}D\frac{P^*}{R^E} = L + SP^*$$

where P^* is the price for which the bank defaults. It is the price of early assets when all early assets are sold ($X = S$).

From equation 1, the default threshold $\overline{\theta}$ writes:

$$\overline{\theta} = \frac{R^E L + R^E SP^* - \overline{c}P^*D}{\overline{c}D(R^E - P^*)} \tag{2}$$

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

P^*, the price for which the bank defaults, does not depend on θ as will be explained later on.

When the bank does not default, it pays the promised rate on deposits:

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

In case of default, the bank sells all assets and divides the profits among all depositors[9]. Let c_B be the payment by banks in case of default. The liquidation value of the portfolio is equal to all liquid reserves put in storage plus early assets sold at price P^*. The payment in case of default is then:

$$c_B = L + SP^*$$

2.5 Funds

The payment K made to funds in period 1 (but decided by households in period 0) can be interpreted as a pension contribution or a payment to an insurance contract. More broadly, they capture any illiquid contracts from which households cannot withdraw which allows the financial institutions holding those contracts to be in a position to buy assets. The total amount corresponds to the liquidity available to funds to buy assets and hence to the supply of liquidity in the economy. I make a simplifying assumption on the funds balance sheet: funds do not have any other type of assets. In reality, pension funds balance sheet is more complex as they hold some long-term assets. But allowing for funds to have some long-term assets in the model would not materially change the analysis conducted here which focuses on the supply of and the demand for liquidity.

In period 1, funds invest $Y(\theta)$ in late assets which 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 benefit from the exact same gross return on these sold early assets as banks: the early assets they buy yield the same return R_E as originators (banks). This is a fundamental assumption that rules out any technological externality.

The funds' budget constraint in period 1 writes:

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

The liquidity available to funds K 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. A market is missing here.

[9] I do not adopt a Diamond and Dybvig (1983) sequential service constraint rule. As a bankruptcy rule, I rather choose an equal division of the bank portfolio among all depositors, as Allen and Gale (1998, 2004).

2.6 Timing

The timing is as follows. Period 0 is divided into two sub-periods. First, households allocate their endowment between depositing D in the liquid bank contract in period 0 and investing K in the illiquid contract in period 1. The payment to the illiquid contract offered by funds is decided in period 0 and made in period 1: it is similar to a pension contribution or a payment to insurance contracts which are decided contractually *ex ante*. After the portfolio allocation of households has been made, banks receive deposits D. Second, banks decide how much to invest in early assets S and how much liquid reserves L to keep. Banks design an 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.

In period 1, funds collect the payments K by households. For some realizations of the liquidity shock, banks need to sell some early assets to be able to pay impatient the fixed rate. In period 1, funds can buy back these early assets at a price $P(\theta)$ or buy $Y(\theta)$ *late* assets. In some state of the world, banks might default when it cannot pay impatient the fixed rate in period 1 and ensure at least the same payment to patient households in period 2.

3 Equilibrium

In this section, I study two types of economies and solve for the equilibrium in each case. First, in the decentralized economy, banks and households do not internalize the effect of their choice on the price. Second, in the constrained efficient economy, the price of sold early project $P(\theta)$ is no longer taken as given while institutional and technological assumptions are maintained.

3.1 Decentralized economy

The model is solved backward. In period 2, no decisions are taken. First, the period 1 is studied: I solve for funds' problem and define fire sales. I then solve for the payments received by households, which depend on the size of the shock. Second, the period 0 decisions are solved for: the households' and then the banks' problems.

3.1.1 Period 1

Funds' program and fire sales

The funds' choice of (Y, X) is made after the realization of the shock. Funds receive the contribution K from households to their illiquid contracts in period 1. They choose between buying back X early assets sold at a price P by the banks or investing Y 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. The funds' program writes:

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

subject to:

$$K = Y(\theta) + PX(\theta)$$

The first order condition of this program gives the price of the early assets sold by banks 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 early assets and new late assets and it happens when $X \geq 0$ and $Y \geq 0$. The fundamental price P^F is the price that prevails when in equilibrium funds invest both in early and late assets. Therefore, it is defined by the relative productivity of early and late assets:

$$P = \frac{R^E}{R^L} \equiv P^F \tag{3}$$

Crucially, I assume away short selling: $Y(\theta) \geq 0$. Funds cannot take on debt either. Therefore, the wealth available to buy back early assets is fixed, due to markets incompleteness. It is equal to K which was chosen before the shock in period 0 by households.

$PX(\theta) = \theta\bar{c}D - L$ is the exact quantity that the bank is missing to be able to pay impatient consumers the promised rate: it is equal to the supply of assets by banks. For the secondary market of early assets to clear, it is necessary that there is sufficient funds available liquidity K to buy back all assets sold by the bank:

$$PX(\theta) \leq K$$

A cash-in-the-market episode occurs when the wealth of funds and the reserves hold by banks are too low such that $K \leq P^F X$. The cash-in-the-market condition allows defining a threshold θ^* which is the liquidity shock threshold above which the price falls to the cash-in-the-market value. As θ increases, the number of early assets sold $X(\theta)$ increases since the fundamental price is equal to $\frac{R^E}{R^L}$, i.e. does not depend on θ, and the investment on late assets by funds $Y(\theta)$ decreases to zero, up to the point where the market cannot clear anymore. Above a certain threshold, the liquidity shock is too high for the market to absorb all early assets sold at their fundamental price. Then, the price has to fall below its fundamental value for the market to clear. The price is no longer determined by relative productivities but rather by the amount of cash available to buy assets, K, and by the total amount of assets sold.

I focus on cases that are relevant to our analysis, i.e. cases in which cash-in-the-market episodes can arise for certain realization of the liquidity shock. Those are cases for which in equilibrium $K/S < P^F$, where K and S are endogenous variables.

Assumption 2. *A sufficient condition to ensure that the cash-in-the market pricing happens when the bank is defaulting, i.e. selling all early assets $X = S$, is to choose R^E sufficiently high compared to R^L, i.e. the return of early assets sufficiently higher than the return of late assets.*

I now turn to the definition of fire sales for which I refer to Schleifer and Vishny (1992) approach: fire sales are situations in which an agent is *forced* to sell an asset at a *dislocated price*. This definition contains two fundamental ingredients. First, the bank is forced to sell all its assets to pay back impatient consumers and ensure at least the same payment to patient: this is the solvency condition (equation 1). Second, the price brutally falls below the fundamental value.

Definition 1. *Fire sales are situations in which banks are forced to sell assets to funds at a price strictly smaller than their fundamental value P^F because they are insolvent.*

Furthermore, it is easily shown that $\theta^* \geq \overline{\theta}$ is satisfied, in which recall that θ^* is the default threshold. It is the insolvency as governed by the solvency condition of the bank that causes the cash-in-the-market pricing and thus the fire sales. In other words, it cannot happen that a cash-in-the-market situation arises while the bank is still solvent as governed by the solvency equation, making it selling its whole portfolio ($X = S$) whereas the bank is solvent.

Theorem 1 (Fire sales)**.** *Under assumption 2, when $\theta \geq \overline{\theta}$ i.e. when the bank is insolvent, all early assets are sold $X = S$, and the price falls to its fire-sales level so that:*

$$P^* = \frac{K}{S} \tag{4}$$

P^* is called the fire-sales price.

Proof. The proof follows directly from *i*) the fact that $\theta^* \geq \overline{\theta}$ and *ii*) from assumption 2. When $\theta = \overline{\theta} + \epsilon$, the cask-in-the-market condition is not yet satisfied. As the bank is insolvent, it sells all early asset, $X = S$ and due to 2, the price falls below the fundamental price. □

This theorem states that bank insolvency caused by a sufficiently high liquidity shock triggers cash-in-the-market pricing. The amount of early assets sold jumps to S, and the price falls from P^F to its fire sale level K/S, implying a discontinuity in the price at P^*. It is straightforward from this theorem that keeping more reserves reduces the probability of defaulting. Keeping more reserves L and investing less in early assets S reduces both the probability of experiencing fire sales and the severity of these episodes, should they occur.

Proposition 1.

$$P^* \leq 1 \tag{5}$$

Proof. See the proof in appendix, section .1. □

Thus, in case of fire sales, early assets return is lower than storage for the bank: they are risky for the bank.

Payments to households

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 some early assets and the funds liquidity is sufficient for the price to be at its fundamental value P^F; *iii*) the liquidity shock is so high that the funds liquidity is not sufficient for the price to remain at the fundamental value: it falls at its fire-sales value P^*, and the bank needs to default.

In each three cases, the payments to households vary. Remember that c_1 is the payment per capita by banks to impatient households and C_2 the total payment per capita to patient households which includes i) the payment by banks and ii) payments by funds. Two thresholds $\underline{\theta}$ and $\overline{\theta}$ are needed. $\overline{\theta}$ has been defined in section 2 and I now turn to the definition of $\underline{\theta}$.

The threshold $\underline{\theta}$ is the threshold below which the bank does not need to sell any early assets. Impatient consumers get the promised rate on their deposits. Patient consumers get the remaining reserves after payment of impatient 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 liquidity K in new late assets. $\underline{\theta}$ is defined by:

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

The consumptions per capita in the first case $\theta < \underline{\theta}$ are then:

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

The second threshold $\overline{\theta}$, already defined, is the threshold above which the bank has to default. $(1 - \overline{\theta})$ is the probability of default. Therefore, the larger is $\overline{\theta}$, the smaller is the probability of default.

