Savings, Efficiency and the Nature of Bank Runs *

Agnese Leonello[†] Caterina Mendicino[‡] Ettore Panetti[§]
Davide Porcellacchia[¶]

June 10, 2021

Abstract

Does the level of deposits matter for bank fragility and efficiency? In a banking model with a consumption-saving decision and endogenous runs, we show that the level of deposits has opposite effects on bank fragility depending on the nature of bank runs. In an economy with fundamental-driven runs only, higher deposits make banks more fragile, while the opposite is true in the presence of panic runs. Depositors fail to internalize this effect. A saving externality arises, leading to excessive fragility and insufficient liquidity provision. The economy features over-saving when runs are only driven by fundamentals, and under-saving with panic runs.

JEL codes: G01, G21, G28

Keywords: endogenous bank runs, fundamental runs, liquidity provision, panic runs, saving externality.

*We are grateful to E. Carletti, T. Eisenbach, M. Gertler, W. Den Haan, A. Kashyap, T. Keister, N. Kiyotaki, A. Krishnamurty, G. Ordoñez, M. Ravn, R. Repullo, J.C. Rochet, I. Schnabel and J. Suarez for useful comments and discussions, as well as to conference and seminar participants at 2020 EEA summer meeting, 2020 CEBRA, Banco de Portugal, Riksbank, European Central Bank, University of Bristol and Rutgers University. The opinions and findings of this paper reflect only the views of the authors, and do not necessarily reflect those of the European Central Bank. A special thank to Luigi Falasconi for outstanding research assistance.

[†]European Central Bank - Directorate General Research - Financial Research.
[‡]European Central Bank - Directorate General Research - Monetary Policy Research.
[§]University of Naples Federico II, UECE and SUERF.
[¶]European Central Bank - Directorate General Research - Monetary Policy Research.

1 Introduction

Banks and their health are important for economic outcomes and have attracted a great deal of attention in policy and academic debates over the years. Bank's reliance on short-term debt as a source of funding has been considered an important source of fragility (e.g., Diamond and Dybvig, 1983; Allen and Gale, 2004; Brunnermeier and Oehmke, 2013; Krishnamurthy, 2010). Noteworthy is the unprecedented increase in total deposits that banks experienced at the start of the COVID-19 crisis (see Li et al., 2020; Levine et al., 2021). Figure 1 illustrates the evolution of total bank deposits in the last years for the U.S. and the Euro Area. Over time, bank deposits featured an increasing trend and experienced a remarkable jump in the first months of 2020. From January to May 2020, bank deposits increased by around \$2 trillion in the U.S. and €1.5 trillion in the Euro Area. In light of this fact, two questions naturally arise. Do more deposits affect bank fragility? If so, do agents correctly internalize this effect when deciding how much to save?

In this paper, we address these two questions through the lens of a bank-run model augmented with a consumption-saving decision. First, we show that the level of deposits has an effect on the probability of a bank run. Moreover, the sign of this effect depends on the nature of the run. More deposits make banks more fragile when runs are only driven by fundamentals. On the contrary, it may make them more stable when runs are driven by panics. Second, we characterize the existence of a *saving externality*: Individual depositors fail to fully internalize the impact of their saving decision on the probability of a bank run. As a result, the allocation is inefficient. The economy features excessive financial fragility, too little bank liquidity provision and an inefficient bank size. In particular, the saving externality leads to over-saving and too large banks when runs are fundamental-driven, while under-saving and excessively small banks may emerge when runs are panic-driven.

To carry out this analysis, we build a model in which the bank run probability is endogenous and, as in Goldstein and Pauzner (2005), it is uniquely determined using the global-game methodology. Moreover, the run probability depends on the terms of the deposit contract. We extend this framework by adding an initial consumption-saving decision. This allows us to endogenize the level of deposits and study its implications for banks' fragility and the welfare properties of the decentralized equilibrium.

To the best of our knowledge, this is the first attempt to study the interaction between consumption-saving decisions and endogenous bank runs. In the bank-run literature it is standard to take as given the amount of deposits and therefore the funds intermediated by banks (e.g. Diamond and Dybvig, 1983; Goldstein and Pauzner, 2005). We show that this apparently innocuous assumption has important implications for the efficiency of the equilibrium. While in standard bank-run frameworks banks issuing demandable deposit contracts can achieve the constrained efficient allocation despite a positive probability

(a) U.S. (b) Euro Area

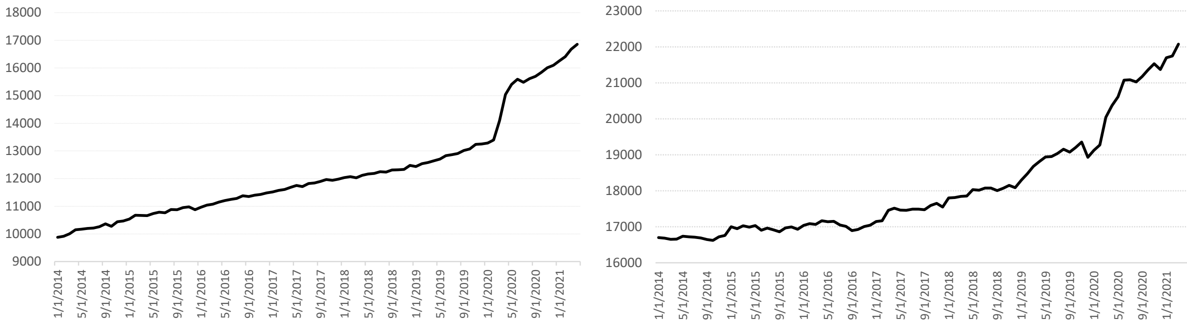

Figure 1: Total deposits in commercial banks, in billions of U.S. dollars in Panel (a) and billions of euros in Panel (b).

of runs, this is not true in our framework. The allocation is inefficient because financial fragility is endogenous to the level of savings and depositors do not fully internalize the effect of their individual saving decisions. This provides a novel rationale for policy intervention.

The model features three dates. At the initial date, ex-ante identical risk-averse agents decide how much to consume and how much to deposit in the banking sector. Aggregate deposits fully determine bank size. Competitive banks issue demand deposits and invest them in a profitable risky project whose returns at the final date depend on the fundamental of the economy. In exchange for the funds provided to banks, depositors are promised a positive deposit rate if they withdraw at an interim date (run) and a higher one if they withdraw at the final date and the bank's investment project is successful. Banks meet early withdrawals by liquidating a share of their long-term investment and, in case they fail to repay the promised deposit rate, depositors receive a pro-rata share of the available resources. Depositors take their individual withdrawal decisions at the interim date on the basis of an imperfect signal on the realization of the economy's fundamental. The signal provides information about both the fundamental and the proportion of depositors running. Depositors run if the fundamental of the economy falls below a unique threshold, which is a function of the terms of the deposit contract and the level of deposits. We can distinguish two types of runs: fundamental- or panic-driven. The former are only due to low realization of the fundamental of the economy. In contrast, the latter are the result of a coordination failure hinging on the fact that depositors' withdrawal decisions are strategic complements. In other words, depositors run out of the fear that others will do the same and so there will not be sufficient resources left for the bank to repay those who decide to wait.

Our analysis provides novel insights about the sources of financial fragility and the efficiency of the decentralized allocation. First, depositors' incentives to run are a function of the level of deposits. When runs are only driven by fundamentals, the probability of a bank run is increasing in the level of deposits. A larger amount of deposits increases the payoffs both at date 1 and 2. Since depositors are risk averse, the increase in the level of deposits is more valuable to them at the date in which their level of consumption is lower. Therefore, as the deposit rate at date 1 is lower than at date 2, higher savings increase the incentives to run more than the incentives to wait.

In the presence of panic runs, the run behavior of each individual depositor depends on how many other depositors she expects will run. In particular, when calculating the expected value of waiting, each depositor assigns a positive probability to the event of a run by almost all other depositors. In this case, banks liquidate almost all their investment at date 1. Hence, the resources left for the depositors that wait are low, as well as their consumption level. Since, due to risk aversion, depositors value more the increase in savings at the date in which consumption is lower, higher savings increase the incentives to wait more than to run. Hence, the probability of a panic run is decreasing in the level of deposits.

Second, the economy exhibits a saving externality. While the social planner internalizes the effect of deposits on the incentive to run, individual depositors do not. In the decentralized economy, banks pool depositors' resources to provide liquidity. Hence, the resources available to service a depositor's withdrawal crucially depend on bank solvency, and not just on the individual's level of deposits. This weakens depositors' incentives to internalize the effects of their saving decision on financial fragility. The saving externality leads to an excessively high probability of bank runs, too little liquidity provision and an inefficient bank size. Crucially, the implications of the saving externality for efficiency depend on the nature of bank runs. When runs are only driven by fundamentals, since the depositors do not internalize that the probability of a panic run is increasing in the level of deposits, the decentralized economy features over-saving with respect to the constrained efficient benchmark. In contrast, in the presence of panic runs the depositors do not internalize that the probability of a panic run may be decreasing in the level of deposits, hence the decentralized economy features under-saving.

This last result brings about a novel motive for public intervention. Since the inefficiency is rooted in the consumption-saving decision, taxes on deposits are an effective tool to reduce the inefficiently high levels of savings in the decentralized equilibrium with only fundamental runs. By reducing aggregate savings ex-ante, a social planner can reduce the payoff at the interim date, and in turn the incentives to run. On the contrary, a subsidy would be optimal in the presence of panic runs.

According to Kashyap and Stein (2012) banks that perform maturity transformation and are subject to runs should always be taxed. We complement their findings by showing that in the presence of a saving externality a tax on financial intermediation is indeed optimal under fundamental-driven runs. However, incentivizing deposits via a subsidy is desirable in the presence of panic runs. Overall, our analysis highlights that the nature of bank runs – whether they are due to low fundamentals or panics – is crucial to understand the inefficiencies associated with the saving externality and the design of optimal policy.

Literature Review. Our paper contributes to three strands of the literature. First, our analysis takes a step forward in understanding the trade-offs associated with the role of banks as liquidity providers (e.g. Diamond and Dybvig, 1983; Goldstein and Pauzner, 2005; Ennis and Keister, 2009). The novelty of our paper is to endogenize consumers' saving decision and study its implication for fragility and efficiency. While financial intermediation in previous contributions is constrained efficient, the saving externality that arises from the depositors' saving decisions leads to an inefficiency which justifies government intervention. Hence, our paper also connects to the literature that studies the efficiency of decentralized banking economies (e.g. Allen and Gale, 2004; Allen et al., 2014).

More closely related to our paper is Peck and Setayesh (2019). In a Diamond-Dybvig framework, they find that a reduction in the share of savings intermediated by banks leads to more financial instability, although the equilibrium allocation remains constrained efficient. Their result differs from ours because they assume a fixed quantity of aggregate savings. Furthermore, an important difference is that our analysis relies on a model of endogenous runs, which allows us to capture both fundamental- and panic-driven runs and their different implications for financial fragility and bank size.

The ability to endogenize the probability of a run relies on the use of global games techniques (e.g., Carlsson and van Damme, 1993; Morris and Shin, 2011). We share with a growing number of papers (e.g. Choi, 2014; Vives, 2014; Eisenbach, 2017; Allen et al., 2018; Ahnert et al., 2019) the use of global games to highlight the inefficiencies associated with the role of banks as financial intermediaries and the desirability of policy intervention.

