Savings, Efficiency and Bank Runs*

Agnese Leonello[†] Caterina Mendicino[‡] Ettore Panetti[§]

Davide Porcellacchia[¶]

Abstract

Does the level of deposits matter for bank fragility and efficiency? By augmenting a standard model of endogenous bank runs with a consumption-saving decision, we obtain two results. First, depositors' incentives to run are a function of savings held as bank deposits. Second, a saving externality emerges since individual depositors do not internalize the effect of their savings on the bank-run probability. Therefore, the economy features inefficient savings and bank liquidity provision, as well as excessive bank fragility. Finally, we characterize the optimal policy to implement the efficient allocation.

JEL codes: G01, G21, G28

Keywords: endogenous bank runs, liquidity provision, financial crises, saving externality.

*We are grateful to E. Carletti, T. Eisenbach, M. Gertler, W. Den Haan, A. Kashyap, T. Keister, N. Kiyotaki, A. Krishnamurthy, G. Ordoñez, M. Ravn, R. Repullo, J.C. Rochet, I. Schnabel and J. Suarez for useful comments, and to participants at 2020 EEA Summer Meeting, 2020 CEBRA, Banco de Portugal, Sverige Riksbank, European Central Bank, University of Bristol and Rutgers University. A special thank to Luigi Falasconi for outstanding research assistance. 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.

[†]European Central Bank, DG-Research, Financial Research. Email: agnese.leonello@ecb.europa.eu.

[‡]European Central Bank, DG-Monetary Policy, Monetary Policy Analysis and UECE. Email: caterina.mendicino1@ecb.europa.eu.

[§]University of Naples Federico II, CSEF, SUERF and UECE. Email: ettore.panetti@unina.it.

[¶]European Central Bank, DG-Research, Monetary Policy Research. Email: davide.porcellacchia@ecb.europa.eu.

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. Banks' reliance on short-term debt as a source of funding has been considered a key source of fragility (e.g., Diamond and Dybvig, 1983; Allen and Gale, 2004; Krishnamurthy, 2010; Brunnermeier and Oehmke, 2013). 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).[1] Figure 1 illustrates the evolution of total bank deposits in the last years for the U.S. and the Euro Area. They 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. A series of bank failures in early 2023 – in particular, Silicon Valley Bank and Signature Bank in the U.S. – suggests a connection between large inflows of uninsured deposits and banks' resilience to panic-driven depositors' runs.[2] In light of this evidence, two questions naturally arise. Does the size of a bank's deposit base matter for its 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. Second, we characterize the existence of a *saving externality*: Individual investors fail to fully internalize the impact of their decision to save in bank deposits on the probability of a bank run. As a result, the allocation is constrained inefficient. The economy features excessive financial fragility, and inefficient liquidity provision and bank size.

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

[1] Deposit decisions have always been a key determinant of the size of financial intermediation. In fact, in the period 1896-2012, deposits represented on average around 80 per cent of U.S. banks' liabilities (Hanson et al., 2015). The literature has also recognized savers' demand for money-like assets as a key driver of financial crises and macroeconomic activity (Gorton et al., 2012; Dang et al., 2017).

[2] It is worth noting that to date uninsured deposits still represent about half of the total deposits in the largest commercial banks both in the U.S. (Egan et al., 2017) and in the Euro Area (source: ECB Bank Balance Sheet Items and European Banking Authority data).

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

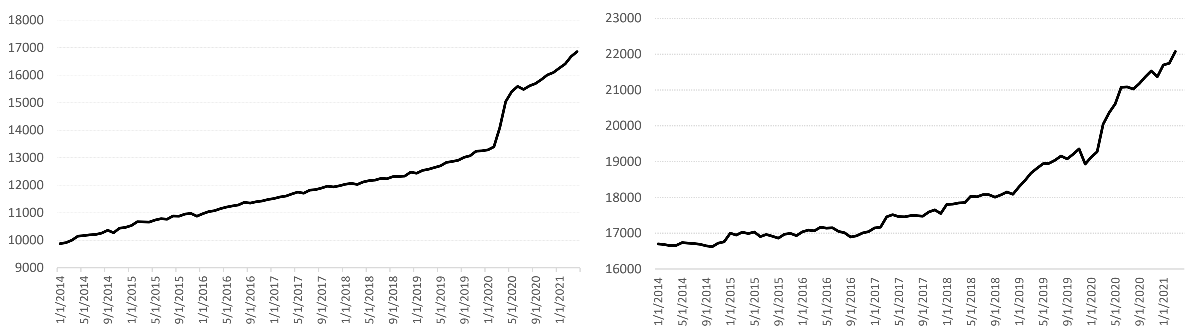

Figure 1: Total deposits in commercial banks, in billions of U.S. dollars in Panel (a) and billions of euros in Panel (b).

global-game methodology. 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 of runs, this is not true in our framework. The allocation is constrained inefficient because financial fragility is endogenous to the level of savings and consumers 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 banks 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 based on 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. Runs are the result of a coordination failure: Depositors run out of fear that others will do the same and there will not be enough resources left in the bank to repay those who wait. We refer to these events as panic-driven runs.