For $\underline{\theta} \leq \theta < \overline{\theta}$, the price is at its fundamental value and the bank is solvent. Early assets are sold at the fundamental price because the liquidity of funds is sufficient to buy back all assets sold at the fundamental price. 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 $\left(R^E\left[S - X(\theta)\right]\right)$ divided among them. Notice that a fraction of early assets' return $X(\theta)R^E$ is allocated to impatient consumers: the bank is operating a redistribution from patient towards impatient depositors. The consumptions per capita in the second case are then:

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

To understand the mechanism of the externality at work in the model, it is crucial to notice that the threshold $\overline{\theta} = \frac{R^E L + R^E SP^* - \overline{c}P^* D}{\overline{c}D(R^E - P^*)}$ depends on the fire-sales price $P^* = K/S$.

When $\theta \geq \overline{\theta}$, the early asset price falls at the fire-sales price $P^* = K/S$ and the bank defaults. In that case, the bank defaults because it cannot both pay the promised rate to impatient depositors in period 1 and make at least the same payment to patient depositors in period 0: the solvency condition 1 is no longer satisfied. Both patient and impatient consumers receive their equal share of the liquidation value. On top of the bank's payment, patient consumers get their share of funds' profits.

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

Proposition 2. $\lim_{\theta \to \overline{\theta}^-} C_2 < \lim_{\theta \to \overline{\theta}^+} C_2$. *Therefore, patient consumers benefit from a default when the size of the shock approaches $\overline{\theta}$ from the right.*

Proof. See proof in appendix .2. □

The first important result of this theorem is that $\lim_{\theta \to \overline{\theta}^-} C_2 \neq \lim_{\theta \to \overline{\theta}^+} C_2$: the payment function is not continuous at the default threshold $\overline{\theta}$ as there is a discontinuity in the price at that point. It directly follows from assuming R^E sufficiently high compared to R^L so that at the equilibrium $K/S \leq P^F$.

The second part of proposition 2 states that, as θ increases, the number of early assets sold X increases and the consumption of patient consumers without default decreases as C_2 is decreasing in X for $\theta \in [\underline{\theta}, \overline{\theta}]$. The bank's default corresponds to a reduction of what is payed to impatient whose payment falls below the promised rate $\overline{c}$. Therefore, default allows to increase what banks pay to patient consumers compared to a situation where the bank would not have defaulted, keeping on paying the promised rate to impatient: patient now get the whole return on early assets SR^E. Default allows to restore some contingency in the incomplete banking contract: it is a redistributive device between patient and impatient and the bank must use it efficiently to optimally insure depositors. In a setting with risk averse depositors, redistribution between the two liquidity types matters.

As we shall see, the inefficiency arises form the failure of banks to internalize the impact of their choice on the probability of default and thus on the redistribution operated between patient and impatient, and not from the default itself as default is optimal *per se.*

3.1.2 Period 0

Period 0 decisions - i.e. the households' choice variables (K, 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. Households choice takes place before bank's choice. Therefore, I first solve for the bank's problem - who take the deposits D received by households as given - and second for the households' problem.

Banks' problem

The banking sector is perfectly competitive so that the bank maximizes its depositors' utility. The bank chooses how much to promise impatient depositors, $\overline{c}$, how much to keep in reserves, L, and how much to invest in productive assets, S, after households have made their deposits: given the deposits D received, the bank chooses L, S, and $\overline{c}$ to maximize its depositors' utility. The problem then writes:

$$\max_{\overline{c},L,S} \left\{ E_\theta \left\{ u \left[\theta c_1(\theta) + (1-\theta) C_2(\theta) \right] \right\} \right\} \tag{6}$$

subject to the budget constraint, whose Lagrange multiplier is μ_1:

$$D = L + S$$

The contract needs to be incentive compatible: when the bank can no longer satisfy the solvency condition stated in equation 1, it defaults. 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 fire-sale price $P^* = K/S$. The price-taking behavior of bank is theoretically grounded on the perfect competition assumed in the banking sector. This assumption seems realistic considering the real world, since banks seem to be indeed acting as price-takers on the international financial markets. In a setting with incomplete markets, this price-taking feature will be the source of the pecuniary externality as will be made clear in section 4 with thorough details.

The Lagrangian of the problem writes:

$$
\begin{aligned}
\mathcal{L}_{bank} &= E_\theta[u(c_1) + (1-\theta)u(C_2)] + \mu_1[D - L - S] \\
&= \int_0^{\underline{\theta}} [u(c_1) + (1-\theta)u(C_2)]\, d\theta \\
&+ \int_{\underline{\theta}}^{\overline{\theta}} [\theta u(c_1) + (1-\theta)u(C_2)]\, d\theta \\
&+ \int_{\overline{\theta}}^{1} \left[\theta u(L + SP^*) + (1-\theta)u\left(L + SP^* + \frac{\pi(\theta)}{1-\theta}\right)\right] d\theta + \mu_1[D - L - S]
\end{aligned}
$$

Recall that the payment function by the bank is not continuous in $\overline{\theta}$ due to default. Then it is useful to define $E[U_{ND}]$ the expected utility when θ tends to $\overline{\theta}$ from the left so that the bank remains solvent - where the subscript $_{ND}$ stands for 'non default':

$$
E[U_{ND}] = \overline{\theta}u(\overline{c_1}) + (1-\overline{\theta})u\left[\frac{R^L K}{1-\overline{\theta}}\right]
$$

$E[U_D]$ is the expected utility when θ tends to $\overline{\theta}$ from the right so that the bank is defaulting - where the subscript $_D$ stands for 'default':

$$
E[U_D] = \overline{\theta}u(L+K) + (1-\overline{\theta})u\left[L + K + \frac{R^E S}{1-\overline{\theta}}\right]
$$

The first order condition with respect to S writes:

$$
\begin{aligned}
&\int_0^{\overline{\theta}} R^E u'(C_2) d\theta + \int_{\overline{\theta}}^{1} \left[\theta P^* u'(C_1^B) + [(1-\theta)P^* + R^E]u'(C_2^B)\right] d\theta \\
&+ \frac{R^E P^*}{\overline{c}D(R^E - P^*)}\left[E(U_{ND}) - E(U_D)\right] = \mu_1
\end{aligned} \tag{7}
$$

The optimality condition 7 ensures that the expected marginal return of an additional unit of early assets expressed in utility terms is exactly equal to its marginal cost - the Lagrange multiplier.

For a liquidity shock between 0 and $\overline{\theta}$, i.e. when the bank is solvent, the asset yields R^E which accrue to impatient agents amounting to $R^E u'(C_2)$ in utility terms. This is the first integral on the left hand side. Notice that the asset yields the exact same return whether it is sold to funds or held until maturity by banks. Therefore, as long as the bank is solvent, sales of assets are immaterial for asset marginal return.

For a liquidity shock between $\overline{\theta}$ and 1, i.e. when the bank is insolvent, the asset yields P^* for the bank and R^E for the funds. This is the second integral on the left hand side. The impatient only perceive the return generated by the bank while the patient perceive both returns, from the banks and from the funds. Hence, the return of the early assets expressed in utility terms is exactly the expectation of $\theta P^* u'(C_1^B)+$ $[(1-\theta)P^* + R^E]u'(C_2^B)$. The structure of the marginal return of early assets clearly shows that default implies a shift on the redistribution operated between patient and impatient households.

Finally the last term on the left hand side reflects the fact that at the solvency threshold $\overline{\theta}$, the payments by the bank function is not continuous (see proposition 2) so that at the threshold, an additional unit of S has an impact on the level of the utility. Recall that $\overline{\theta}$ is also the probability of being solvent. The marginal impact of increasing S by one unit on the probability of being solvent is $\frac{\partial \overline{\theta}}{\partial S} = \frac{R^E P^*}{\overline{c}D(R^E - P^*)}$ and the difference in levels of utility is denoted by $E(U_{ND}) - E(U_D)$. When choosing S, the bank needs to take into account the impact of S on the level of utility at that threshold.

The first order condition with respect to L writes:

$$\begin{aligned}&\int_0^{\underline{\theta}} u'(\overline{C_2})d\theta + \int_{\underline{\theta}}^{\overline{\theta}} R^L u'(C_2)d\theta + \int_{\overline{\theta}}^1 [\theta u'(C_1^B) + (1-\theta)u'(C_2^B)]d\theta \\ &+ \frac{R^E}{\overline{c}D(R^E - P^*)}\left[E(U_{ND}) - E(U_D)\right] = \mu\end{aligned} \tag{8}$$

The interpretation of the optimality condition 8 is similar to the optimality condition 7. Notice that for the return of reserves, the sales of assets now matter. Between 0 and $\underline{\theta}$, the marginal return of reserves is 1 as reserves are put in storage. Between $\underline{\theta}$ and $\overline{\theta}$, i.e. when there are sales of assets and the bank is solvent, the marginal return is equal to the marginal return of late assets R^L: one more unit of reserves held by banks means that one additional unit of liquidity of funds needs not be used to buy back assets and can be invested instead in late assets. R^L can be interpreted here is as the opportunity cost of early assets when those assets are sold.

Combining those two first order conditions shows that at the optimum the bank equates the marginal return expressed in utility terms of investing in the two assets at its disposal: reserves and early assets.