Second, our analysis is strictly related to the literature that studies the constrained efficiency of decentralized economies in the presence of externalities. Several papers build on the financial frictions as the source of externalities (Hart, 1975; Stiglitz, 1982). The resulting constrained inefficient allocations can be improved upon by policy interventions in financial markets (Geanakoplos and Polemarchakis, 1985). Recent papers have studied the role of pecuniary externalities (Caballero and Krishnamurthy, 2001; Lorenzoni, 2008; Davila and Korinek, 2018), aggregate-demand externalities (Farhi and Werning, 2016; Caballero and Simsek, 2019) and run externalities (Gertler et al., 2020). Still missing from the current debate is an exploration of the role of savers' decisions for fragility and the efficiency of market outcomes. Our work complements existing papers by filling this gap.[1]

[1]Savers' decisions are a key determinant of the size of financial intermediation. In fact, in the period 1896-2012, deposits have represented on average around 80 per cent of U.S. banks' liabilities (Hanson

We share the focus on the role of consumption-saving decisions for the efficiency of the decentralized equilibrium with Davila et al. (2012). They show that in an economy with incomplete markets and idiosyncratic risk the competitive equilibrium is inefficient because the agents do not internalize the effect of their saving choices on the return from capital. We, instead, analyze an economy in which a mechanism to insure against idiosyncratic shocks (i.e. banks) is readily available, but does not ensure the efficiency of the competitive equilibrium. In fact, idiosyncratic-risk pooling brings about the saving externality.

Third, our paper also connects to the literature on the saving glut and financial fragility. Excessive savings around the world bring about excessive leverage, bubbles in asset markets, and other imbalances (Kindleberger and Aliber, 1978; Bernanke, 2005; Caballero and Krishnamurthy, 2009). A handful of papers link over-saving to financial fragility through lower bank incentives to monitor borrowers (Bolton et al., 2016; Martinez-Miera and Repullo, 2017). In our framework, it is instead the nature of bank runs that determines whether higher financial fragility is associated with over- or under-saving. This has important implications for the design of policies.

Paper outline. The paper proceeds as follows. Section 2 describes the baseline model. Section 3 considers the equilibrium in the economy with fundamental runs. We characterize the decentralized economy and then solve for the constrained efficient allocation in order to identify the inefficiency. In Section 4, we follow the same structure and present the economy with panic runs. Section 5 characterizes the optimal policy in the economy with fundamental and panic runs, while Section 6 illustrates the main results through a numerical example. Finally, Section 7 concludes. All proofs are in the Appendix.

2 The baseline model

Our model builds on Goldstein and Pauzner (2005), augmented to include a consumption-saving decision. There are three dates ($t = 0, 1, 2$) and a single good that can be used for consumption and saving. The economy is populated by a continuum of measure one of banks, operating in a competitive market with free entry, and a continuum of measure one of depositors for each bank.

Consumers. Consumers have a unitary endowment of the good at date 0 and nothing thereafter. They can consume at date 0, 1 or 2. At date 1, they face an idiosyncratic liquidity shock: Each of them has a probability λ of being an early consumer (impatient) and a probability $1 - \lambda$ of being a late consumer (patient). Early consumers only want

et al., 2015). Savers' demand for money-like assets has also been recognized as a driver of financial crises and macroeconomic activity (Gorton et al., 2012; Dang et al., 2017).

to consume at date 1, while late consumers are indifferent between consuming at either date. The consumers learn their own realization of the shock privately. The law of large numbers holds, so λ and $1-\lambda$ are also the fraction of consumers who turn out to be early and late, respectively.

The expected utility of a consumer i is given by:

$$U(c_{0i}, c_{1i}, c_{2i}) = u\left(c_{0i}\right) + \lambda u(c_{1i}) + (1-\lambda)u(c_{1i} + c_{2i}), \tag{1}$$

where the utility function satisfies $u'(c) > 0$, $u''(c) < 0$, , $u(0) = 0$ and the coefficient of relative risk aversion $-cu''(c)/u'(c)$ is greater than 1 for any $c \geq \overline{c}$, with $\overline{c} > 0$.[2]

At date 0, each consumer i takes a consumption-saving decision subject to the budget constraint $c_{0i} + d_i = 1$, where c_{0i} is date-0 consumption, and d_i the amount that she deposits in a bank. In exchange for the funds deposited, each bank promises a gross fixed deposit rate r_1 if the consumer withdraws at date 1, and $r_2 > r_1$ if she withdraws at date 2 and the bank is solvent. Banks offers deposit contracts competitively. Thus, they maximize depositors' expected welfare, subject to the budget constraint. This implies that depositors are residual claimants of banks' available resources at date 2, and the repayment r_2 is equal to the return of the non-liquidated units of the bank investment.

Banks. At date 0, banks use total collected deposits D to make an investment I in a productive investment technology, with $I = D$.[3] For each unit invested at date 0, the investment returns 1 if liquidated at date 1 and a stochastic return $\tilde{R}$ at date 2 given by:

$$\tilde{R} = \begin{cases} R > 1 & \text{with prob.} \quad \theta, \\ \quad 0 & \text{with prob.} \quad 1-\theta. \end{cases} \tag{2}$$

We assume that $\mathbb{E}[\theta]R > 1$, which implies that the expected long-term return of the investment is higher than its short-term return. The variable θ, which represents the fundamental of the economy, is uniformly distributed over the interval $[0, 1]$.

Banks satisfy withdrawal demand at date 1 by liquidating the productive investment. So, the per-unit promised repayment at date 2 is a function of the deposit rate r_1, and is given by $r_2 = R\frac{1-\lambda r_1}{1-\lambda}$. Finally, if the liquidation proceeds are not enough to repay the promised deposit rate r_1 to all the withdrawing depositors, a bank is insolvent. In this case, it liquidates all its investment, and distributes the proceeds pro-rata to all the withdrawing depositors at date 1.

[2]We consider very small values of $\overline{c}$ so that relative risk aversion is greater than 1 in equilibrium. In the numerical analysis, we provide an example of utility function that satisfies all the aforementioned hypotheses.

[3]Lower case letters indicate individual variables, and upper case ones aggregate variables.

Information. The fundamental of the economy θ is realized at the beginning of date 1, but publicly revealed only at date 2. At date 1, each consumer receives a private signal x_i about it and decides whether to withdraw at date 1 or wait until date 2. The private signal x_i is of the form:

$$x_i = \theta + \eta_i, \tag{3}$$

where η_i are small error terms, indistinguishable from the true value of the fundamental θ and independently and uniformly distributed over the interval $[-\varepsilon, +\varepsilon]$. A depositor uses her signal to infer both the fundamental of the economy and the probability that other depositors will run on the bank. On this basis, she decides whether to withdraw at date 1 or wait until date 2. As we will shown in detail below, depositors run if the fundamental of the economy θ falls below a unique threshold. In the region in which runs occur, they can be classified either as fundamental-driven, meaning that they are only due to a low realization of θ, or panic-driven, meaning that depositors run lest others do the same. In this case, there will be no resources left for a bank to repay those who decided to wait.

Timing. At date 0, consumers choose how to allocate their unitary endowment between consumption c_{0i} and deposits d_i, and banks set the deposit rate r_1. At date 1, after receiving idiosyncratic liquidity shocks and private signals about the fundamental of the economy θ, early depositors withdraw and late depositors decide whether to withdraw or wait until date 2. At date 2, the banks' investment return is realized and those late depositors who have not withdrawn at date 1 get an equal share of the available resources.

Discussion of the assumptions. As standard in the bank-run literature, the deposit rate r_1 that banks pay at date 1 depends neither on the fundamental nor on the realization of a bank run. Equally, the deposit rate is not a function of the individual amount of deposits. It is conceivable that the repayment offered to depositors could be a function of the amount deposited. For instance, the deposit contract could take the form of a schedule, in which depositors accrue a positive repayment until a certain amount deposited and nothing thereafter. While possible, this is inconsistent with the assumption of a competitive banking sector. Such repayment schedule would create a supply of deposits that are not served. Other banks could attract these with the offer of a lower but positive repayment, thus making strictly positive profits.

In our framework, savings are fully intermediated by banks. Alternatively, one could let consumers invest their savings directly into storage or in the investment technology. In this case, our results would still hold. This is due to the fact that banks provide liquidity insurance. Hence, these alternative investments would be dominated by depositing into a bank.

More generally, as long as we interpret the undeposited endowment as date-0 consumption, it is natural to assume that consumers enjoy a separable utility from it. Alternatively, we could interpret the undeposited endowment as being invested in a different asset. In this case, all our results would still hold as long as utility is separable in bank deposits. This would be akin to modeling deposits in the utility function (e.g. Van Den Heuvel, 2008), and could be rationalized by depositors' preference for liquidity. Introducing non-separable utility would instead require depositors to solve a more involved portfolio choice. This extension would considerably complicate the analysis without affecting its main qualitative insights. In particular, both fundamental- and panic-driven runs would still occur, and depend on how consumers' endowments are allocated. This implies that the introduction of a portfolio choice would not affect the existence of the saving externality either.

3 The economy with fundamental runs

In this section, we study an economy in which panic runs are ruled out by assumption. This case can be thought of as an economy in which an authority, e.g. a lender of last resort, promises to transfer resources to banks in need as long as they are solvent, thus preventing them from fully liquidating their investment to meet early withdrawals. Depositors are guaranteed to receive the promised repayments at both dates, regardless of the share of depositors withdrawing early. We refer to runs occurring in this setting as "fundamental".

We solve the model by backward induction and characterize the symmetric equilibrium. A representative bank chooses the deposit contract, while consumers take their consumption-saving decisions and choose whether to withdraw at date 1 (i.e., "run to the bank") following the threshold strategy:[4]

$$a_i(\underline{\theta}_i) = \begin{cases} \text{withdraw at date 1} & \text{if } \theta \leq \underline{\theta}_i, \\ \text{withdraw at date 2} & \text{if } \theta > \underline{\theta}_i. \end{cases} \tag{4}$$

The definition of equilibrium is as follows:

Definition 1. *A decentralized equilibrium with fundamental runs consists of a set of withdrawal strategies* $\{a_i\}_{i\in[0,1]}$*, vectors of quantities* $\{c_{0i}, d_i\}_{i\in[0,1]}$ *and* $\{D, I\}$*, and a deposit rate* r_1 *such that:*

- *For a given deposit rate* r_1 *and deposits* $\{d_i\}_{i\in[0,1]}$*, the withdrawal strategies* $\{a_i\}_{i\in[0,1]}$ *are chosen optimally;*
- *For given* $\{d_i\}_{i\in[0,1]}$*, the deposit rate* r_1 *maximizes the depositors' expected utility at date 1, subject to the budget constraint* $D = I$*;*

[4] Since the only relevant information for the withdrawal decision is the fundamental θ and the signals are arbitrarily precise, assuming that the withdrawal strategies directly depend on the realization of θ comes at no loss of generality.

- *The consumption-saving decisions $\{c_{0i}, d_i\}_{i\in[0,1]}$ maximize consumers' expected utility at date 0, subject to the budget constraint $c_{0i} + d_i = 1$;*

- *The deposit market clears: $D = \int_i d_i di$.*

3.1 Depositors' withdrawal decisions

We analyze the withdrawal decision of a late depositor i who holds deposit d_i. In the decision, the deposit rate r_1 as well as the amount deposited by others d_{-i} are taken as given. A depositor's run decision depends on whether the bank is solvent. Thus, a depositor i is not concerned about other depositors' withdrawals unless enough depositors run and force the bank into insolvency. For this reason, we start by characterizing the run behaviour of other depositors $-i$.