Our analysis provides novel insights into 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 deciding whether to run, depositors compare the expected utility from running with that from waiting. Since depositors are risk averse and consumption differs in the two dates, a marginal increase in the level of deposits is valued differently at dates 1 and 2. Hence, the level of deposits affects their incentives to run. In a reasonable parameter space, we further show that a marginal increase in the level of deposits is valued more at date 1 than at date 2. Hence, an increase in the level of deposits leads to an increase in depositors' incentives to run. This is in line with the empirical evidence of Iyer and Puri (2012) that shows a positive relation between depositors' account balances and their likelihood of running.

Second, the economy exhibits a saving externality. While a constrained social planner subject to panic-driven runs internalizes the effect of the level of deposits on the incentives to withdraw, individual depositors do not. In the decentralized economy, each depositor finds it optimal to follow the run behavior of all others, and therefore takes as given the level of the fundamental below which a bank run occurs. Due to the saving externality, in the decentralized economy individual saving decisions are socially costly. They lead to excessive financial fragility, as well as constrained inefficient liquidity provision and bank size.

Overall, our results bring about a novel motive for public intervention. Since the inefficiency of the decentralized economy is rooted in the depositors' consumption-saving decisions, fiscal policy (i.e. a deposit tax) could represent an effective tool to control the inefficient levels of savings in the economy, thereby contributing to the reduction of financial fragility. This represents a new rationale for coordination between prudential and fiscal policy.

Literature Review. Our paper contributes to three strands of the literature. 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; Ahnert et al., 2019). The novelty of our paper is to endogenize consumers' saving decisions and study their implications for fragility and efficiency. While in previous contributions financial intermediation is constrained efficient, the saving externality that arises from the depositors' saving decisions leads to an inefficiency that 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; Farhi et al., 2009; Stein, 2012; Allen et al., 2014).

Closely related to our paper is Peck and Setayesh (2022). In a Diamond-Dybvig framework, they study how deposit size affects the feasibility of the efficient allocation in the decentralized economy, and its fragility. In particular, they show that there exists a whole interval of deposit levels that can yield the efficient allocation in equilibrium. Among those equilibria, the ones with lower deposit levels are more fragile, even though deposits do not affect the probability of a bank run, which is sunspot-driven. In our framework instead, the probability of a bank run is fully endogenous. Moreover, it is also a function of deposits. More importantly, depositors' failure to recognize the effect of their saving choices on financial fragility brings about the saving externality, that distorts the constrained efficiency of the decentralized equilibrium. This further allows us to uniquely determine the optimal policy to implement the constrained efficient allocation.

The ability to endogenize the probability of a run relies on the use of global-game 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.

Our analysis also connects to the literature that studies the constrained efficiency of decentralized economies in the presence of externalities. Several papers build on 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 interaction of coordination failures and deposit decisions for the fragility and efficiency of market outcomes. Our work complements existing papers by filling this gap.

We share the focus on the role of consumption-saving decisions for the efficiency of decentralized equilibria with Davila et al. (2012). They show that in an economy with idiosyncratic risk and incomplete markets, 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 via banks is what brings about the saving externality.

Finally, our paper contributes to the literature on the saving glut and financial fragility. Excessive savings around the world arguably 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). Our framework instead, focuses on the liability side of banks' balance sheets and so links the inefficient level of savings to the occurrence of bank runs.

2 A model with panic runs

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). Consumers learn their own realization of the shock privately. The law of large numbers holds, so λ and $1 - \lambda$ are also the fractions of consumers who turn out to be early and late, respectively. Early consumers only want to consume at date 1, while late consumers are indifferent between consuming at date 1 or 2. 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 is continuous and satisfies $u'(c) > 0$, $u''(c) < 0$, and $u(0) = 0$. The coefficient of relative risk aversion $-cu''(c)/u'(c)$ is greater than 1 for any $c > 0$.

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 line with the literature, the relationship between banks and depositors is exclusive, in the sense that a depositor only has one 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's project is successful. Banks offer 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:

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

The variable θ represents the fundamental of the economy and is uniformly distributed over the interval $[0, 1]$. We assume that $p(\theta) = \theta$ and $\mathbb{E}[\theta]R > 1$, which implies that the expected long-term return of the investment is higher than its short-term return.[4] 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 liquidates all its investment and distributes the proceeds pro-rata to all the withdrawing depositors at date 1.