The first order condition with respect to $\overline{c}$ writes:

$$\begin{aligned} &\int_0^{\overline{\theta}} \theta D u'(\overline{c_1}) d\theta = \\ &\frac{R^E(L+SP^*)}{\overline{c}D(R^E-P^*)}\left[E(U_{ND})-E(U_D)\right] + \int_0^{\underline{\theta}} \theta D u'(C_2) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} R^L \theta D u'(C_2) d\theta \end{aligned} \tag{9}$$

The choice of $\overline{c}$ determines the redistribution operated by the bank between patient and impatient depositors. Therefore, the optimality condition 9 requires the bank to operate a redistribution which ensures that the marginal utility of impatient households equals the marginal utility of patient.

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

$$\begin{aligned} &\frac{\overline{c}}{L+SP^*}(1-P^*)\int_0^{\overline{\theta}} \theta D u'(\overline{c_1}) d\theta + \int_{\overline{\theta}}^1 \theta[1-P^*]u'(c_1^B)d\theta = \\ &\int_0^{\underline{\theta}} \left[R^E - 1 + \frac{\theta D\overline{c}}{L+SP^*}(1-P^*)\right] u'(C_2) d\theta \\ &+ \int_{\underline{\theta}}^{\overline{\theta}} \left[(R^E-R^L) + \frac{R^L\theta D\overline{c}}{L+SP^*}(1-P^*)\right] u'(C_2) d\theta \\ &+ \int_{\overline{\theta}}^1 [R^E - (1-P^*)(1-\theta)]u'(C_2^B)d\theta \end{aligned} \tag{10}$$

Equation 10 is an implicit equation in $\overline{c}$, L and S ensuring that the expected marginal utility of an impatient consumer equals the expected marginal utility of a patient consumer, taking into account the marginal returns of assets. Crucially, the fire-sales price P^* enters this decision rule through the default threshold $\overline{\theta}$ and through the level of the payments in case of default. Due to market incompleteness, the overall amount of S chosen will have an impact on the fire-sales price $P^* = K/S$. The failure of the price-taking bank to take into account the impact of its choice of S on P^* explains the inefficient choice as will be made clearer in section 4.

Households' problem

A given household i chooses deposits D^i to invest into liquid contract and the contribution K^i to invest into funds illiquid contract knowing that his choice does not have any impact on aggregate variables D, L, Y, X. All aggregate variables are denoted thereafter without any subscript. Individual variables are denoted with the subscript i.

In period 0, households do not know whether they will become impatient or patient consumer in period 1. They know that banks maximize their utility with respect to $\overline{c}, L, S$ taking deposits as given. The problem writes:

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

subject to the budget constraint $E^i = D^i + K^i$.

If patient, each individual consumer gets in period 2 a share of banks and funds portfolio. The share of bank portfolio is equal to $\frac{D^i}{(1-\theta)D}$ and the share of funds portfolio to $\frac{K^i}{(1-\theta)K}$ as there are $(1-\theta)$ patient.

Using the envelop theorem, the first order condition with respect to D^i yields:

$$\int_0^{\overline{\theta}}\left[\theta\overline{c}u'(\overline{c_1}) + \frac{L-\theta\overline{c}D+SR^E}{D}u'(C_2)\right]d\theta + \int_{\underline{\theta}}^{\overline{\theta}}\left[\theta\overline{c}u'(\overline{c_1}) + \frac{(S-X)R^E}{D}u'(C_2)\right]d\theta$$
$$+\int_{\overline{\theta}}^{1}\left[\theta\frac{L+SP^*}{D}u'(c_1^B) + \frac{(1-\theta)(L+SP^*)}{D}u'(C_2^B)\right]d\theta = \lambda^i \tag{11}$$

The optimality condition 11 ensures that the marginal return of investing into deposits expressed in utility terms equals the marginal cost. When the bank is solvent and there is no sales of assets, one unit of deposits yields $\overline{c}$ for impatient, so that this return is expressed in impatient utility terms, and $\frac{L-\theta\overline{c}D+SR^E}{D}$ for patient by unit of deposits, so that this return is expressed in impatient utility terms. When the bank is solvent and there are some sales of assets, the payment to patient only changes. When the bank defaults, the payments to both types of depositors is the same $\frac{(L+SP^*)}{D}$. Notice that the Lagrange multiplier from the banks' problem μ_1 does not enter the decentralized households problem as they take aggregate variables D as given.

The first order condition with respect to K^i yields:

$$\int_0^{\underline{\theta}} \frac{R^L}{K} u'(C_2) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{R^E X + (K - PX) R^L}{K} \right] u'(C_2) d\theta + \int_{\overline{\theta}}^1 \frac{R^E S}{K} u'(C_2) d\theta = \lambda^i \tag{12}$$

Similarly, the optimality condition 12 ensures that the marginal return of investing into the illiquid contract of funds expressed in utility terms equals the marginal cost. Only patient consumers benefit from these returns so that returns from the illiquid contract offered by funds are expressed only in marginal utility of patient.

The two first order conditions combined with respect to D^i and to K^i write:

$$\begin{aligned}
&\int_0^{\overline{\theta}} \theta \overline{c} u'(\overline{c_1}) d\theta + \int_{\overline{\theta}}^1 \theta \frac{L + SP^*}{D} u'(c_1^B) d\theta + \mu_1 = \\
&\int_0^{\underline{\theta}} \left[\frac{R^E X + (K - PX) R^L}{K} - \frac{L - \theta \overline{c} D + SR^E}{D} u'(C_2) \right] d\theta \\
&+ \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{R^L}{K} - \frac{(S - X) R^E}{D} \right] u'(C_2) d\theta + \int_{\overline{\theta}}^1 \left[\frac{R^E S}{K} - \frac{(1-\theta)(L + SP^*)}{D} u'(C_2^B) \right] d\theta
\end{aligned} \tag{13}$$

The left hand side of equation 13 is the marginal utility of an additional unit of deposit into banks. Thus, at the optimum, households choose D^i and K^i such that the expected marginal utility gain if they are impatient is exactly compensated by the marginal utility gain if they are patient. Indeed, for each realization of θ, 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.

3.1.3 Equilibrium definition

An equilibrium is an allocation made of period 0 choice variables $(D, K, S, L, \overline{c})$, of the solvency threshold $\overline{\theta}$, and for each state θ of period 1 choice variables (X, Y) and of price of early assets P.

Definition 2. *A decentralized perfectly competitive and rational expectation equilibrium is defined as the equilibrium of an economy in which:*
i) depositors' type (impatient / patient) is private information;
ii) the solvency condition of the bank gives the default threshold $\overline{\theta}$*;*
iii) banks and households are price-takers on the secondary market of early assets; in particular, they take the fire-sales price as given;
iv) the fund chooses X *and* Y *after the liquidity shock is realized such that the optimality condition 3 and its budget constraint are satisfied; the price of early assets* P *is given by the optimality condition of funds when the liquidity shock is below the solvency threshold and by equation 4 when the shock is higher;*
v) households choose K *and* D *such that the optimality conditions 11 and 12 and the budget constraint are satisfied;*
vi) with perfect competition in the banking sector, the bank maximizes their depositors' utility (see equation 6) taking the deposits D *by households as given; the bank chooses* S *and* L *such that optimality conditions 7, 8, and 9 and the budget constraint are satisfied.*

3.2 Constrained efficient economy

I now turn to the constrained efficient problem. The difference is that the price is no longer taken as given. The planner maximizes period 0 welfare by choosing the period 0 allocation: how much households invest into the liquid bank contract D and in the illiquid fund contract K, how much banks invest into early assets S and reserves L and the bank deposit contract fixed rate $\overline{c}$.

To solve for this problem, it should be first recognized that all period 1 decisions are identical to the decentralized economy's decisions: the funds' problem and the different payments made by banks and funds to households, 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. Period 0 decisions, i.e. households' and banks' problems, differ from the decentralized equilibrium. I first solve for the bank choice variable as chosen by the constrained social planner and second for the households choice variables as chosen by the constrained social planner.

Banks' problem by the constrained social planner

The bank maximizes its depositors' utility with respect to S and L, taking deposits D as given. The bank's problem is modified with respect to the decentralized economy because the social planner now internalizes the effect of its choice of S, $\overline{c}$, and L on the price of early assets on the secondary market.

The expression of the first order conditions with respects to $\overline{c}$ and L are identical in the decentralized and the centralized economy as shown in the appendix in .4. At the equilibrium, due to the inefficiency in S, they will differ but for given S, K and D, they are identical. The first order condition with respect to S differs, it writes:

$$\begin{aligned} &\int_0^{\overline{\theta}} R^E u'(C_2) d\theta + \int_{\overline{\theta}}^1 R^E u'(C_2^B) d\theta \\ &+ \frac{R^E P^*}{\overline{c}D(R^E - P^*)} \frac{\overline{c}D - K - L}{R^E S - K} \left[E(U_{ND}) - E(U_D) \right] = \mu_1 \end{aligned} \tag{14}$$

where differences with the optimality condition in the decentralized economy have been highlighted in bold characters and where μ_1 is the Lagrange multiplier of the bank problem. These differences imply an externality in the choice of the decentralized bank. The right hand side is the expected marginal return of an additional unit of S expressed in utility terms and the left hand site is the marginal cost. The last term on the left hand side captures the marginal effect of an additional unit of S on the probability of default and is expressed in level of utilities as the payment function is not continuous at $\overline{\theta}$.

The social planner realizes that the choice of aggregate S has an impact on the price $P^* = K/S$: therefore, she knows that his choice of S has an impact on first i) the probability of default $\overline{\theta}$ and second ii) on the payment he can make to depositors in case of default, c^B. First, the impact of S on the probability of default differs from the decentralized economy by a term $\frac{\overline{c}D-K-L}{R^E S-K}$. Second, the expected return of early assets in case of default is $R^E u'(C_2)$ rather than $\theta P^* u'(C_1^B) + [(1-\theta)P^* + R^E]u'(C_2^B)$ in the decentralized economy. These two differences are examined in section 4.