As panic runs are ruled out by assumption, late depositors choose when to withdraw by comparing the expected utility at date 1 with that at date 2 under the assumption that only early depositors withdraw at date 1. Hence, depositors $-i$ withdraw when θ falls below the threshold $\underline{\theta}(r_1, d_{-i})$ that solves:

$$u(r_1 d_{-i}) = \theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_{-i}\right), \tag{5}$$

and so equals:

$$\underline{\theta}(r_1, d_{-i}) = \frac{u(r_1 d_{-i})}{u\left(R\frac{1-\lambda r_1}{1-\lambda}d_{-i}\right)}. \tag{6}$$

Given (6), the following proposition characterizes depositor i's run strategy.

Proposition 1. *In the economy with only fundamental runs, a late depositor i withdraws at date 1 when θ falls below the threshold:*

$$\underline{\theta}_i = \max\left\{\underline{\theta}(r_1, d_{-i}), \underline{\theta}(r_1, d_i)\right\}, \tag{7}$$

with $\underline{\theta}(r_1, d_{-i})$ as given by (6) and $\underline{\theta}(r_1, d_i) = \frac{u(r_1 d_i)}{u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)}$. The run threshold $\underline{\theta}_i$ is non-decreasing in the amount deposited d_i, i.e., $\frac{\partial \underline{\theta}_i}{\partial d_i} \geq 0$.

The proposition highlights two results. First, depositor i's run decision is driven by the run strategy $\underline{\theta}(r_1, d_{-i})$ of all other depositors. In other words, depositor i has an incentive to run at least as often as others. If everybody else withdraws, depositor i is certain to receive no repayment at date 2, because the bank is insolvent and liquidates all its assets prematurely to serve the other depositors. This case is depicted in the top panel of Figure 2. If the fundamental θ falls in the region $[\underline{\theta}_i, \underline{\theta}_{-i}]$, depositor i does not run while all other depositors $-i$ run. However, waiting until date 2 cannot be optimal since depositor i would be better off by joining the run and withdrawing d_i at date 1. In contrast, depositor i

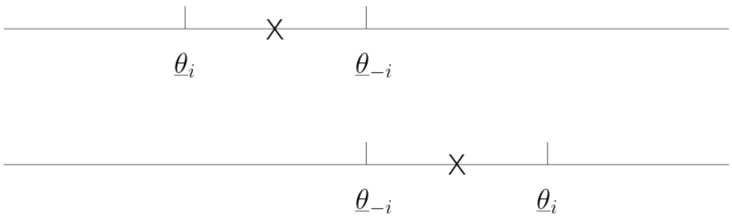

Figure 2: The withdrawal strategy of a late depositor i compared to all other depositors $-i$ in the economy with fundamental runs.

might have incentives to run more often than other depositors. When depositor i is the only late depositor running, the bank is solvent and depositor i is guaranteed to receive positive repayments both at date 1 and 2. As long as $u(r_1 d_i) > u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)$, withdrawing at date 1 when all $-i$ depositors wait until date 2 is optimal. This case is depicted in the bottom panel of Figure 2. If the fundamental θ falls in the region $[\underline{\theta}_{-i}, \underline{\theta}_i]$, the depositor i runs while all other depositors $-i$ do not.

The second result of the proposition is that the run threshold is non-decreasing in the amount deposited. This happens because a higher deposit increases both date-1 and date-2 consumption. However, since depositors are risk averse and their relative risk aversion is larger than 1, they value the increase in consumption more in the state when they are poorer. This happens when they withdraw early, as $r_1 < r_2$. Hence, higher deposits increase the incentives of running over waiting.

3.2 Decentralized economy: saving and deposit rate decisions

Having characterized depositors' withdrawal decisions at date 1, we now solve for the bank's and consumers' decisions at date 0. They choose respectively the deposit rate r_1 and the amount to deposit into a bank d_i.

Bank. The bank chooses the deposit rate r_1 to maximize the utility of a representative depositor i. Thus, it solves the following problem:

$$\max_{r_1} \int_0^{\underline{\theta}(r_1, d_{-i})} u(d_i) d\theta + \int_{\underline{\theta}(r_1, d_{-i})}^{\underline{\theta}(r_1, d_i)} u(r_1 d_i) d\theta + \\ + \int_{\underline{\theta}(r_1, d_i)}^1 \left[\lambda u(r_1 d_i) + (1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)\right] d\theta. \tag{8}$$

The first term represents depositor i's utility when all other depositors run, i.e., when $\theta \leq \underline{\theta}(r_1, d_{-i})$. In this case, the per-unit liquidation value of the bank's investment is 1 and each depositor receives a pro-rata share. Hence, she consumes d_i. The second term represents depositor i's utility when she is the only one to run, i.e. when $\underline{\theta}(r_1, d_{-i}) < \theta \leq \underline{\theta}(r_1, d_i)$. In this case, she obtains $r_1 d_i$. Finally, the third term captures the utility in the absence of runs. When no depositor runs, i.e. for $\theta > \underline{\theta}(r_1, d_i)$, a depositor i receives $r_1 d_i$

if impatient, while if patient she receives a share of bank's available resources $R\frac{1-\lambda r_1}{1-\lambda}d_i$ with probability θ, and zero otherwise.

Consumers. At date 0, each consumer chooses d_i to maximize her utility by solving:

$$\max_{d_i} u(1-d_i) + \int_0^{\underline{\theta}(r_1,d_{-i})} u(d_i)d\theta + \int_{\underline{\theta}(r_1,d_{-i})}^{\underline{\theta}(r_1,d_i)} u(r_1 d_i)d\theta + \\ + \int_{\underline{\theta}(r_1,d_i)}^{1} \left[\lambda u(r_1 d_i) + (1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)\right] d\theta, \tag{9}$$

subject to the budget constraint $d_i = 1 - c_{0i}$, with $\underline{\theta}(r_1, d_{-i})$ as given by (6).

The following proposition characterizes the decentralized equilibrium with fundamental runs.

Proposition 2. *The decentralized equilibrium with fundamental runs is given by $r_1 > 1$ and $d > 0$ that solve:*

$$\int_{\underline{\theta}(r_1,d)}^{1} \left[u'(r_1 d) - \theta R u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right] d\theta - \frac{\partial\underline{\theta}(r_1,d)}{\partial r_1}\frac{\Delta}{\lambda d} = 0, \tag{10}$$

$$u'(1-d) = \int_0^{\underline{\theta}(r_1,d)} u'(d)d\theta + \\ + \int_{\underline{\theta}(r_1,d)}^{1} \left[\lambda r_1 u'(r_1 d) + \theta R(1-\lambda r_1)u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right] d\theta, \tag{11}$$

$$d_i = d_{-i} = d = D, \tag{12}$$

where $\Delta = \lambda u(r_1 d) + (1-\lambda)\underline{\theta}(r_1,d)u\left(R\frac{1-\lambda r_1}{1-\lambda}d\right) - u(d)$*, and* $\underline{\theta}(r_1,d)$ *comes from* (6) *evaluated at* $d_{-i} = d_i = d$*.*

In choosing the deposit rate r_1, the bank compares marginal benefit and marginal cost. The marginal benefit, represented by the first term in (10), captures improved risk sharing owing to the transfer of consumption from late to early consumers. Hence, the deposit rate r_1 can be interpreted as a measure of liquidity insurance. In the equilibrium, the bank finds it optimal to set $r_1 > 1$, thus providing liquidity insurance to depositors. The marginal cost, represented by the second term of (10), is instead the loss in expected utility Δ due to the increased probability of a run, as measured by the derivative of the run threshold $\underline{\theta}(r_1, d)$ with respect to r_1.

The choice of the deposit d again trades off marginal cost and marginal benefit. The former comes from less consumption at time 0, as captured by the left-hand side of (11). The latter comes from more consumption at date 1 and 2, as captured by the right-hand side of (11). Importantly, in (11) there is no term capturing the effect of the amount deposited d_i on the run threshold. The reason is twofold. First, the individual depositor

i cannot influence the run threshold of all other depositors $\underline{\theta}(r_1, d_{-i})$, i.e., $\frac{\partial \underline{\theta}(r_1, d_{-i})}{\partial d_i} = 0$. Second, the depositor can choose to run more often than all other depositors, with $\theta(d_i)$ being the relevant run threshold. In this case, the amount she deposits directly affects the threshold as shown in Proposition 1. However, in this case the cost of the increased run probability for the individual in terms of lost expected utility is zero. In other words, an individual depositor does not perceive a marginal increase in the probability of withdrawing as costly for her, because she is withdrawing optimally given the deposit rate r_1. In summary, consumers do not internalize the effect of the quantity of deposits on the probability of a fundamental run. This highlights the existence of a "saving externality", which, as we show in details below, has important implications for the efficiency of the allocation.

We can substitute (12) and (10) into (11) and obtain an expression summarizing the decentralized equilibrium with fundamental runs:

$$u'(1-D) = \int_0^{\underline{\theta}(r_1,D)} u'(D) d\theta + \int_{\underline{\theta}(r_1,D)}^1 u'(r_1 D) d\theta - (1-\lambda r_1) \frac{\Delta}{\lambda D} \frac{\partial \underline{\theta}(r_1, D)}{\partial r_1}. \tag{13}$$

The above equation resembles an Euler equation, as typically used in dynamic macroeconomic models: It determines the equilibrium level of savings as the quantity that equates their marginal cost and benefit in terms of present vs. expected future consumption. In the rest of the analysis, we use this equation to compare the decentralized equilibrium with the constrained efficient allocation.

3.3 Constrained efficient allocation

In order to study the efficiency of the decentralized equilibrium with fundamental runs, we characterize a constrained-efficient benchmark. We consider a social planner who can only offer demand-deposit contracts like the banks. As a consequence, the planner is subject to runs in the same way as banks: It takes as given depositors' withdrawal strategies, as characterized by the run threshold $\underline{\theta}_i$ in (7), when $i = -i$.

At date 0, the planner allocates $C_0 = 1 - D$ resources to consumption, the remaining D units to bank deposits and chooses the deposit rate r_1 to maximize expected welfare:

$$u(1-D) + \int_0^{\underline{\theta}(r_1,D)} u(D)\, \mathrm{d}\theta + \int_{\underline{\theta}(r_1,D)}^1 \lambda u\left(r_1 D\right) + (1-\lambda)\theta u \left(R \frac{1-\lambda r_1}{1-\lambda} D \right) \mathrm{d}\theta. \tag{14}$$

The following lemma characterizes the constrained-efficient allocation with fundamental runs:

Lemma 1. *The constrained-efficient equilibrium with fundamental runs is given by* $r_1 > 1$

and $D > 0$ that solve:

$$\int_{\underline{\theta}(r_1,D)}^{1} \left[u'(r_1 D) - \theta R u' \left(R \frac{1-\lambda r_1}{1-\lambda} D \right) \right] d\theta - \frac{\partial \underline{\theta}(r_1,D)}{\partial r_1} \frac{\Delta}{\lambda D} = 0, \tag{15}$$

$$\begin{aligned} u'(1-D) = & \int_{0}^{\underline{\theta}(r_1,D)} u'(D)\,\mathrm{d}\theta + \\ & + \int_{\underline{\theta}(r_1,D)}^{1} \left[\lambda r_1 u'(r_1 D) + \theta R (1-\lambda r_1) u' \left(R \frac{1-\lambda r_1}{1-\lambda} D \right) \right] \mathrm{d}\theta - \frac{\partial \underline{\theta}(r_1,D)}{\partial D} \Delta, \end{aligned} \tag{16}$$

where $\Delta = \lambda u\left(r_1 D\right) + (1-\lambda)\underline{\theta}(r_1,D) u\left(R \frac{1-\lambda r_1}{1-\lambda} D\right) - u(D)$

The planner chooses the optimal deposit rate $r_1 > 1$ in the same way as the bank. In other words, for given amount of aggregate deposits D, liquidity insurance in the decentralized equilibrium is constrained efficient. The constrained-efficient allocation differs from the equilibrium of the decentralized economy only for the last term on the right-hand side of equation (16). Relative to the decentralized economy, the social planner internalized the saving externality, accounting for the effect of deposits on the likelihood of fundamental runs and the costs associated with it.