Households' problem

The problem is identical to the decentralized one except for two major differences. First, the constrained social planner now chooses directly the aggregate variables D and K rather than the individual ones D^i and K^i. He recognizes that his decisions have an impact on the insolvency threshold $\overline{\theta}$. Second, the social planner now recognizes the impact of its choice on the fire-sales price and no longer takes it as given.

$E[U_{ND}]$ and $E[U_D]$ defined above are used.

The first order condition with respect to D writes:

$$\begin{aligned} &\int_0^{\overline{\theta}} \theta\overline{c}u'(\overline{c_1})d\theta + \mu_1 = \\ &\int_0^{\underline{\theta}} \theta\overline{c}u'(C_2)d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \theta\overline{c}R^L u'(C_2)d\theta + \lambda + \frac{SR^E(L+K)}{\overline{c}D^2(R^E S - K)}\left(E[U_{ND}] - E[U_D]\right) \end{aligned} \tag{15}$$

The optimality condition 15 ensures that the marginal return expressed in utility terms from investing in deposits (left hand side) is equal to the marginal cost (right hand side). The social planner understands that increasing D relaxes the constraint of the bank: the bank's Lagrange multiplier μ_1 appears in the left hand side as a positive effect on the marginal return of deposits.

The social planner also recognizes the marginal cost of deposits expressed in the opportunity cost of deposits in terms of patient consumers: the two first terms on the right hand side captures this marginal cost. For a liquidity shock between 0 and $\underline{\theta}$, i.e. when there are no sales of assets, an additional unit of deposits reduces marginal patient consumers utility by $\theta\overline{c}u'(C_2)$. Between $\underline{\theta}$ and $\overline{\theta}$, i.e. when there are sales of assets but no default by the bank, the marginal opportunity cost of deposits is higher as an additional units of deposits will be bought by funds and not invested at a rate R^L, hence the additional term R^L, capturing the opportunity cost.

Finally, the social planner takes into account the impact of D on the probability of default by the bank. He recognizes the impact on the fire sale price $P^* = K/S$. The marginal impact of increasing D at the optimum is an increase in the probability of default. This cost is expressed in terms of difference in levels of utility before and after default $E[U_{ND}] - E[U_D]$. This is the last term on the right.

The first order condition with respect to K yields:

$$\int_0^{\overline{\theta}} R^L u'(C_2)d\theta + \int_{\underline{\theta}}^1 \theta u'(c_2^B) + (1-\theta)u'(C_2^B)d\theta$$
$$+ \frac{SR^E(SR^E - \overline{c}D + L)}{\overline{c}D^2(R^E S - K)} \left(E[U_{ND}] - E[U_D]\right) = \lambda \tag{16}$$

Similarly, the optimality condition 16 ensures that the marginal return expressed in utility terms from investing in deposits (left hand side) is equal to the marginal cost (right hand side). The last term on the left hand side represents the marginal impact of increasing K on the probability of default. The social planner understands that its impact on the fire-sales price implies that increasing K decreases the probability of default, hence the additional marginal return of K.

A comparison between the first order conditions in the two economies makes it clear that an inefficiency arises in the households' choice (compare equations 15 and 11 and equations 16 and 12). They do not choose the optimal levels of K and D.

Equilibrium definition

An equilibrium is an allocation made of period 0 choice variables $(D, K, S, L, \overline{c})$, of the solvency threshold $\overline{\theta}$, and in each state θ of period 1 choice variables (X, Y) and of price of early assets P.

Definition 3. *A constrained efficient and rational expectation equilibrium is defined as the equilibrium of an economy in which:*
i) depositors' type (impatient / patient) is private information;
ii) the social planner is submitted to the same technological and institutional constraints as in the decentralized economy;
ii) the solvency condition gives the default threshold of the bank $\overline{\theta}$*;*
iv) the social planner does not take the price of early assets as given;
v) the fund chooses X *and* Y *after the liquidity shock is realized such that the optimality condition 3 and its budget constraint are satisfied; the price of early assets* P *is given by the optimality condition of the funds program when the liquidity shock is below the solvency threshold and by equation 4 when the shock is higher;*
vi) the social planner choose K *and* D *such that the optimality conditions 15 and 16 and the budget constraint are satisfied;*

vii) the social planner chooses S and L such that optimality conditions 8, 9, 14, and the budget constraint are satisfied[10].

4 Pecuniary externality and policy

4.1 The pecuniary externality

In this section, I consider the two inefficiencies sequentially: the distorted choice made by banks and then the distorted choice made by households. I introduce two new - partially efficient - fictitious economies that I compare to the purely inefficient economy. In each fictitious economy, one of the two inefficiencies is shut down in order to be able to draw welfare general equilibrium results.

First, in order to focus on the banks inefficient choice, I compare the purely inefficient economy, i.e. the decentralized economy studied above, called E_{dec} thereafter, to a fictitious economy which includes efficient banks and inefficient households. It is then a partially efficient economy called thereafter E_1. Second, in order to focus on the households inefficient choice, I compare the purely inefficient economy to a second fictitious economy which includes efficient households and inefficient banks. This economy is called thereafter E_2. This method allows to isolate one inefficiency at a time and to understand precisely why each choice is distorted, abstracting from the other efficiency.

4.1.1 Inefficient choice by the bank

I first consider the economy E_1 with efficient banks and inefficient households. The fact that households are inefficient in E_1 and in E_{dec} allows to make the two economies directly comparable and to focus on the inefficient bank choice. The first order conditions with respect to $\overline{c}$ and L are identical in both economies for given S, D and K. Only the first order condition with respect to S differs in the decentralized and in the economy E_1. Therefore, the equilibrium allocation would be optimal if the bank were not choosing an inefficient level of S. They are reproduced here.

The first order condition in the decentralized economy writes:

[10]The first order conditions for $\overline{c}$ and for L have the same expressions than in the decentralized economy: see equations 8, and 9.

$$
\int_0^{\overline{\theta}} R^E u'(C_2) d\theta + \int_{\overline{\theta}}^1 \left[\theta P^* u'(C_1^B) + [(1-\theta)P^* + R^E] u'(C_2^B)\right] d\theta
+ \frac{R^E P^*}{\overline{c}D(R^E - P^*)} \left[E(U_{ND}) - E(U_D)\right] = \mu
$$

And writes in E_1:

$$
\int_0^{\overline{\theta}} R^E u'(C_2) d\theta + \int_{\overline{\theta}}^1 R^E u'(C_2^B) d\theta
+ \frac{R^E P^*}{\overline{c}D(R^E - P^*)} \frac{\overline{c}D - K - L}{R^E S - K} \left[E(U_{ND}) - E(U_D)\right] = \mu
$$

The externality only lies in the choice of S because the bank takes as given 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, when the wealth of funds is too low, the price falls to a level, the fire-sales price $P^* = \frac{K}{S}$, which decreases with the overall amount of S matters.

A preliminary result is:

Lemma 1. $E[U_D] < E[U_{ND}]$.

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

For a known size of the liquidity shock close to $\overline{\theta}$, the expected utility of a given depositor who *ex ante* does not know if he will be impatient or patient is higher before default than after default. Risk averse households fear default *ex ante* because default increases the difference in level of consumption between patient and impatient households.

This result allows stating the following theorem. Recall that the bank's choice happens after the choice of households. Then, banks make their choice for a given D and a given K. In the fictitious efficient-bank-only economy, households are inefficient so that we can directly compare it to the purely inefficient economy. In both economies, households choose the same value of D and of K, before the bank's choice happens.

Lemma 2. *The value of the partial derivative of the Lagrangian with respect to S is greater in the decentralized economy E_{dec} than in the partially efficient economy E_1 for a given choice of households, i.e. for a given value of D, $D = \overline{D}$ and of K, $K = \overline{K}$.*

$$\frac{\partial \mathcal{L}_{dec}}{\partial S}(\overline{D}, \overline{K}) > \frac{\partial \mathcal{L}_{E_1}}{\partial S}(\overline{D}, \overline{K})$$

Proof See proof in appendix .5. □

The partial derivative of the Lagrangian with respect to S is lower once the price is taken into account because increasing S can increase the probability of default by lowering the fire-sales price.

Hence, the following theorem can now be stated:

Theorem 2. *The decentralized bank invests too much in early asset S and does not keep sufficient liquidity buffer compared to the partially efficient economy E_1: $S_{dec} \geq S_{E_1}$*

Proof The result follows directly from *i*) lemma 2, *ii*) from the fact that the derivatives of $\mathcal{L}_{dec}$ and $\mathcal{L}_{E_1}$ with respect to S are decreasing as the utility function U is concave, and *iii*) from the fact that the two order first conditions, with respect to L and $\overline{c}$ are identical in the two economies for given S, D, and K.

□

The externality has two dimensions. The price-taking bank fails to internalize the impact of its choice of S on the price and ultimately on the payment it will be able to make to depositors in case of default; and second on the probability of default. Indeed, both depend on the fire-sales price P^* which depends on S.

Let first consider the payments made by the bank to depositors in case of default. Whenever the bank is defaulting, early assets are sold at a price below the fundamental value due to fire sales. Therefore, the bank gets poorer. When choosing S *ex ante*, the bank cannot realize what is the amount $L + SP^*$ it will be able to pay depositors, as P^* is taken as given. Therefore, the degree of insurance provided is not efficient.