To ease the comparison with the decentralized economy, it is useful to substitute (15) into (16) and obtain:

$$u'(1-D) = \int_{0}^{\underline{\theta}(r_1,D)} u'(D) d\theta + \int_{\underline{\theta}(r_1,D)}^{1} u'(r_1 D) d\theta - (1-\lambda r_1) \frac{\Delta}{\lambda D} \frac{\partial \underline{\theta}(r_1,D)}{\partial r_1} - \frac{\partial \underline{\theta}(r_1,D)}{\partial D} \Delta. \tag{17}$$

The following proposition compares the social planner allocation with the decentralized equilibrium.

Proposition 3. *The decentralized equilibrium with fundamental runs is not constrained efficient. It exhibits over-saving, excessive financial instability and an inefficient level of bank liquidity insurance.*

By internalizing the effects of aggregate savings D on financial instability, the social planner chooses a lower level of savings than in the decentralized equilibrium. In other words, in the decentralized equilibrium, consumers save too much, because they do not internalize the adverse effect of their saving decision on financial stability and runs are too frequent. The inefficiency emerging in the decentralized economy represents a novel result relative to the existing literature on bank runs. In Diamond and Dybvig (1983) and subsequent related papers (e.g., Goldstein and Pauzner, 2005), banks achieve the constrained-efficient allocation by providing liquidity insurance to risk-averse depositors. In our framework, banks provide liquidity insurance to depositors. However, the equilibrium level of insurance is not constrained efficient.

4 The economy with panic runs

In this section, we relax the assumption that the economy suffers no panic runs and consider the strategic complementarities in withdrawing decisions. Late depositors are residual claimants of bank's available resources, i.e. $r_2 = R\frac{1-nr_1}{1-n}$, with $n \geq \lambda$ denoting the proportion of early withdrawing depositors. As long as the deposit rate r_1 is larger than 1, more depositors withdrawing lowers the per-unit investment return r_2 at date 2. Therefore, the incentives for a late consumer to join a run increase with the proportion of early withdrawing depositors n.

As in the previous section, we solve the model by backward induction and characterize the symmetric equilibrium. A representative bank chooses the deposit contract, all consumers take the consumption-saving decision, and late ones, based on their signals, decide when to withdraw following the threshold strategy:[5]

$$a_i(x_i) = \begin{cases} \text{withdraw at date 1} & \text{if } x_i \leq x_i^*, \\ \text{withdraw at date 2} & \text{if } x_i > x_i^*. \end{cases} \tag{18}$$

The definition of equilibrium is as follows:

Definition 2. *A decentralized equilibrium with panic runs consists of a set of withdrawal strategies $\{a_i\}_{i\in[0,1]}$, vectors of quantities $\{c_{0i}, d_i\}_{i\in[0,1]}$ and $\{D, I\}$ and a deposit rate r_1 such that:*

- *For a given deposit rate r_1 and deposits $\{d_i\}_{i\in[0,1]}$, upon receiving the signal x_i, depositors' beliefs about early withdrawals are updated according to the Bayes rule, and the withdrawal strategies $\{a_i\}_{i\in[0,1]}$ are chosen optimally;*
- *For given $\{d_i\}_{i\in[0,1]}$, the deposit rate r_1 maximizes the depositors' expected utility at date 1, subject to the budget constraint $D = I$;*
- *The consumption-saving choices $\{c_{0i}, d_i\}_{i\in[0,1]}$ maximize depositors' expected utility at date 0, subject to the budget constraint $c_{0i} + d_i = 1$;*
- *The deposit market clears: $D = \int_i d_i di$.*

4.1 Depositors' withdrawal decision

We analyze depositors' withdrawal decisions at date 1 for a given deposit rate r_1 and amount deposited d_i. Early consumers always withdraw at date 1 to satisfy their consumption needs. In contrast, late consumers decide whether to withdraw at date 1 based

[5] Selecting threshold strategies comes at no loss of generality, as Goldstein and Pauzner (2005) show in a similar environment that every equilibrium strategy is a threshold strategy.

on the signal x_i that they receive, since this provides information on both the fundamental θ and other depositors' actions.

Upon receiving a high signal, a late consumer attributes a high posterior probability to a positive bank project return R at date 2, and infers that the other late consumers have also received a high signal. This lowers her belief about the likelihood of a run and thus her own incentive to withdraw at date 1. Conversely, when the signal is low, the opposite happens and a late consumer has a high incentive to withdraw early. This suggests that late consumers withdraw at date 1 when the signal is sufficiently low, and wait until date 2 when the signal is sufficiently high.

To show this formally, we first examine two regions of extremely bad and extremely good fundamentals, where each late consumer's action is based on the realization of the fundamental irrespective of beliefs about other agents' behavior.

Lower dominance region. The lower dominance region of θ corresponds to the range $[0, \underline{\theta}]$ in which fundamentals are so bad that running is a dominant strategy. Upon receiving a signal indicating that the fundamentals are in the lower dominance region, a late consumer is certain that the expected utility from waiting until date 2 is lower than that from withdrawing at date 1, even if only λ early depositors were to withdraw. The expected utility from waiting equals $\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)$, given that $\frac{R(1-\lambda r_1)}{1-\lambda}$ is the per-unit return of deposit when only λ depositors withdraw. The expected utility from withdrawing at date 1 instead equals $u(r_1 d_i)$. Then, we denote by $\underline{\theta}(r_1, d_i)$ the value of θ that solves:

$$u(r_1 d_i) = \theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right), \tag{19}$$

that is:

$$\underline{\theta}(r_1, d_i) = \frac{u(r_1 d_i)}{u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)}. \tag{20}$$

We refer to the interval $[0, \underline{\theta}(r_1, d_i)]$ as the lower dominance region, where runs are only driven by bad fundamentals. For the lower dominance region to exist for any $r_1 \geq 1$, there must be feasible values of θ for which all late depositors receive signals that assure them to be in this region. Since the noise contained in the signal x_i is at most ε, each late depositor withdraws at date 1 if she observes $x_i < \underline{\theta}(r_1, d_i) - \varepsilon$. It follows that all depositors receive signals that assure them that θ is in the lower dominance region when $\theta < \underline{\theta}(r_1, d_i) - 2\varepsilon$. Given that $\underline{\theta}$ is increasing in r_1, the condition for the lower dominance region to exist is satisfied for any $r_1 \geq 1$ if:

$$\underline{\theta}\left(1, d_i\right) = \frac{u\left(d_i\right)}{u\left(Rd_i\right)} > 2\varepsilon. \tag{21}$$

Upper dominance region. The upper dominance region of θ corresponds to the range $(\overline{\theta}, 1]$ in which fundamentals are so good that waiting is a dominant strategy. As Goldstein and Pauzner (2005), we construct this region by assuming that in the range $(\overline{\theta}, 1]$ the investment is safe, i.e. $\theta = 1$, and yields the same return $R > 1$ at dates 1 and 2. This means that, given that n depositors run, a late depositor expects to receive a repayment $\frac{R-nr_1}{1-n}d_i > r_1 d_i$ since $R - r_1 > 0$ is required for the contract to be incentive compatible (i.e. $R - r_1 > 0$ is implied by $r_1 < r_2 \equiv \frac{R(1-\lambda r_1)}{(1-\lambda))}$). Then, upon observing a signal indicating that the fundamentals θ are in the upper dominance region, a late consumer is certain to receive her payment $\frac{R(1-\lambda r_1)}{(1-\lambda)}D$ at date 2, irrespective of her beliefs about other depositors' actions, and thus she has no incentives to run. As before, the upper dominance region exists if there are feasible values of θ for which all late depositors receive signals that assure them to be in this range. This is the case if $\overline{\theta} < 1 - 2\varepsilon$.

The intermediate region. The existence of the lower and upper dominance region guarantees the existence of a threshold θ^* in the intermediate region $(\underline{\theta}(r_1, d_i), \overline{\theta}]$, in which a depositor's decision to withdraw early depends on the realization of θ as well as on her beliefs regarding other late consumers' actions.

The characterization of the equilibrium run threshold θ^* consists of two steps. First, we show that no depositor has an incentive to deviate from the run strategy of all the others. Second, we characterize the run threshold θ^*. We have the following lemma.

Lemma 2. *Assume all depositors $-i$ run when their signals $x_{-i} \leq x^*_{-i}$. Then, a depositor i follows the same withdrawal strategy, i.e. she withdraws if $x_i \leq x^*_{-i}$.*

The above lemma shows that, from the point of view of a single depositor i, when the fundamental lies in the intermediate region, it is optimal to follow the withdrawal strategy x^*_{-i} of all the other depositors $-i$. It follows that all depositors withdraw if their signals are lower than a common threshold x^*_{-i} which everyone takes as given. This result hinges on two arguments. First, large withdrawals of deposits at date 1 force the bank to liquidate its assets prematurely, leaving no resources for those who wait and thus bringing about strategic complementarities between depositors' actions. Second, being in the region above the fundamental run threshold implies that it is never optimal for a depositor to run when she expects all other late depositors to withdraw at date 2.

Having established that the relevant run threshold is x^*_{-i}, we now compute it. We start by specifying the utility differential between withdrawing at date 2 and at date 1 for a representative late consumer with deposit d_{-i}. This is given by:

$$\mathcal{V}_{-i}(\theta, n) = \begin{cases} \theta u\left(R\frac{1-nr_1}{1-n}d_{-i}\right) - u(r_1 d_{-i}) & \text{if } \lambda \leq n \leq \overline{n}, \\ 0 - u(\frac{d_{-i}}{n}) & \text{if } \overline{n} \leq n \leq 1, \end{cases} \tag{22}$$

where n represents the proportion of depositors withdrawing at date 1 and $\overline{n} = 1/r_1$ is the value of n at which the bank exhausts its resources if it pays $r_1 > 1$ to all withdrawing depositors. For $n \leq \overline{n}$, a depositor who waits obtains $\frac{R(1-nr_1)}{(1-n)}$ with probability θ for each unit d_{-i} deposited, while an early withdrawer obtains r_1. By contrast, for $n \geq \overline{n}$ the bank liquidates its entire investment at date 1. Late depositors receive either nothing if they wait until date 2 or the pro-rata share $\frac{d_{-i}}{n}$ if they withdraw early.

The function $\mathcal{V}_{-i}(\theta, n)$ decreases in n for $n \leq \overline{n}$ and increases in it afterwards, crossing zero once for $n \leq \overline{n}$ and remaining always below afterwards. Thus, the model exhibits the property of one-sided strategic complementarity and there exists a unique equilibrium in which a late depositor $-i$ runs if and only if her signal is below the threshold $x^*(r_1, d_{-i})$. At this signal value, a late depositor is indifferent between withdrawing at date 1 and waiting until date 2. The following proposition holds.

Proposition 4. *In the economy with panic runs, each late depositor i runs if she observes a signal below the threshold $x^*(r_1, d_{-i})$ and does not run above. At the limit, as the error term $\varepsilon \to 0$, the threshold $x^*(r_1, d_{-i})$ simplifies to:*

$$\theta^*(r_1, d_{-i}) = \frac{\displaystyle\int_{\lambda}^{\overline{n}} u(r_1 d_{-i}) dn + \int_{\overline{n}}^{1} u\left(\frac{d_{-i}}{n}\right) dn}{\displaystyle\int_{\lambda}^{\overline{n}} u\left(R\frac{1-nr_1}{1-n} d_{-i}\right) dn}. \tag{23}$$

The threshold $\theta^*(r_1, d_{-i})$ is increasing in r_1 and decreasing in d_{-i} if $\lim_{c \to 0^+} u'(c)c$ is sufficiently large.

The proposition states that in the intermediate region a late depositor's action depends uniquely on the signal that she receives, as this provides information both on the fundamental of the economy θ and on the other depositors' actions. For θ in the interval $(\underline{\theta}(r_1, d_{-i}), \theta^*(r_1, d_{-i})]$ there are strategic complementarities in depositors' withdrawal decisions. If $r_1 > 1$, the bank has to liquidate more than one unit for each withdrawing depositor, which implies that late depositors' incentives to run increase with the proportion n of depositors withdrawing early. In the limit case when $\varepsilon \to 0$, all late depositors behave alike as they receive approximately the same signal and take the same action. This implies that only complete runs, where all late depositors withdraw at date 1, occur. In what follows, we focus on this limit case, and so the run threshold θ^* is the probability of a run.[6]

In this economy, late depositors run because they fear that other depositors would withdraw early, thus leaving no resources for the bank to pay them. Put differently, in the intermediate region of fundamentals, runs are due to a coordination failure among depositors, and thus we refer to them as "panic-driven".

[6] In the limit case $\varepsilon \to 0$, the probability of a run is equal to the probability that θ falls below θ^*. Since $\theta \sim U[0, 1]$, $Pr(\theta \leq \theta^*) = \theta^*$.

As in the economy with fundamental runs only, the run threshold $\theta^*(r_1, d_{-i})$ increases with the deposit rate r_1 offered by banks. In fact, an increase in r_1 increases depositors' repayment at date 1, while decreasing that at date 2. As a consequence, depositors' incentive to run becomes higher.

Unlike the probability of a fundamental run, the probability of a panic run is not always increasing in the size of the individual deposit d_{-i}. An increase in d_{-i} may reduce the incentive to run and so the threshold $\theta^*(r_1, d_{-i})$. As in Section 3, a rise in the deposited amount increases depositors' repayment at both date 1 and 2. However, in the context of panic runs, a late depositor attaches a positive probability to the possibility of receiving almost zero consumption at date 2 when the proportion of depositors withdrawing early approaches $n = \overline{n}$. In other words, differently from fundamental runs, a depositor expects her date-2 consumption and utility to fall below those at date 1 when a large proportion of depositors runs. As a result, the marginal effect on the run threshold of an increase in the amount deposited, as measured by $u'(c)c$, is high in such states. Overall, when $u'(c)c$ becomes very large as c approaches zero, the increase in deposit has a stronger marginal effect on the expected utility of withdrawing at date 2 than at date 1, thus inducing depositors to run less. In what follows, we focus on the case when $\lim_{c\to 0^+} u'(c)c$ is sufficiently large, so that an increase in the level of deposits leads to a reduction in the likelihood of panic runs θ^*. In Section 6, we provide an example of a functional form for utility that satisfies this property.

4.2 Decentralized economy: saving and deposit rate decisions

Having analyzed the depositors' decision to run, we now characterize the terms of the deposit contract r_1, and the consumption-saving decision at date 0.

Bank. Given the aggregate amount deposited and anticipating depositors' withdrawal decision, as summarized by the run threshold $\theta^*(r_1, d_{-i})$, the bank chooses r_1 to maximize the expected utility of a representative depositor i by solving the following problem:

$$\max_{r_1} \int_0^{\theta^*(r_1,d_{-i})} u(d_i)d\theta + \int_{\theta^*(r_1,d_{-i})}^1 \left[\lambda u(r_1 d_i) + (1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)\right] d\theta. \tag{24}$$

The first term represents the expected utility from depositing at a bank, when the fundamental of the economy lies below θ^*. In this case, all depositors run and receive back their initial deposits d_i. The second term is the expected utility when θ is above θ^*. In this case the bank continues operating until date 2, λ early depositors receive $r_1 d_i$, and $1-\lambda$ late depositors receive a pro-rata share of the residual resources with probability θ and zero otherwise.

Consumers. At date 0, each consumer i chooses the amount to deposit d_i and the date-0 consumption c_{0i} to maximize her utility by solving:

$$\max_{d_i, c_{0i}} u(c_{0i}) + \int_0^{\theta^*(r_1, d_{-i})} u(d_i) d\theta + \int_{\theta^*(r_1, d_{-i})}^1 \left[\lambda u(r_1 d_i) + (1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)\right] d\theta, \tag{25}$$

subject to the budget constraint $d_i = 1 - c_{0i}$. At date 0, higher d_i reduces the amount c_{0i} available for consumption. At date 1, if there is a run all consumers get back the deposit d_i. If there is no run, impatient depositors get $r_1 d_i$ at date 1, while patient depositors receive a share of the residual banks' resources at date 2. Notice that, as proved in Proposition 4, from the point of view of a single depositor i the run threshold is only a function of the deposit rate r_1 and of the deposit decisions d_{-i} of everybody else, and not of the individual amount deposited d_i. Therefore, when deciding how much to deposit, the consumer does not internalize the impact of her own savings on the probability of a run.

Having described the bank's and consumers' problems, the following proposition characterizes the decentralized equilibrium with panic runs.

Proposition 5. *The decentralized equilibrium with panic runs is given by $r_1 > 1$ and $d > 0$ that solve:*

$$\int_{\theta^*(r_1,d)}^1 \left[u'(r_1 d) - \theta R u'\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] d\theta - \frac{\partial \theta^*(r_1, d)}{\partial r_1}\frac{\Delta}{\lambda d} = 0, \tag{26}$$

$$\begin{aligned} u'(1-d) = &\int_0^{\theta^*(r_1,d)} u'(d) d\theta + \\ &+ \int_{\theta^*(r_1,d)}^1 \left[\lambda r_1 u'(r_1 d) + \theta R(1-\lambda r_1) u'\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] d\theta, \end{aligned} \tag{27}$$

respectively and

$$d_i = d_{-i} = d = D, \tag{28}$$

where $\Delta = \lambda u(r_1 d) + (1-\lambda)\theta^* u\left(R\frac{1-\lambda r_1}{1-\lambda} d\right) - u(d)$, *and* $\theta^*(r_1, d)$ *comes from* (23) *when* $d_{-i} = d$.

In choosing r_1, the bank trades off its marginal benefit with its marginal cost. The former, represented by the first term in (26), captures improved risk sharing obtained from the transfer of consumption from late to early consumers. The latter, represented by the second term of (26), is the loss in expected utility Δ due to the increased probability of a run, as measured by the derivative of the panic-run threshold θ^* with respect to r_1.

The provision of bank liquidity insurance to depositors is captured by $r_1 > 1$. As in Diamond and Dybvig (1983) and subsequent papers, being risk averse and exposed to the risk of being impatient, depositors value the possibility of obtaining an amount of consumption higher than their original deposit at date 1, even if this implies a lower amount of consumption at date 2. Setting $r_1 = 1$ would rule out panics (i.e., $\theta^* = \underline{\theta}$). This implies the the utility loss of a run, as captured by Δ, becomes zero. However, the marginal benefit of risk-sharing remains positive, so this cannot be an equilibrium.

In choosing the deposit d, a consumer again trades off marginal cost and marginal benefit. The former comes from less consumption at time 0, as captured by the left-hand side of (27). The latter comes from more consumption at date 1 and 2, as captured by the right-hand side of (27).

We can substitute (28) and (26) into (27) and obtain an expression summarizing the decentralized equilibrium:

$$u'(1-D) = \int_0^{\theta^*(r_1,D)} u'(D)d\theta + \int_{\theta^*(r_1,D)}^1 u'(r_1 D)d\theta - (1-\lambda r_1)\frac{\Delta}{\lambda D}\frac{\partial \theta^*(r_1,D)}{\partial r_1}. \tag{29}$$

As in Section 3.3, this equation resembles an Euler equation and we use it to compare the decentralized economy with the constrained efficient allocation.

4.3 Constrained efficient allocation

In order to study the efficiency of the decentralized equilibrium, we characterize the constrained-efficient benchmark. As in the previous section, we consider a social planner who can only offer demand-deposit contracts like the banks. Hence, the planner is subject to panic runs in the same way as banks, and takes as given depositors' withdrawal strategies, as characterized by the run threshold θ^* in (23), evaluated at $d_i = d_{-i} = D$.

At date 0, the planner allocates $C_0 = 1 - D$ resources to consumption, and uses all deposits to finance investment. Since, as in the decentralized economy, the investment technology yields a unitary return at date 1, all consumers receive $C_1^{\text{run}} = D$ if there is a run at date 1. If there is no run, early consumers receive $C_1 = r_1 D$, while late consumers obtain C_2 that clears the planner's resource constraint:

$$\lambda C_1 + (1-\lambda)\frac{C_2}{R} = 1 - C_0. \tag{30}$$

The planner chooses r_1 and D to maximize the economy's expected aggregate welfare:

$$u(C_0) + \int_0^{\theta^*(r_1,D)} u(C_1^{\text{run}})d\theta + \int_{\theta^*(r_1,D)}^1 \left[\lambda u(C_1) + (1-\lambda)\theta u\left(C_2\right)\right] d\theta. \tag{31}$$

The following lemma characterizes the constrained efficient allocation.

Lemma 3. *The constrained-efficient equilibrium with panic runs is given by $r_1 > 1$ and $D > 0$ that solve:*

$$\int_{\theta^*(r_1,D)}^1 \left[u'(r_1 D) - \theta R u'\left(R\frac{1-\lambda r_1}{1-\lambda}D\right)\right] d\theta - \frac{\partial \theta^*(r_1,D)}{\partial r_1}\frac{\Delta}{\lambda D} = 0, \tag{32}$$

$$u'(1-D) = \int_0^{\theta^*(r_1,D)} u'(D)d\theta +$$
$$+ \int_{\theta^*(r_1,D)}^1 \left[\lambda r_1 u'(r_1 D) + \theta R(1-\lambda r_1)u'\left(R\frac{1-\lambda r_1}{1-\lambda}D\right)\right] d\theta - \frac{\partial \theta^*(r_1,D)}{\partial D}\Delta, \tag{33}$$

where $\Delta = \lambda u(r_1 D) + (1-\lambda)\theta^* u\left(R\frac{1-\lambda r_1}{1-\lambda}D\right) - u(D)$*, and* $\theta^*(r_1, D)$ *comes from* (23)*.*

The planner chooses the optimal level of liquidity insurance r_1 in the same way as banks in the decentralized economy. In doing so, it leaves the economy exposed to panic-driven runs, i.e. $r_1 > 1$, as this entails first-order benefits in terms of liquidity insurance. Regarding the savings choice, the planner trades off its marginal cost, in terms of lower date-0 consumption, with its marginal benefit, in terms of higher date-1 and date-2 consumption. However, unlike individual consumers in the decentralized equilibrium, the planner takes into account the effect of the level of deposits on the probability of a run. This is captured by the last term on the right-hand side of (33). In other words, relative to the decentralized economy, there is no saving externality in the sense that a planner fully internalizes the effect that the amount deposited has on the likelihood of panic runs. To ease the comparison with the decentralized economy, it is useful to substitute (32) into (33) and obtain:

$$u'(1-D) = \int_0^{\theta^*(r_1,D)} u'(D)d\theta +$$
$$+ \int_{\theta^*(r_1,D)}^1 u'(r_1 D)d\theta - (1-\lambda r_1)\frac{\Delta}{\lambda D}\frac{\partial \theta^*(r_1,D)}{\partial r_1} - \frac{\partial \theta^*(r_1,D)}{\partial D}\Delta. \tag{34}$$

The following proposition compares the social planner allocation with the decentralized equilibrium. This boils down to the comparison between (34) and (29), as the other equations that pin down the allocation are the same under the social planner as in the decentralized economy.