Proposition 3. *The funds' profits increase in case of default, i.e. during a fire sales episode:* $\Pi_{ND} \geq \Pi_D$.

Proof. For $\theta \in [\underline{\theta}, \overline{\theta}]$, i.e. when there is no default, $\Pi_{ND} = KR^L$. For $\theta \in [\overline{\theta}, 1]$, i.e. when the bank defaults, $\Pi_D = SR^E$. As $P^* = \frac{K}{S}$, $KR^L \geq \frac{K}{S}$ so that funds becomes richer in case of default by banks. □

A transfer of wealth from impatient consumers to patient consumers happens through the collapse of the price to P^* that makes funds richer to the detriment of the bank. The price-taking decentralized bank cannot fully realize the magnitude of this transfer of wealth as it takes the price as given, whereas this redistribution matters for social welfare.

Second, consider the probability of default by comparing the partial derivative of the insolvency threshold $\overline{\theta}$ - which is also the probability of being solvent - with respect to S in the decentralized and in the constrained efficient equilibrium. Recall that $(1 - \overline{\theta})$ is the probability of default. The marginal impact of the choice of S by bank in the decentralized economy, knowing that banks take deposits as given is:

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

The marginal impact of the choice of S by the bank, knowing that banks take deposits as given is:

$$\frac{\partial \overline{\theta}}{\partial S_{E_1}} = \frac{\partial \overline{\theta}}{\partial S_{dec}} \frac{\overline{c}_{E_1} D_{E_1} - L_{E_1} - K_{E_1}}{R^E S_{E_1} - K_{E_1}}$$

They only differ by one term, $\frac{\overline{c}D - L - K}{R^E S - K}$ which is the ratio of what the bank is missing if $\theta = 1$, i.e. for the highest liquidity shock, $\overline{c}D - L - P^*S$, to pay out depositors over a measure of the gap between the total liquidity of funds available to buy back early assets K and the whole return on assets sold in case of default $R^E S$. The ratio measures the potential maximum severeness of fire sales in case of the highest liquidity shock $\theta = 1$. The highest the ratio, the highest the externality.

$$\frac{\partial \overline{\theta}}{\partial S_{E_1}} \leq \frac{\partial \overline{\theta}}{\partial S}$$

In the decentralized equilibrium, the partial derivative is always inferior to the partial derivative when the effect on the price is taken into account. The welfare loss arises from a different impact of the choice of S on the probability of default once the effect on the price is internalized. The more early assets S are done in period 0, the more the fire-sales price P^* in period 1 falls in case of default, and consequently, the more likely the default is *ex ante*.

At the equilibrium, the fact that the choice of S is distorted implies an inefficient equilibrium level of $\overline{c}$. The redistribution of wealth between patient and impatient depositors is not efficient even when fire sales do not arise because the bank does not choose the efficient level of $\overline{c}$. Fire sales imply welfare costs even when they do not actually occur. The bank does not achieve the efficient redistribution between patient and impatient in both cases.

4.1.2 Inefficient choice by the households

I now consider a fictitious economy with inefficient banks and efficient households, the economy E_2. There are two first conditions to consider, with respect to K and with respect to D. Again, this is a sequential choice, with the households maximizing over D and K the utility that is maximized by banks over the bank choice variables on a second step. Then, the households choose before the banks, anticipating that the banks will optimally choose $S, L, \overline{c}$ after that. They know the optimal function decisions of the households: $S^* = S^*(D, K)$, $\overline{c}^* = \overline{c}^*(D, K)$ and $L^* = L^*(D, K)$.

Lemma 3. *The value of the partial derivative of the Lagrangian of households with respect to K is smaller in the decentralized economy than in the fictitious partially efficient economy E_2 for given optimal function decisions of the banks: $S^* = S^*(D, K)$, $\overline{c}^* = \overline{c}^*(D, K)$ and $L^* = L^*(D, K)$ and for a given value of the other choice variables of the households, D.*

$$\frac{\partial \mathcal{L}_{dec}}{\partial K}(S^*(D,K), \overline{c}^*(D,K), L^*(D,K), \overline{D}) \leq \frac{\partial \mathcal{L}_{E_2}}{\partial K}(S^*(D,K), \overline{c}^*(D,K), L^*(D,K), \overline{D})$$

Proof See proof in appendix .6. $\square$

Then, for a given D and given choices by banks, the households tends to invest too little in the illiquid contract offered by funds.

Lemma 4. *The value of the partial derivative of the Lagrangian with respect to D is greater in the decentralized economy than in the fictitious partially efficient economy E_2 for given optimal function decisions of the households: $S^* = S^*(D,K)$, $\overline{c}^* = \overline{c}^*(D,K)$ and $L^* = L^*(D,K)$ and for a given value of the other choice variable of the households, K for many specifications of the utility function.*

$$\frac{\partial \mathcal{L}_{dec}}{\partial D}(S^*(D,K), \overline{c}^*(D,K), L^*(D,K), \overline{K}) > \frac{\partial \mathcal{L}_{soc}}{\partial D}(S^*(D,K), \overline{c}^*(D,K), L^*(D,K), \overline{K})$$

Proof See proof in appendix .7. □

Theorem 3. *The decentralized households invests too much in deposits S and too little in the illiquid contracts offered by funds $D_{dec} \geq D_{E_2}$ and $K_{dec} \leq K_{E_2}$.*

Proof. The result follows directly from lemmas 3 and 4, and from the fact that the derivatives of $\mathcal{L}_{dec}$ and $\mathcal{L}_{E_2}$ with respect to D and K are decreasing as the utility function U is concave.

□

now turn to the inefficiency lying the households choice. Two lemmas are needed as intermediary results to establish the theorem.

4.1.3 A numerical example

I now turn to a numerical example. The utility is given by a constant relative risk aversion functional form: $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$.

Parameter	Value
Risk aversion parameter γ	2
Early asset return R^E	1.6
Late asset return R^L	1.2
Households' endowment E	50

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

This over investment of households in the liquid deposit contract offered by banks and the too scarce investment in the illiquid contract offered by funds is due to the fact that households ignore the impact of the fire sale price on the expected payments the bank can make to them. Taking into account the impact of their decisions on the price, the expected payment by bank is lower. The allocation of households between D and K has an impact on S, as S depends on D, and ultimately on the fire-sales price $P^* = K/S$. Indeed, the fire-sales price is exactly determined by the ratio between K and S so depends indirectly on the ratio between K and D. By taking the fire-sales price as given, households fail to internalize the impact of their choice of deposits on the price and then on i) the probability of default and ii) the level of the rates served by banks in case of default. Hence, they invest too much in deposits as they neglect the impact of their choice on P^*.

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 third result is that the fire-sales price in the decentralized economy is lower than in the efficient constrained economy: $P^*_{dec} = 0,01768$ and $P^*_{soc} = 0,304511$ in the constrained efficient economy. This reflects the *ex post* cost: when fire sales happen in the decentralized economy, the situation is worsened as the fire-sales price is lower. A lower fire-sales price indeed implies a lower payment to impatient households.

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. This is the source of the externality: it leads banks to over invest to S as atomistic banks do not realize that every other banks will make the same choice of S, lowering the fire-sales price P^* and increasing the probability of default. The social planner on the contrary looks at the aggregate level of S and internalizes the impact of his choice of S on the price P^*.

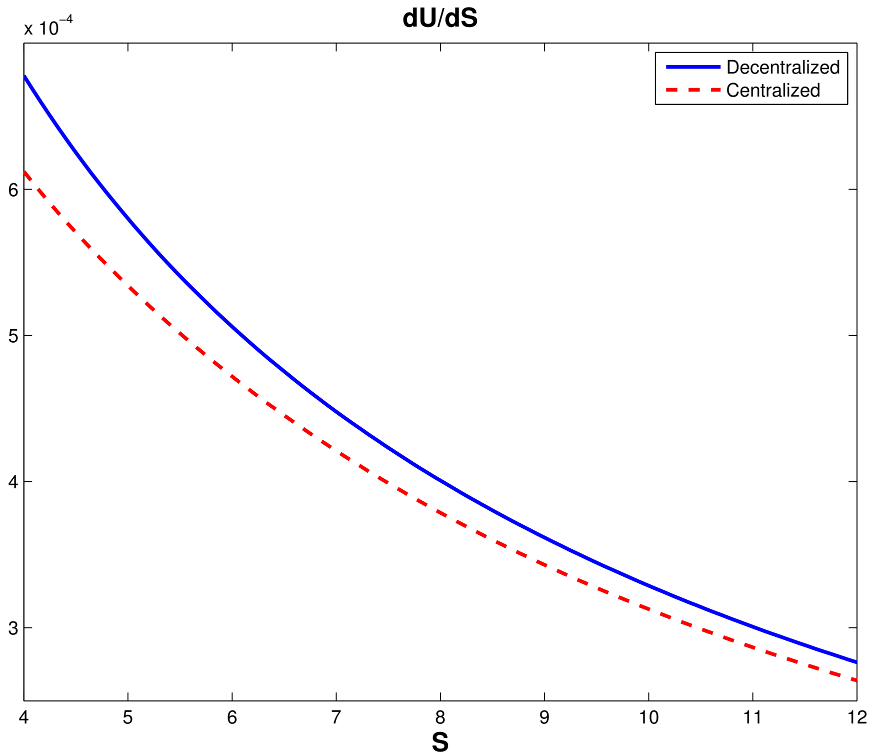

Figure 1: Derivative of the utility with respect to S

4.2 Policy

4.2.1 Liquidity ratios

I first investigate liquidity ratios that constitute an *ex ante* policy. I do not seek to design the optimal policy but rather study the effect of the Basel III Liquidity Coverage Ratio.