Proposition 6. *The decentralized equilibrium with panic runs is not constrained efficient. It exhibits under-saving, excessive financial instability and an inefficient level of bank liquidity insurance.*

By internalizing the effects of savings on the likelihood of panic runs, the social planner chooses a higher level of savings than in the decentralized equilibrium. Hence, in the decentralized equilibrium, there are too few deposits and runs are therefore too frequent. This result hinges directly on Proposition 5, which highlights that θ^* is decreasing in the level of deposits. In other words, the excessive fragility of the decentralized economy is not driven by the bank's distorted incentives, but rather relies on the saving externality:

The individual depositor fails to internalize the effect that her saving decision has on her own and other depositors' withdrawal decisions.

Interestingly, one implication of the comparison between the constrained efficient allocation and the decentralized economy is that the level of bank liquidity insurance, as measured by $r_1 > 1$, is also inefficient. As mentioned above, this is at odds with the results in Goldstein and Pauzner (2005), and is due to the fact that banks intermediate an inefficient amount of deposits. For a given aggregate level of deposits, r_1 is the same in the decentralized economy and in the constrained efficient one, since (26) and (32) are identical. Thus, if depositors saved the constrained efficient amount, banks would provide the constrained efficient level of liquidity insurance.

The result of Proposition 6 is in contrast with that of Proposition 3. In the economy with fundamental runs, the decentralized allocation features over-saving, while it exhibits under-saving in the economy with panic runs. This difference depends on the different nature of runs in the two frameworks and the resulting different sign of the saving externality. As shown in Propositions 4 and 1, when panic runs are possible, depositors attach a positive probability to the event that their date-2 consumption falls to zero, while with fundamental runs, date-2 consumption always stays positive and larger than date-1 consumption.

5 Optimal policy

The previous sections have shown that the decentralized equilibrium features a saving externality both with fundamental and panic runs. The resulting inefficiency creates a motive for public intervention. The aim of this section is to show how the constrained-efficient allocation can be implemented in the decentralized economy. To this end, we introduce a policy-maker who can impose proportional taxes on deposit holdings τ. The government collects taxes and rebates revenues to consumers as a lump-sum transfer T to clear its budget constraint:

$$T = \tau D. \tag{35}$$

The consumer's date-0 budget constraint reads:

$$c_{0i} + (1+\tau)\, d_i = 1 + T. \tag{36}$$

With the exception of the above budget constraints, the economy is the same as described in Sections 3 and 4. Denote a general run threshold as $\tilde{\theta}(r_1, d)$, with $\tilde{\theta}(r_1, d) = \underline{\theta}(r_1, d)$ in the economy with fundamental runs and $\tilde{\theta}(r_1, d) = \theta^*(r_1, d)$ in the economy with panic runs. The following lemma characterizes the equilibrium conditions of the economy with taxes.

Lemma 4. *Given a tax on deposit holdings τ, the decentralized equilibrium is characterized by:*

$$\int_{\tilde{\theta}(r_1,d)}^{1}\left[u'(r_1 d)-\theta R u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta-\frac{\partial\tilde{\theta}(r_1,d)}{\partial r_1}\frac{\Delta}{\lambda d}=0, \tag{37}$$

$$\begin{aligned}(1+\tau)u'(1-d)=&\int_{0}^{\tilde{\theta}(r_1,d)}u'(d)\,d\theta+\\&+\int_{\tilde{\theta}(r_1,d)}^{1}\left[\lambda r_1 u'(r_1 d)+\theta R(1-\lambda r_1)u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta,\end{aligned} \tag{38}$$

$$d_i=d_{-i}=d=D=I, \tag{39}$$

where $\Delta=\lambda u(r_1 d)+(1-\lambda)\tilde{\theta}(r_1,d)u\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)-u(d)$.

The tax policy creates a wedge in the intertemporal consumption-savings decision, thereby discouraging or encouraging savings. This can be seen by comparing (38) with (11) and (27). Optimal taxation is characterized in the following proposition.

Proposition 7. *The tax on deposit holdings that decentralizes the constrained efficient allocation solves:*

$$\tau^{opt}=\frac{\Delta}{u'(1-I)}\frac{\partial\tilde{\theta}(r_1,D)}{\partial D}. \tag{40}$$

It is positive in the economy with fundamental runs and negative in the economy with panic runs.

The optimal wedge is increasing in the marginal effect of deposits on the run probability $\frac{\partial\tilde{\theta}(r_1,D)}{\partial D}$ and the cost of bank runs Δ. The former indicates the strength of the saving externality and the latter the benefit of reducing the probability of bank runs. The optimal wedge is also decreasing in the marginal utility of date-0 consumption. This reflects a wealth effect: The cost of reducing bank intermediation is larger in a poorer economy. Hence, a benevolent policy-maker should intervene less.

As shown in Propositions 3 and 6, the sign of the saving externality is different in the economy with fundamental runs and in the economy with panic runs. This has interesting implications for the optimal policy: While fundamental runs imply a positive optimal wedge, in an economy with panic runs the optimal wedge is negative. Hence, in an economy with fundamental runs a benevolent policy-maker should tax deposits. On the contrary, in an economy with panic runs deposits should be subsidized.

6 A numerical illustration

In this section, we illustrate the properties of the model using a numerical example. In particular, we study how the severity of the inefficiency stemming from the saving externality and the optimal policy vary with R, i.e. the investment return in case of success.

We assume the following functional form for the depositors' utility function:

$$u(c) = \begin{cases} c & \text{if } c \leq \bar{c}, \\ \frac{c^{1-\sigma}-F}{1-\sigma} & \text{otherwise,} \end{cases} \tag{41}$$

where $\bar{c}$ is a small positive constant. In this way, $u(0) = 0$ and the utility function exhibits constant relative risk aversion σ for $c > \bar{c}$. We set $\sigma = 2$ and the scale parameter F to 2.8. The threshold of the upper dominance region is set to $\bar{\theta} = 1$ and the probability of being an early consumer λ to 0.02 as in Mattana and Panetti (2020). We provide results for values of R ranging between 2.02 and 2.10, so that the expected net return on the risky investment $\mathbb{E}[\theta]R$ lies between 1 and 5 per cent. Table 1a and 1b provide the characterization of the decentralized equilibrium of the economy with fundamental runs and panic runs, as depicted in Sections 3 and 4, as well as the comparison with the relevant constrained efficient allocation.

In line with Propositions 2 and 5, the (gross) deposit rate r_1 is larger than 1 in both economies, as it captures the provision of liquidity insurance to the depositors. There exists a positive relation between R and r_1 (column 2), and a negative one between R and d (column 3). The per-unit return on the productive asset R affects the intertemporal allocation of resources in the decentralized equilibrium in a non-trivial way. At date 0, a higher R triggers both an income and a substitution effect. On the one hand, through the substitution effect, higher R induces consumers to deposit more in the bank and consume less. On the other hand, through the income effect, a higher R leads to an increase in date 0 consumption. At date 1, similar forces also affect the allocation of resources and, in turn, consumption between date 1 and date 2 via a change in r_1. In our numerical illustration the income effect dominates the substitution effect both at date 0 and date 1, for any value of R. Thus, higher R leads to higher r_1 and lower d.

The relation between R and both the fundamental-run and the panic-run thresholds $\underline{\theta}$ and θ^* (column 4) is negative, since the investment return in case of success R increases late consumption, and thus lowers the incentives to withdraw early, as shown in equations (7) and (23). Interestingly, comparing Table 1a and 1b, the deposit rate is more than five times larger in the economy with fundamental runs than in the economy with panic runs. A consequence of this is that the run threshold in the economy with fundamental runs is higher than in the economy with panic runs, despite the former economy not suffering from coordination failures as the latter.

Column 5 of Table 1a highlights the existence of a positive saving externality since, as shown in Proposition 1, the fundamental-run threshold is increasing in the equilibrium deposit d. This implies that the decentralized equilibrium with fundamental runs exhibits

Table 1: The decentralized equilibrium and the comparison with the constrained efficient allocation for different values of $\mathbb{E}(R)$.

(a) Economy with fundamental runs

$\mathbb{E}(R)$	r_1(net, bps)	d(%)	$\underline{\theta}$	$\partial\underline{\theta}/\partial d$	Δd(bps)	$\Delta\underline{\theta}$(bps)	Δr_1(bps)	τ(bps)
1.01	66.230	42.855	0.333	5.747	175.699	625.156	-0.5022	288.120
1.02	67.352	42.700	0.328	5.840	190.610	699.494	-0.5485	300.432
1.03	68.462	42.544	0.322	5.936	207.473	786.512	-0.6011	313.168
1.04	69.560	42.386	0.316	6.033	226.765	889.829	-0.6616	326.355
1.05	70.647	42.226	0.310	6.132	249.178	1014.701	-0.7322	340.028

(b) Economy with panic runs

$\mathbb{E}(R)$	r_1(net, bps)	d(%)	θ^*	$\partial\theta^*/\partial d$	Δd(bps)	$\Delta\theta^*$(bps)	Δr_1(bps)	τ(bps)
1.01	12.682	33.660	0.251	-0.114	-1.063	0.067	-0.005	-2.963
1.02	12.806	33.556	0.249	-0.095	-0.889	0.041	-0.004	-2.475
1.03	12.929	33.453	0.246	-0.077	-0.718	0.022	-0.004	-1.996
1.04	13.052	33.351	0.244	-0.059	-0.549	0.008	-0.004	-1.525
1.05	13.173	33.249	0.241	-0.041	-0.383	0.001	-0.002	-1.063

over-saving (column 6), excessive fragility (column 7) and low liquidity insurance (column 8). As the distortion of the decentralized equilibrium is increasing in R, an increasing tax on deposits of between 2.9 and 3.4 per cent is needed to correct the inefficiency (column 9).

Table 1b also reports the comparison between the decentralized equilibrium with panic runs and the constrained efficient allocation. For a given value of R, the derivative of θ^* with respect to d is negative (column 5), as proved in Proposition 5. This confirms that the decentralized equilibrium with panic runs exhibits under-saving (column 6) and excessive financial fragility (column 7) with respect to the constrained efficient allocation. Interestingly, compared to the decentralized equilibrium, a social planner commanding higher savings not only brings about lower financial fragility, but is also able to provide higher liquidity insurance than the banks (column 8). Moreover, the distortion of the decentralized equilibrium is decreasing in R. Therefore, the implementation of the constrained efficient allocation in the decentralized economy is ensured by a subsidy to deposits (column 9) that is also decreasing in R.

7 Conclusions

In this paper, we develop a banking model with endogenous depositor runs and consumption-saving decisions. Our contribution is twofold. First, we find that the probability of runs is affected by the level of deposits in the economy. Second, we show that individual depositors do not internalize the effect on fragility when choosing how much to deposit into a bank. The resulting saving externality represents a novelty in the bank-run literature and has important implications for the efficiency of the competitive equilibrium.

The inefficiency associated with the saving externality represents a rationale for public intervention. Policy-makers should induce individual depositors to internalize the effect of their consumption-saving decision on financial stability. The design of the optimal policy depends on the nature of bank runs, namely on whether banks are subject to fundamental- or panic-driven runs. In particular, the former leads to over-saving, which can be corrected with a tax on deposits. The latter leads to under-saving, which requires a subsidy on deposits.

Our results show that intermediaries that are not facing the risk of panic runs tend to grow excessively large, provide an inefficient level of liquidity and are too fragile. This further suggests that policies like deposit insurance and/or liquidity support by central banks should be complemented by other interventions meant to reduce the incentives of depositors to over-save. In this respect, our paper highlights an additional potential drawback associated with bank guarantees. Besides the well-known moral hazard problems on the side of the bank, deposit insurance and emergency liquidity provision by central banks may also distort savers' incentives, and translate into an excessively large and fragile financial sector.

References

Ahnert, T., K. Anand, P. Gai, and J. Chapman (2019). Asset encumbrance, bank funding, and fragility. *The Review of Financial Studies 32*(6), 2422–2455.

Allen, F., E. Carletti, and D. Gale (2014). Money, financial stability and efficiency. *Journal of Economic Theory 149*, 100–127.

Allen, F., E. Carletti, I. Goldstein, and A. Leonello (2018). Government guarantees and financial stability. *Journal of Economic Theory 177*, 518–557.

Allen, F. and D. Gale (2004, July). Financial Intermediaries and Markets. *Econometrica 72*(4), 1023–1061.

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

Bolton, P., T. Santos, and J. A. Scheinkman (2016, March). Savings Gluts and Financial Fragility. mimeo.

Brunnermeier, M. K. and M. Oehmke (2013). The maturity rat race. *The Journal of Finance 68*(2), 483–521.

Caballero, R. J. and A. Krishnamurthy (2001). International and domestic collateral constraints in a model of emerging market crises. *Journal of monetary Economics 48*(3), 513–548.

Caballero, R. J. and A. Krishnamurthy (2009). Global imbalances and financial fragility. *American Economic Review 99*(2), 584–88.

Caballero, R. J. and A. Simsek (2019, June). Prudential Monetary Policy. CEPR Discussion Paper No. 13832.

Carlsson, H. and E. van Damme (1993, July). Global Games and Equilibrium Selection. *Econometrica 61*(4), 989–1018.

Choi, D. B. (2014). Heterogeneity and stability: Bolster the strong, not the weak. *The Review of Financial Studies 27*(6), 1830–1867.

Dang, T. V., G. B. Gorton, B. Holmstrom, and G. Ordonez (2017, April). Banks as Secret Keepers. *American Economic Review 107*(4), 1005–1029.

Davila, E. and A. Korinek (2018, January). Pecuniary Externalities in Economies with Financial Frictions. *Review of Economic Studies 85*(1), 352–395.

Davila, J., J. H. Hong, P. Krusell, and J.-V. Rios-Rull (2012, November). Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks. *Econometrica 80*(6), 2431–2467.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political economy 91*(3), 401–419.

Eisenbach, T. M. (2017). Rollover risk as market discipline: A two-sided inefficiency. *Journal of Financial Economics 126*(2), 252–269.

Ennis, H. M. and T. Keister (2009). Bank runs and institutions: The perils of intervention. *American Economic Review 99*(4), 1588–1607.

Farhi, E. and I. Werning (2016, September). A Theory of Macroprudential Policies in the Presence of Nominal Rigidities. *Econometrica 84*(5), 1645–1704.

Geanakoplos, J. and H. Polemarchakis (1985, August). Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete. Cowles Foundation Discussion Paper No. 764.

Gertler, M., N. Kiyotaki, and A. Prestipino (2020). Banking panics as endogenous disasters and the welfare gains from macroprudential policy. *American Economic Review Forthcoming.*

Goldstein, I. and A. Pauzner (2005). Demand–deposit contracts and the probability of bank runs. *Journal of Finance 60*(3), 1293–1327.

Gorton, G. B., S. Lewellen, and A. Metrick (2012). The Safe-Asset Share. *American Economic Review AEA Papers and Proceedings 102*(3), 101–106.

Hanson, S., A. Shleifer, J. C. Stein, and R. W. Vishny (2015, September). Banks as patient fixed-income investors. *Journal of Financial Economics 117*(3), 449–469.

Hart, O. (1975). On the Optimality of Equilibrium When the Market Structure Is Incomplete. *Journal of Economic Theory 11*(3), 418–443.

Kashyap, A. K. and J. C. Stein (2012, January). The Optimal Conduct of Monetary Policy with Interest on Reserves. *American Economic Journal: Macroeconomics 4*(1), 266–282.

Kindleberger, C. P. and R. Z. Aliber (1978). *Manias, Panics, and Crashes: A history of financial crises* (5th ed.). John Wiley & Sons.

Krishnamurthy, A. (2010, March). How debt markets have malfunctioned in the crisis. *Journal of Economic Perspectives 24*(1), 3–28.

Levine, R., C. Lin, M. Tai, and W. Xie (2021). How Did Depositors Respond to COVID-19? *Review of Financial Studies forthcoming.*

Li, L., P. E. Strahan, and S. Zhang (2020). Banks as lenders of first resort: Evidence from the covid-19 crisis. *The Review of Corporate Finance Studies 9*(3), 472–500.

Lorenzoni, G. (2008, July). Inefficient Credit Booms. *Review of Economic Studies 75*(3), 809–833.

Martinez-Miera, D. and R. Repullo (2017, March). Search for Yield. *Econometrica 85*(2), 351–378.

Mattana, E. and E. Panetti (2020, March-April). The Welfare Costs of Self-Fulfilling Bank Runs. *Journal of Money, Credit and Banking 53*(2-3), 401–440.

Morris, S. and H. S. Shin (2011, August). Global Games: Theories and Applications. Cowles Foundation Discussion Paper No. 1275R.

Peck, J. and A. Setayesh (2019). A diamond-dybvig model in which the level of deposits is endogenous.

Stiglitz, J. (1982). The Inefficiency of the Stock Market Equilibrium. *Review of Economic Studies 49*(2), 241–261.

Van Den Heuvel, S. (2008). The welfare cost of bank capital requirements. *Journal of Monetary Economics 55*, 298–320.

Vives, X. (2014). Strategic complementarity, fragility, and regulation. *The Review of Financial Studies 27*(12), 3547–3592.

A Proofs

Proof of Proposition 1. Denote as $v\left(\theta, \underline{\theta}(r_1, d_{-i})\right)$ the net benefit of waiting until period 2 as a function of the economy's fundamental θ and of the fraction of depositors $-i$ who withdraw at date 1:

$$v\left(\theta, \underline{\theta}(r_1, d_{-i})\right) = \begin{cases} \theta u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right) - u(r_1 d_i) & \text{if } \theta > \underline{\theta}(r_1, d_{-i}), \\ 0 - u\left(d_i\right) & \text{if } \theta \leq \underline{\theta}(r_1, d_{-i}). \end{cases} \tag{42}$$

If $\theta > \underline{\theta}(r_1, d_{-i})$, all $-i$ depositors do not run and depositor i expects to receive the promised repayment at either dates. If $\theta \leq \underline{\theta}(r_1, d_{-i})$, all $-i$ depositors run, the bank is forced to liquidate its investment at date 1, it becomes insolvent and so depositor i expects to receive nothing if she withdraw at date 2. If she withdraws at date 1, instead, depositor i receives back her deposit d_i. The threshold in the proposition follows directly from the function $v(\theta, \underline{\theta}(r_1, d_{-i})$. When $\theta \leq \underline{\theta}(r_1, d_{-i})$, depositor i is better off withdrawing and the proposition follows. When $\theta > \underline{\theta}(r_1, d_{-i})$, it is optimal for depositor i to withdraw as long as

$$\theta u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right) < u(r_1 d_i), \tag{43}$$

which is the case for any $\theta < \underline{\theta}(r_1, d_i) = \frac{u(r_1 d_i)}{u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)}$.

For the second part of the proof, taking the derivative of $\underline{\theta}(d_i)$ with respect to d_i, we obtain:

$$\frac{\partial \underline{\theta}(d_i)}{\partial d_i} = \frac{1}{u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)} \left[u'(r_1 d_i) r_1 - \underline{\theta}(d_i) u'\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right) R\frac{1-\lambda r_1}{1-\lambda} \right], \tag{44}$$

Multiply and divide by d_i and collect $u(r_1 d_i)$ to obtain:

$$\frac{\partial \underline{\theta}(d_i)}{\partial d_i} = \frac{u(r_1 d_i)}{u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right) d_i} \left[\frac{u'(r_1 d_i) r_1 d_i}{u(r_1 d_i)} - \frac{u'\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right) R\frac{1-\lambda r_1}{1-\lambda} d_i}{u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)} \right]. \tag{45}$$

The expression in the square brackets is positive if $u'(c)c/u(c)$ (i.e. the semi-elasticity of consumption) is decreasing in c, that is:

$$\frac{[u''(c)c + u'(c)]u(c) - [u'(c)]^2 c}{[u(c)]^2} < 0. \tag{46}$$

A sufficient condition for this to be true is that the coefficient of relative risk aversion $RRA = -u''(c)c/u'(c)$ is larger than 1, as assumed. Hence, the proposition follows. □

Proof of Proposition 2. Taking the derivative of (8) with respect to r_1 and substituting $d_i = d_{-i} = d$ we obtain expression (10) as in the proposition. The condition that pins down the equilibrium amount of deposits is obtained similarly by differentiating (9) with respect to d_i and evaluating it at $d_i = d_{-i} = d$. Thus, we obtain:

$$u'(1-d) = \int_0^{\underline{\theta}(r_1,d)} u'(d)\,\mathrm{d}\theta + \int_{\underline{\theta}(r_1,d)}^1 \left[\lambda r_1 u'(r_1 d) + \theta R(1-\lambda r_1) u'\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] \mathrm{d}\theta - \frac{\partial \underline{\theta}(r_1, d_i)}{\partial d_i}\Delta, \tag{47}$$

where $\Delta = (1-\lambda)\underline{\theta}(r_1,d)\left[u\left(R\frac{1-\lambda r_1}{1-\lambda}d\right) - u(r_1 d)\right] = 0$ because of the definition of $\underline{\theta}(r_1, d_i)$. Hence, we obtain the same expression as in (9). Finally, we prove that $r_1 > 1$ by contradiction. Assume that $r_1 = 1$. When this is the case, the right-hand side of (10) is zero as $\Delta = 0$, while the left-hand side of (10) is positive as $u'(d) > Ru'(Rd)$. Hence, since r_1 is an interior solution, it follows that $r_1 > 1$ holds in equilibrium and this completes the proof. □

Proof of Lemma 1. The two conditions in the lemma are obtained by simply differentiating (14) with respect to both r_1 and D. The proof of $r_1 > 1$ is analogous to the one in Proposition 2. Hence, the lemma follows. □

Proof of Proposition 3. For given $d = D$, the deposit rate chosen by banks is the same as the one chosen by the planner, as it can be easily seen by comparing (10) with (15). Hence, the comparison between the decentralized and the constrained efficient allocation boils down to the comparison of (13) with (17). It is easy to see that the former is larger than the latter since $\frac{\partial \underline{\theta}(r_1,d)}{\partial d} > 0$ as shown in Proposition 1. Hence, it follows that the quantity of deposit D in the decentralized allocation is larger than the constrained efficient one and the proposition follows. □

Proof of Lemma 2. The proof is done by contradiction. Assume first that depositor i finds it optimal not run when the other depositors run, i.e., $x_i^* < x_{-i}^*$. Then, depositor i receives 0 in the range $\left(x_i^*, x_{-i}^*\right)$ at date 2, while she could get d_i if joining the run. Hence, $x_i^* < x_{-i}^*$ cannot hold. Assume now that depositor i finds it optimal to run when the others do not run, i.e., $x_i^* > x_{-i}^*$. Then, depositor i receives $u(r_1 d_i)$ in the range (x_{-i}^*, x_i^*) when she runs, while she expects to receive $u(r_2 d_i) = u\left(R\frac{1-\lambda r_1}{1-\lambda} d_i\right)$ at date 2. Yet, $u(r_1 d_i)/u(r_2 d_i) = \underline{\theta}(r_1, d_i)$ by definition, and $\underline{\theta}(r_1, d_i) < x_i^*$ by construction. Hence, $x_i^* > x_{-i}^*$ cannot be optimal and the lemma follows. □

Proof of Proposition 4. The proof follows closely the one in Goldstein and Pauzner (2005) since our model also exhibits one-sided strategic complementarities.