Ex post policies could be contemplated, such as a lender of last resort policy. However, beyond moral hazard issue, in this setting, *ex post* policies implemented only in cases of fire sales could not help alleviating the fact that the fixed payment $\bar{c}$ has not been chosen optimally by the bank due to a non optimal choice of S. Therefore, even when fire sales do not occur, the bank operates an inefficient redistribution between patient and impatient. Thus, implementing a policy only when fire sales occur is not sufficient.

Liquidity ratios suggested in the Basel III framework aim at forcing banks to hold more liquid assets. The paper suggests an explanation why ratio might be useful: banks neglect the impact of their decision on the price of assets in case of fire sales and therefore on the probability of default. In the model, the proposed liquidity coverage ratio would be equivalent to forcing bank to hold more reserves L and to 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.

Liquidity ratio and bank choice

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

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

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

$$D = L + S$$

and to the 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 ratios, I first focus on the bank's problem and will then examine the effect on the households' choice. To study how constraining the choice of S can help alleviating the inefficiency in the bank choice, I compare the purely inefficient economy with ratio to the fictitious efficient-banks-only economy, E_1.

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

Theorem 4. *Imposing binding liquidity ratios allows to get the decentralized allocation closer to the constrained efficient allocation. Therefore, binding liquidity ratios increase welfare.*

Proof See proof in appendix .8. □

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. Eliminating the inefficiency lying in bank's choice requires the liquidity ratio to be binding ($\alpha < 1$).

Liquidity ratios cannot reach the constrained efficiency

I now examine the impact of a liquidity regulation on the whole allocation at equilibrium. I will investigate numerically the optimality of liquidity ratios to show that such a regulation cannot reach the first best because liquidity ratios are not equipped to deal with the inefficiency lying in the households' choice.

Theorem 5. *Liquidity ratio cannot reach the constrained efficiency as such an instrument cannot close the inefficiency lying in the households choice.*

This statement is shown numerically, using the same calibration as above for which the utility is given by a constant relative risk aversion functional form: $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ and with:

Parameter	Value
Risk aversion parameter γ	2
Early asset return R^E	1.6
Late asset return R^L	1.2
Households' endowment E	50

The figure 2 presents the expected utility of households - who are *ex ante* identical - in the centralized economy (black line), in the decentralized economy (red line) and in the decentralized economy with liquidity ratio (blue line) on the y-axis as a function of liquidity ratio parameter α on the x-axis. 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. The constrained efficient level of S is set at the vertical red dashed line.

The fact that liquidity ratio do not allow to reach the first best is explained by the fact that ratios cannot solve for the inefficiency lying in the households choice. Decentralized households do not take into account such a constraint that has an impact on aggregate quantities. Therefore, the distortion in the households choice remains when the liquidity ratios constraint is imposed - even if banks are efficient.

4.2.2 Non contingent tax policy

Since liquidity ratios cannot solve for the inefficiency lying in households' choice, another policy may be contemplated to deal with the over investment of households in deposits. The government could implement a non contingent tax t on households withdrawing in period 1 that would be paid to patient households remaining in period 2. This policy is simply a tax on deposits and a subsidy of the illiquid contracts offered by funds and is equivalent to a redistribution between households types (impatient and patient). As the inefficiency in the model takes the form of a failure of insurance between types, such a tax seem well-fitted to restore efficiency. Note that such a policy does not require to observe the type, only to observe which household is withdrawing.

This policy could help solve for the inefficient investment in the illiquid contract because it would increase the expected payment that households receive from the bank. By increasing the expected payment of the illiquid contract, the tax can increase the investment into the illiquid contract of funds by price-taking households to a more efficient level. It aims at creating an incentive for households to invest in

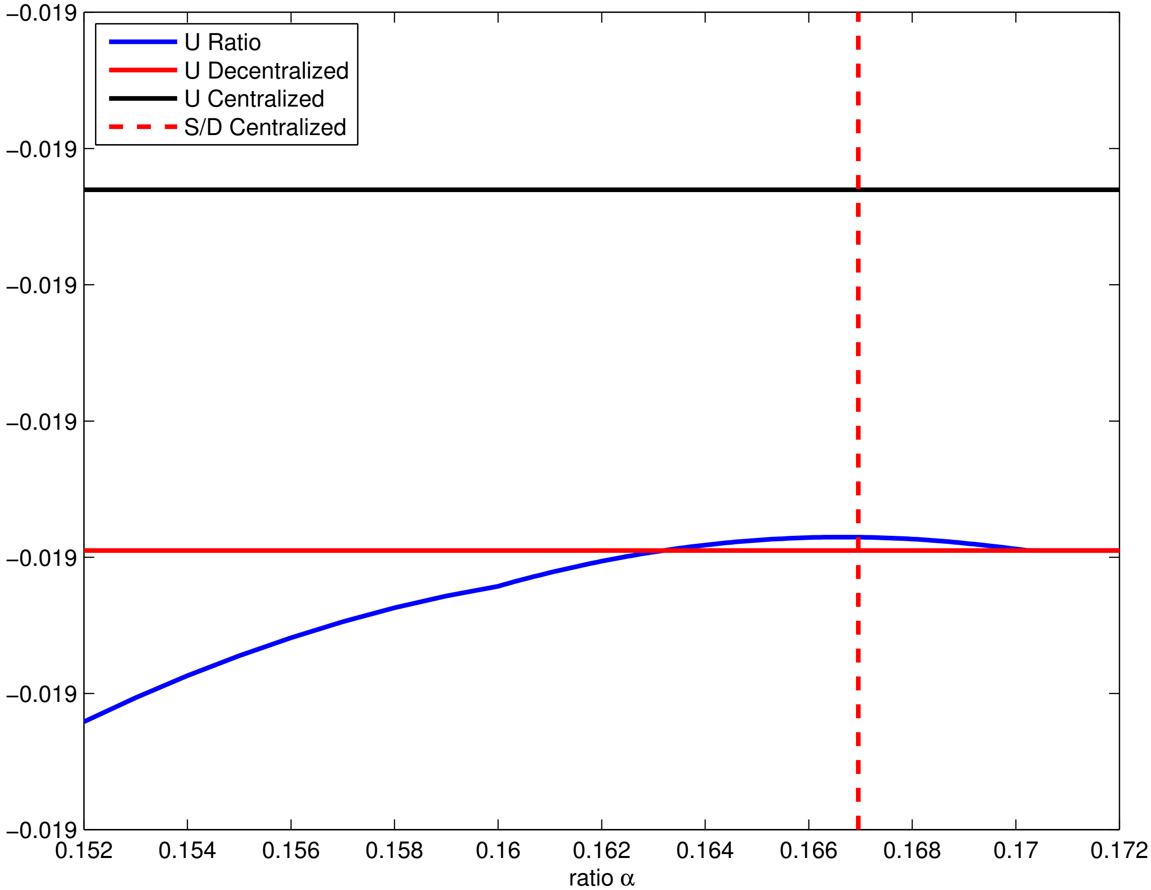

Figure 2: Utility as a function of ratios α

long-term investments as it subsidies such instruments. A positive tax t should lower the level of D and increase the level of K chosen by households, bringing it closer to the efficiency. The tax allows to rebalance the incentives of households that fail to make the efficient choice because they take the price as given, by modifying the expected payments of both assets (deposits or funds shares of profits).

The households withdrawing in period 1 now receive $c_1 = \overline{c}D(1-t)$ when the bank is solvent and the patient receive a subsidy equal to $\frac{\theta\overline{c}Dt}{1-\theta}$. As above, we compare the purely inefficient economy to the fictitious efficient-households-only economy E_2 so that banks have the same behavior: we can focus on the inefficiency lying in households choice. We then obtain the following result by differentiating the modified Lagrangian including the new payments taking into account the tax:

Lemma 5. *For a given allocation,*

$$\frac{\frac{\partial \mathcal{L}_{dec}}{\partial K}}{\partial \tau} \geq 0 \text{ and } \frac{\frac{\partial \mathcal{L}_{dec}}{\partial D}}{\partial \tau} \leq 0.$$

Proof. See the problem of households in presence of a tax t in appendix .9 and the proof of this statement. □

We look at those signs for a given allocation and not at the equilibrium as we are interested in understanding the impact of the modification of the tax on the first order conditions in a first step, to understand in a second step how the tax affects the equilibrium choices.

Recall that $\frac{\partial \mathcal{L}_{dec}}{\partial K} \leq \frac{\partial U_{soc}}{\partial K}$ so that households do not choose to invest enough into the illiquid contract offered by funds K. Therefore, a positive tax $\tau \geq 0$ can help with the distortion.

.1 Proof of proposition 1

As banks maximize their depositors utility, it must be true that the promised rate on deposits weakly dominates the storage marginal return: $\overline{c} \geq 1$. Then using the cash-in-the-market condition, there exists some θ_0 for which $L + K \leq \theta_0\overline{c}D$. Using the fact that $L = D - S$, we get $K < D(\theta\overline{c} - 1) + S \leq S$ as $\overline{c} \geq 1$ and $0 \leq \theta_0 \leq 1$.

The result obtains: $K < S$ so that $P^* = K/S \leq 1$

.2 Proof of proposition 2

$$\lim_{\theta \to \overline{\theta}^-} C_2 = \frac{R^L K}{1-\theta}$$

$$\lim_{\theta \to \overline{\theta}^+} C_2 = L + K + \frac{R^E S}{1-\theta}$$

Using the fact that when θ approaches $\overline{\theta}$, we have $\lim_{\theta \to \overline{\theta}^-}(R^E S - R^L K) \geq 0$ - which is the definition of fire sales -, we have established that:

$$\lim_{\theta \to \overline{\theta}^+} C_2 - lim_{\theta \to \overline{\theta}^-} C_2 \geq 0$$

Therefore, the consumption of patient consumers before default is smaller than the consumption of patient after default as θ approaches $\overline{\theta}$. □

.3 Proof of theorem 3

$R^E - P^* > 0$ as $P \leq R^E/R^L$ using the fact that $R^L \geq 1$. □

.4 Bank's problem in the centralized economy

The first order condition with respect to L writes:

$$\int_0^{\underline{\theta}} u'(\overline{C_2}) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} R^L u'(C_2) d\theta + \int_{\overline{\theta}}^1 [\theta u'(C_1^B) + (1-\theta) u'(C_2^B)] d\theta$$
$$+ \frac{S R^E}{\overline{c} D (R^E S - K)} [E(U_{ND}) - E(U_D)] = \mu$$

Replacing P^* by its value K/S, the optimality condition obtained is exactly equal to the optimality condition in the decentralized economy (see equation 8). At the optimum, they will differ to the inefficient choice of S but everything else held equal, they are similar: the inefficiency does not arise from the choice of $\overline{c}$.

The first order condition with respect to $\overline{c}$ writes:

$$
\begin{aligned}
&\int_0^{\overline{\theta}} \theta D u'(\overline{c_1}) d\theta + \frac{SR^E(L+K)}{\overline{c}^2 D(R^E S - K)} \left[E(U_{ND}) - E(U_D)\right] \\
&= \int_0^{\underline{\theta}} \theta D u'(C_2) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} R^L \theta D u'(C_2) d\theta
\end{aligned}
$$

Replacing P^* by its value K/S, the optimality condition obtained is exactly equal to the optimality condition in the decentralized economy (see equation 9).

.5 Proof of lemma 2

In the decentralized economy E_{dec}, it writes:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{dec}}{\partial S} &= \int_0^{\overline{\theta}} R^E u'(C_2) d\theta + \int_{\overline{\theta}}^1 \left[\theta P^* u'(C_1^B) + [(1-\theta)P^* + R^E] u'(C_2^B)\right] d\theta \\
&+ \frac{R^E P^*}{\overline{c} D(R^E - P^*)} \left[E(U_{ND}) - E(U_D)\right] + \mu
\end{aligned}
$$

And in the fictitious economy E_1, it writes:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{E_1}}{\partial S} &= \int_0^{\overline{\theta}} R^E u'(C_2) d\theta + \int_{\overline{\theta}}^1 R^E u'(C_2^B) d\theta \\
&+ \frac{R^E P^*}{\overline{c} D(R^E - P^*)} \frac{\overline{c} D - K - L}{R^E S - K} \left[E(U_{ND}) - E(U_D)\right] + \mu
\end{aligned}
$$

Recall that the last expression is considered in the fictitious economy E_1 in which bank are inefficient but households are efficient. Then, the value of D and K are the same for the two economies compared. Note that I have used the fact that $\mu_{dec} = \mu_{soc} = \mu$ as households are efficient in the two economies studied for now.

Furthermore, to compare those two first order condition, the allocation is directly comparable everything as the two other first order condition are identical in both economies. Therefore, the threshold are identical, except for the impact a marginal increase in S which is capture by the last term.

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 $\frac{\overline{c}D - K - L}{R^E S - K} \leq 1$. 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[11]. Hence the result for a given value D of K using the lemma 4.1 to compare the two expressions. □

.6 Proof of lemma 3

In the decentralized economy E_{dec}, it writes:

$$\frac{\partial \mathcal{L}_{dec}}{\partial K} = \int_0^{\underline{\theta}} \frac{R^L}{K} u'(C_2) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{R^E X + (K - PX) R^L}{K} \right] u'(C_2) d\theta$$
$$+ \int_{\overline{\theta}}^{1} \frac{R^E S}{K} u'(C_2) d\theta - \lambda$$

And in the fictitious economy E_2, it writes:

$$\frac{\partial \mathcal{L}_{E_2}}{\partial D} = \int_0^{\overline{\theta}} R^L u'(C_2) d\theta + \int_{\underline{\theta}}^{1} \theta u'(c_2^B) + (1-\theta) u'(C_2^B) d\theta$$
$$+ \frac{S R^E (S R^E - \overline{c} D + L)}{\overline{c} D^2 (R^E S - K)} \left(E[U_{ND}] - E[U_D] \right) - \lambda$$

These two functions are considered everything else held equal, hence for the same allocation, for the same threshold: I focus on the marginal impact of S. Then, I compare their position for a given K and a given allocation of banks. Therefore, the threshold are identical.

Using the fact that $\overline{c}D - L \leq R^E$, and that $P^F = R^E / R^L$, it is shown easily that the difference $\left[\frac{\partial \mathcal{L}_{soc}}{\partial D} - \frac{\partial \mathcal{L}_{dec}}{\partial K} \right]$ is positive. □

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

.7 Proof of lemma 4

In the decentralized economy E_{dec}, it writes:

$$\frac{\partial \mathcal{L}_{dec}}{\partial D} = \int_0^{\overline{\theta}} \theta \overline{c} u'(\overline{c_1}) d\theta + \int_0^{\underline{\theta}} \frac{L - \theta \overline{c} D + SR^E}{D} u'(c_2) d\theta$$
$$+ \int_{\underline{\theta}}^{\overline{\theta}} \frac{(S-X)R^E}{D} u'(c_2) d\theta + \int_{\overline{\theta}}^{1} [\frac{\theta(L+SP^*)}{D} u'(c_1^B) + \frac{(1-\theta)(L+SP^*)}{D} u'(c_2^B)] d\theta + \lambda$$

And in the fictitious economy E_2, it writes:

$$\frac{\partial \mathcal{L}_{E_2}}{\partial D} = \int_0^{\overline{\theta}} \theta \overline{c} u'(\overline{c_1}) d\theta - \int_0^{\underline{\theta}} \theta \overline{c} u'(c_2) d\theta$$
$$- \int_{\underline{\theta}}^{\overline{\theta}} R^L \theta \overline{c_2} u'(c_2) d\theta - [E(U_{ND}) - E(U_D)] \frac{SR^E(L+K)}{\overline{c} D^2 (R^E S - K)} + \lambda$$

Replacing μ_1 (which is the Lagrange multiplier from the bank's budget constraint) using the first order condition of the decentralized bank (because the economy E_2 includes inefficient bank), we get:

$$\frac{\partial \mathcal{L}_{E_2}}{\partial D} = \int_0^{\overline{\theta}} \theta \overline{c} u'(\overline{c_1}) d\theta + \int_0^{\underline{\theta}} [R^E - \theta \overline{c}] u'(c_2) d\theta$$
$$+ \int_{\underline{\theta}}^{\overline{\theta}} [R^E - R^L \theta \overline{c}] u'(c_2) d\theta - [E(U_{ND}) - E(U_D)] \frac{R^E}{\overline{c} D (R^E S - K)} \frac{L(S-K)}{D} + \lambda$$

I turn to the determination of the sign of the difference $\frac{\partial \mathcal{L}_{dec}}{\partial D} - \frac{\partial \mathcal{L}_{E_2}}{\partial D}$. If this difference is positive, as it has been shown that $\left[\frac{\partial \mathcal{L}_{soc}}{\partial D} - \frac{\partial \mathcal{L}_{dec}}{\partial K}\right]$ for a given D, it means that the decentralized economy tends to over invest in deposits.

$$\frac{\partial \mathcal{L}_{dec}}{\partial D} - \frac{\partial \mathcal{L}_{E_2}}{\partial D} = \int_0^{\underline{\theta}} \frac{L + SR^E - DR^E}{D} u'(c_2) d\theta + \int_{\underline{\theta}}^{\bar{\theta}} \frac{(S-X)R^E - D(R^E - \theta R^L)}{D} u'(c_2) d\theta$$
$$+ \int_{\bar{\theta}}^1 \left[\frac{\theta(L + SP^* - P^*D)}{D} u'(c_1^B) + \frac{(1-\theta)(L + SP^* - P^*D)}{D} u'(c_2^B) \right] d\theta$$
$$+ \left[E(U_{ND}) - E(U_D) \right] \frac{R^E L (S-K)}{\bar{c} D^2 (R^E S - K)}$$

The three first terms (the three integrals on the right hand side) reflect the marginal utility effect from investing into deposits that can be labeled the marginal return effect. They capture the differential marginal of deposits on the return of the bank that can be labeled the return effect. The last term reflects the utility level effect and captures the riskiness of the bank. The marginal utility level terms are negative while the utility level term is positive. Therefore, the return effect is dominated by the riskiness effect for a risk aversion which is high enough, thus making the difference $\frac{\partial \mathcal{L}_{dec}}{\partial D} - \frac{\partial \mathcal{L}_{E_2}}{\partial D}$ positive.