The arguments in the proof in Goldstein and Pauzner (2005) establish that there is a unique equilibrium in which depositors run if and only if the signal they receive is below a common signal x^*. The number n of depositors withdrawing at date 1 is equal to the probability of receiving a signal x_i below x^* and, given that depositors' signals are independent and uniformly distributed over the interval $[\theta-\varepsilon, \theta+\varepsilon]$, it is:

$$n(\theta, x^*) = \begin{cases} 1 & \text{if } \theta \leq x^* - \varepsilon \\ \lambda + (1-\lambda)\left(\frac{x^*-\theta+\varepsilon}{2\varepsilon}\right) & \text{if } x^*-\varepsilon \leq \theta \leq x^*+\varepsilon \\ \lambda & \text{if } \theta \geq x^*+\varepsilon \end{cases} \tag{48}$$

When θ is below $x^* - \varepsilon$, all patient depositors receive a signal below x^* and run. When θ is above $x^* + \varepsilon$, all $1-\lambda$ late depositors wait until date 2 and only the λ early consumers withdraw early. In the intermediate interval, when θ is between $x^* - \varepsilon$ and $x^* + \varepsilon$, there is a partial run as some of the late depositors run. The proportion of late depositors withdrawing early decreases linearly with θ as fewer agents observe a signal below the threshold.

Denote as $\Delta(x_i, n(\theta))$ a depositor's expected utility difference in utility between withdrawing at date 2 and date 1 when he holds beliefs $n(\theta)$ regarding the number of depositors running, which is given in (48) since for any realization of θ, the proportion of depositors running is deterministic. The function $\Delta(x_i, n(\theta))$ is equal to

$$\Delta(x_i, n(\theta)) = \frac{1}{2\epsilon}\int_{x_i-\epsilon}^{x_i+\epsilon} \mathcal{V}(\theta, n(\theta))d\theta, \tag{49}$$

where $\mathcal{V}(\theta, n(\theta))$ is given in (22) and $n(\theta) = n(\theta, x^*)$ as given in (48). The function $\Delta(x_i, n(\theta))$ is continuous in x_i and increases continuously in positive shifts in the signal x_i and proportion of depositors running $n(\theta)$. The proof of the properties of $\Delta(x_i, n(\theta))$, as well as the rest of the proof follows closely Goldstein and Pauzner (2005), thus we omit it for brevity.

Having characterized the proportion of agents withdrawing for any possible value of the fundamentals θ, we can now compute the threshold signal x^*_{-i}. A patient depositor $-i$ who receives the signal x^*_{-i} must be indifferent between withdrawing at date 1 and at date 2. The threshold x^*_{-i} can be then found by equalizing the following expression to zero:

$$f(\theta, r_1, d_{-i}) = \int_{\lambda}^{\frac{1}{r_1}} \left[\theta u\left(R\frac{1-nr_1}{1-n}d_{-i}\right) - u(r_1 d_{-i})\right] dn + \int_{\frac{1}{r_1}}^{1} \left[u(0) - u\left(\frac{d_{-i}}{n}\right)\right] dn, \tag{50}$$

where $\theta(n) = x^*_{-i} + \varepsilon - 2\varepsilon\frac{(n-\lambda)}{1-\lambda}$ from (48). Equation (50) follows from (22) and requires that a late depositor's expected utility when he or she withdraws at date 1 is equal to that when he or she waits until date 2. Note that in the limit, when $\varepsilon \to 0$, $\theta(n) \to x^*_{-i}$, and we denote it as $\theta^*(r_1, d_{-i})$.

To prove that $\theta^*(r_1, d_{-i})$ is increasing in r_1 and decreasing d_{-i}, we use the implicit function theorem on (50) and obtain:

$$\frac{\partial\theta^*(r_1,d_{-i})}{\partial r_1}=-\frac{\frac{\partial f(\cdot)}{\partial r_1}}{\frac{\partial f(\cdot)}{\partial\theta^*}}\qquad\text{and}\qquad\frac{\partial\theta^*(r_1,d_{-i})}{\partial d_{-i}}=-\frac{\frac{\partial f(\cdot)}{\partial d_{-i}}}{\frac{\partial f(\cdot)}{\partial\theta^*}}.\tag{51}$$

It is easy to see that $\partial f(\cdot)/\partial\theta>0$. Thus, the sign of $\partial\theta^*(r_1,d_{-i})/\partial r_1$ and $\partial\theta^*(r_1,d_{-i})/\partial d_{-i}$ are given by the opposite sign of $\partial f(\cdot)/\partial r_1$ and $\partial f(\cdot)/\partial d_{-i}$, respectively. The former is given by:

$$\frac{\partial f(\cdot)}{\partial r_1}=-d_{-i}\int_{\lambda}^{\frac{1}{r_1}}\left[u'(r_1d_{-i})+\theta^*\frac{nR}{1-n}u'\left(R\frac{1-nr_1}{1-n}d_{-i}\right)\right]dn<0.\tag{52}$$

The latter is equal to:

$$\frac{\partial f(\cdot)}{\partial d_{-i}}=\int_{\lambda}^{\overline{n}}\left[\theta^*u'\left(Rd_{-i}\frac{1-nr_1}{1-n}\right)R\frac{1-nr_1}{1-n}-u'(r_1d_{-i})r_1\right]dn-\int_{\overline{n}}^{1}u'\left(\frac{d_{-i}}{n}\right)\frac{1}{n}dn,\tag{53}$$

where $\overline{n}=1/r_1$. Multiply and divide everything by d_{-i} to obtain:

$$\begin{aligned}\frac{\partial f(\cdot)}{\partial d_{-i}}=&\frac{1}{d_{-i}}\left[\int_{\lambda}^{\overline{n}}\left[\theta^*u'\left(R\frac{1-nr_1}{1-n}d_{-i}\right)R\frac{1-nr_1}{1-n}d_{-i}-u'(r_1d_{-i})r_1d_{-i}\right]dn+\right.\\&\left.-\int_{\overline{n}}^{1}u'\left(\frac{d_{-i}}{n}\right)\frac{d_{-i}}{n}dn\right],\end{aligned}\tag{54}$$

and denote $c_1=r_1d_{-i}$ and $c_2(n)=R\frac{1-nr_1}{1-n}d_{-i}$. The expression above can be rewritten as:

$$\frac{\partial f(\cdot)}{\partial d_{-i}}=\frac{1}{d_{-i}}\left[\int_{\lambda}^{\overline{n}}\left[\theta^*u'\left(c_2(n)\right)c_2(n)-u'(c_1)c_1\right]dn-\int_{\overline{n}}^{1}u'\left(\frac{d_{-i}}{n}\right)\frac{d_{-i}}{n}dn\right],\tag{55}$$

The sign of the derivative depends on that of the components inside the square brackets. The second term is negative so we turn to study the sign of the first one.

Since $u'(c)c$ is decreasing in c and $c_2(n)$ is decreasing in n, $u'\left(c_2(n)\right)c_2(n)-u'(c_1)c_1$ is increasing in n. Furthermore, when $n=\overline{n}$, $c_2(\overline{n})=0$. If $\lim_{c\to 0^+}u'(c)c$ is sufficiently large, the first integral in (55) is positive and dominates the other. As a result, $\frac{\partial f(.)}{\partial d_{-i}}>0$ and $\frac{\partial\theta^*(r_1,d_{-i})}{\partial d_{-i}}<0$. Hence, the proposition follows. □

Proof of Proposition 5. Differentiating the bank's objective function in (24) with respect to r_1, we obtain (26). Similarly, differentiating (25) with respect to d yields (27).

To prove that $r_1>1$, evaluate (26) at $r_1=1$ using $d_i=d_{-i}=d=D=I$. This leads to:

$$\lambda\int_{\underline{\theta}}^{1}\left[u'(d)d-\theta Rdu'(Rd)\right],\tag{56}$$

since $\theta^*\to\underline{\theta}$ when $r_1=1$, and $\Delta=0$ by definition of $\underline{\theta}$ in (20). This expression is positive because relative risk aversion is larger than 1 for $c>\overline{c}$ and $\overline{c}<I$. To see that, notice

that $u'(d)d - \theta R d u'(Rd) > u'(d)d - R d u'(Rd)$ and $u'(c)c$ is decreasing in c. This follows directly from $-u''(c)c/u'(c) > 1$. Thus, the proposition follows. □

Proof of Lemma 3. The two conditions in the lemma are obtained by simply differentiating (31) with respect to r_1 and D. The proof of $r_1 > 1$ is analogous to that of Proposition 2. □

Proof of Proposition 6. The proof follows directly from the comparison of (29) and (34). When evaluating (29) at the optimal level of investment solving (34), (29) is positive since the two first-order conditions only differs for the term $\frac{\partial \theta^*}{\partial D}\Delta$, which is negative. This implies that in the decentralized allocation the level of aggregate deposits D is lower than that chosen by the planner. The results about the excessively high level of financial fragility follows directly from the fact that the planner implements a higher level of aggregate deposits than in the decentralized economy in order to limit runs given that $\frac{\partial \theta^*}{\partial D}\Delta < 0$. Finally, the inefficient level of liquidity insurance provided by banks to consumers emerges as the result of the fact that both the banks and the planner takes r_1 as the solution to (26). However, the level of deposits d is not the same in the decentralized allocation and in the planner's one, which determines a difference between the r_1 set by banks in the decentralized allocation and that set by the planner. Thus, the proposition follows. □

Proof of Lemma 4. The derivation follows the steps of the proof of Proposition 5. The tax only affects the consumer's problem. For a general run threshold $\tilde{\theta}$, the problem becomes:

$$\max_{d}\ u\left[1-(1+\tau)d+T\right]+\int_{0}^{\tilde{\theta}(r_1,d)} u\left(d\right)d\theta+\int_{\tilde{\theta}(r_1,d)}^{1}\left[\lambda u(r_1 d)+(1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta. \tag{57}$$

Given that consumers behave symmetrically, we can write the associated first-order condition as

$$\begin{aligned}(1+\tau)\,u'\left(1-d\right) = &\int_{0}^{\tilde{\theta}(r_1,d)} u'\left(d\right)d\theta+\\ &+\int_{\tilde{\theta}(r_1,d)}^{1}\left[\lambda r_1 u'(r_1 d)+\theta R(1-\lambda r_1)u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta.\end{aligned} \tag{58}$$

Hence, the lemma follows. □

Proof of Proposition 7. Constrained efficiency in the case with fundamental and with panic runs is determined by Lemmas 1 and 3, respectively. By simple substitution, we find that the expression in the lemma makes the decentralized equilibrium identical to the constrained efficient one. The sign of the optimal tax is determined by the sensitivity of the threshold with respect to the quantity deposited. This property is verified in Propositions 1 and 4. Hence, the proposition follows. □