We want to know for what type of utility functions the statement $\frac{\partial \mathcal{L}_{dec}}{\partial D} \leq \frac{\partial \mathcal{L}_{E_2}}{\partial D}$ is true. For a risk neutral households, the statement is true as $u'(x) = 0$. For a CRRA function $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, a sufficient condition for the statement to be true is to have γ high enough so that u' is negligible enough compared to u.

.8 Proof of theorem 4

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 economy E_1, we have:

$$\frac{\partial \mathcal{L}_{E_1}}{\partial S} = \frac{\partial U_{soc}}{\partial S} - \mu_1$$

As we have shown that in the economy E_1, the value of the derivative of the Lagrangian with respect to S is too high for a given $\bar{c}$, L, K, and D, it follows that a positive μ_2 allows reducing the distortion arising from the inefficient choice of S.

I compare the purely inefficient economy (inefficient households and inefficient banks) to the fictitious efficient-banks-only economy (efficient banks, inefficient households). The households have the same behavior in both economies which makes the comparison straightforward. As demonstrated above, $\frac{\partial U_{dec}}{\partial S} > \frac{\partial U_{soc}}{\partial S}$ for a given D and a given K. 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, $\overline{c}$, K and D are the same in both economies, decentralized and constrained efficient. Therefore, to have $\frac{\partial \mathcal{L}_{dec}}{\partial S}$ closer to $\frac{\partial \mathcal{L}_{soc}}{\partial S}$ for a given D and a given K, the Lagrange multiplier on the ratio constraint must be strictly positive: $\mu_2 > 0$. $\square$

.9 Tax policy

A tax policy modifies all the payments such that:

$$\text{When } \theta < \underline{\theta} \text{ , } \begin{cases} c_1 = \overline{c}D(1-\tau) \\ C_2 = \frac{L-\theta\overline{c}D+SR^E+KR^L}{1-\theta} + \frac{\theta\overline{c}D\tau}{1-\theta} = \frac{L+SR^E+KR^L}{1-\theta} \end{cases}$$

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

I consider the decentralized economy.

$$\frac{\partial \mathcal{L}_{dec}}{\partial D} = \int_0^{\overline{\theta}} \left[\theta\overline{c}(1-\tau)u'(\overline{c_1})d\theta + \frac{L-\theta\overline{c}D\tau+SR^E}{D}u'(c_2) \right] d\theta$$
$$+ \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta\overline{c}(1-\tau)u'(\overline{c_1}) + \frac{(S-X)R^E}{D} \right] u'(c_2)d\theta +$$
$$\int_{\overline{\theta}}^{1} \left[\frac{\theta(L+SP^*)}{D}u'(c_1^B) + \frac{(1-\theta)(L+SP^*)}{D}u'(c_2^B) \right] d\theta$$

Therefore $\frac{\frac{\partial \mathcal{L}_{dec}}{\partial D}}{\partial \tau}$ is negative for a given allocation (we look at the impact of the tax on the first order conditions here).

$$\frac{\partial \mathcal{L}_{dec}}{\partial K} = \int_0^{\bar{\theta}} \frac{KR^L + \theta \bar{c} D\tau}{K} u'(c_2) d\theta + \int_{\underline{\theta}}^{\bar{\theta}} \frac{R^E X + \theta \bar{c} D\tau}{D} u'(c_2) + \int_{\bar{\theta}}^{1} \frac{\theta(R^E S)}{K} u'(c_2^B) d\theta$$

$$\frac{\frac{\partial \mathcal{L}_{dec}}{\partial K}}{\partial \tau} \geq 0$$

Recall that $\frac{\partial \mathcal{L}_{dec}}{\partial K} \leq \frac{\partial U_{soc}}{\partial K}$ so that households do not choose to invest enough into the illiquid contract offered by funds K. Therefore, a positive tax $\tau \geq 0$ can help with the distortion.

A Conclusion

This model explains why fire sales generate a pecuniary externality that reduces welfare in presence of an incomplete market feature, even when households are *ex ante* identical: the welfare loss arises because of an inefficient redistribution between patient and impatient depositors inside the bank. Price-taking banks fail to internalize the impact of their choice of assets on their probability of default and on the rates they will be able to serve depositors in case of default. However, default *per se* is optimal as a way to restore some contingency in the non-contingent banking contract. The decentralized bank chooses too much assets and not enough liquid reserves. Then, the bank fails as a provider of insurance against the idiosyncratic risk of being impatient. It does not ensure the optimal redistribution of wealth between impatient and patient depositors. Crucially it fails to do so even when fire sales do not happen: therefore, *ex post* policies implemented only in case of fire sales - such as lender of last resort - are not sufficient.

The provision of liquidity by funds is not efficient either, because the portfolio allocation by households between a liquid contract, bank deposits, and an illiquid contract, offered by financial institutions akin to pension funds or insurance companies, is not optimal. Households invest too much in deposits and too little in funds shares because they ignore the impact of their choice of deposits and funds shares on the fire-sales price. Consequently, the banking sector is too big compared to the non-banking sector (the funds).

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 address the inefficiency on the households' side. Liquidity ratios are not sufficient to address the welfare loss of fire sales. A non contingent tax in deposits and subsidy on funds shares of profits can reduce the inefficiency in the households' choice.

References

[1] Acharya, Viral V and Shin, Hyun Song and Yorulmazer, Tanju, 2007, "Fire Sales, Foreign Entry and Bank Liquidity," *CEPR Discussion Papers 6309*, C.E.P.R. Discussion Papers.

[2] Acharya, Viral V and Shin, Hyun Song and Yorulmazer, Tanju, 2011, "Crisis Resolution and Bank Liquidity," *Review of Financial Studies*, Society for Financial Studies, vol. 24(6), pages 2166-2205.

[3] Acharya, Viral V and Shin, and Yorulmazer, Tanju, 2008, "Cash-in-the-Market Pricing and Optimal Resolution of Bank Failures," *Review of Financial Studies*, Society for Financial Studies, vol. 21(6), pages 2705-2742, November.

[4] Acharya, Viral V. and S. Viswanathan, 2011, "Leverage, Moral Hazard, and Liquidity," *Journal of Finance*, American Finance Association, vol. 66(1), pages 99-138, 02.

[5] Allen, Franklin and Gale, Douglas, 1998, "Optimal Financial Crises," *Journal of Finance*, American Finance Association, vol. 53(4), pages 1245-1284, 08.

[6] Allen, Franklin and Gale, Douglas, 2004, "Financial Intermediaries and Markets," *Econometrica*, Econometric Society, vol. 72(4), pages 1023-1061, 07.

[7] Allen, Franklin and Gale, Douglas, 2005, "From Cash-in-the-Market Pricing to Financial Fragility," *Journal of the European Economic Association*, MIT Press, vol. 3(2-3), pages 535-546, 04/05.

[8] Allen, Franklin and Gale, Douglas, 2005, "Leverage, Moral Hazard, and Liquidity," *Journal of the European Economic Associations*, MIT Press, vol. 3(2-3), pages 535-546, 04/05.

[9] Bernanke, Ben and Gertler, Mark, 1989, "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, American Economic Association, vol. 79(1), pages 14-31, March.

[10] Carletti, Elena and Leonello, Agnese, 2014, "Credit Market Competition and Liquidity Crises," *CESifo Working Paper Series 4647*, CESifo Group Munich.

[11] Diamond, Douglas W and Dybvig, Philip H, 1983, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, University of Chicago Press, vol. 91(3), pages 401-19, June.

[12] Diamond, Douglas W and Rajan, Raghuram G., 2011, "Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes, and Liquidity," *The Quarterly Journal of Economics*, Oxford University Press, vol. 126(2), pages 557-591.

[13] Freixas, Xavier and Parigi, Bruno Maria, 2008, "Lender of Last Resort and Bank Closure Policy," *CESifo Working Paper Series 2286*, CESifo Group Munich.

[14] Greenwald, Bruce C and Stiglitz, Joseph E, 1986, "Externalities in Economies with Imperfect Information and Incomplete Markets," *The Quarterly Journal of Economics*, MIT Press, vol. 101(2), pages 229-64, May.

[15] Impavido, Gregorio and Musalem, Alberto R. 2000, "Contractual savings, stock, and asset markets," *Policy Research Working Paper Series 2490*, The World Bank.

[16] Impavido, Gregorio and Musalem, Alberto R. and Tressel, 2003, "The impact of contractual savings institutions on securities markets," *Policy Research Working Paper Series 2948*, MIT Press, vol. 3(2-3), The World Bank.

[17] Jacklin, C.J., 1987, "Demand deposits, trading restrictions and risk sharing," in *Contractual arrangements for intertemporal trade*, University of Minnesota Press, 1987.

[18] Kiyotaki, Nobuhiro and Moore, John, 1997, "Credit Cycles," *Journal of Political Economy*, University of Chicago Press, vol. 105(2), pages 211-48, April.

[19] Lorenzoni, Guido, 2008, "Inefficient Credit Booms," *Review of Economic Studies*, Oxford University Press, vol. 75(3), pages 809-833.

[20] Shleifer, Andrei and Vishny, Robert W, 2011, "Fire Sales in Finance and Macroeconomics," *Journal of Economic Perspectives*, American Economic Association, vol. 25(1), pages 29-48, Winter.

[21] Shleifer, Andrei and Vishny, Robert W, 1992, “Liquidation Values and Debt Capacity: A Market Equilibrium Approach,” *Journal of Finance*, American Finance Association, vol. 47(4), pages 1343-66, September.

[22] Stein, Jeremy, 2012, “Monetary Policy as Financial Stability Regulation,” *The Quarterly Journal of Economics*, 127: 57-95, February.