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, early depositors withdraw to satisfy their consumption needs. Late depositors instead receive a private signal x_i about the

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

[4] The assumption of uniform distribution of fundamentals comes at no loss of generality. As argued by Goldstein and Pauzner (2005), results would hold for any function $p(\theta)$, as long as it is strictly increasing in θ. Under this condition, the probability of obtaining R can take any form.

fundamental of the economy. 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 late depositor uses her signal to infer both the fundamental of the economy and the withdrawal behavior of the others. On this basis, late depositors decide whether to withdraw at date 1 ("run") or wait until date 2. As we will show 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 banking literature, the deposit rate r_1 that banks pay at date 1 does not depend on the realization of the fundamental. Equally standard is our assumption that deposits are priced linearly. Thus, we abstract from the possibility that a bank conditions the deposit rate on the amount deposited.[5] In Appendix B, we instead account for the possibility that the amounts deposited depend on the deposit rate, by modifying the timing of actions so that the banks choose the deposit rate before consumers choose how much to deposit. We show that our results on the

[5]Empirical evidence supports this assumption. The difference between the national rates on jumbo deposits (equal to or above US$100,000) and non-jumbo deposits (below US$100,000) in 12-month CDs has been only 2 basis points on average over the period 2009-2021. Similarly tiny differences in returns apply to different maturities, too. Source: Own calculations on data from the FDIC.

existence of the saving externality are robust to this modification. While banks correctly anticipate that changes in the deposit rate affect the amounts deposited by consumers, they fail to enforce the constrained efficient amount of deposits, as this would distort the provided amount of liquidity insurance.

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 because 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 assume that the undeposited endowment is 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 would considerably complicate the analysis without affecting its main qualitative insights.[6]

3 Decentralized equilibrium

In this section, we characterize the decentralized equilibrium of an economy in which late depositors may run because they expect all the other depositors to do the same, i.e. there is a panic-driven run. In this economy, banks choose the deposit contract, all consumers take the consumption-saving decision, and late ones, based on their signals, decide when

[6] Deidda and Panetti (2017) formally show that introducing a portfolio problem in the Goldstein-Pauzner framework does not alter in any crucial way the characterization of depositors' withdrawal decisions and of the run threshold.

to withdraw following the threshold strategy:[7]

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

We solve the model by backward induction, and characterize a symmetric equilibrium so that we can focus our attention on the behavior of a representative bank. The definition of equilibrium is as follows:

Definition 1. *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 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:* $D = \int_i d_i di$*.*

3.1 Depositors' withdrawal decision

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

[7] 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.

a high posterior probability to a positive bank project return R at date 2, and infers that the other late depositors 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 depositor has a high incentive to withdraw early. This suggests that late depositors 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 fundamentals 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:

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

that is:

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

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.[8]

[8]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 $\underline{\theta}(1, d_i) > 2\varepsilon$.

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 for all late depositors. As in 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 $\frac{R(1-\lambda r_1)}{1-\lambda}d_i$ 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 depositors' 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 1. *Assume that 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 fundamentals lie in the intermediate region, it is optimal to follow the withdrawal strategy x^*_{-i} of all the other depositors $-i$. 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, when the fundamentals are above the lower dominance region, it is never optimal for a late depositor to run when she expects all other late depositors to withdraw at date 2. Since depositors' withdrawal decisions are symmetric, it follows that each depositor withdraws if her signal is lower than x^*_{-i}.

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:

$$\mathcal{V}_{-i}(\theta, n) = \begin{cases} \theta u\left(R\frac{1-nr_1}{1-n}d_{-i}\right) - u(r_1 d_{-i}) & \text{if } \lambda \leq n \leq \overline{n}, \\ 0 - u(\frac{d_{-i}}{n}) & \text{if } \overline{n} \leq n \leq 1, \end{cases} \tag{7}$$

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 1. *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:*

$$\theta^*(r_1, d_{-i}) = \frac{\displaystyle\int_{\lambda}^{\overline{n}} u(r_1 d_{-i})dn + \int_{\overline{n}}^{1} u\left(\frac{d_{-i}}{n}\right)dn}{\displaystyle\int_{\lambda}^{\overline{n}} u\left(R\frac{1-nr_1}{1-n}d_{-i}\right)dn}. \tag{8}$$

The threshold $\theta^*(r_1, d_{-i})$ is increasing in r_1 and is not neutral to a change in the size of *the individual deposit d_{-i}.*

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. This hinges on the existence of 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.[9]

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".

The run threshold $\theta^*(r_1, d_{-i})$ increases with the deposit rate r_1 offered by banks. 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.

Importantly, the run threshold $\theta^*(r_1, d_{-i})$ also depends on the size of the individual deposit d_{-i}. This effect is more involved than the one of r_1. On the one hand, a rise in the deposited amount increases depositors' repayment at date 1 thereby increasing incentives to run. On the other hand, it also increases the repayment at date 2 thereby lowering incentives to run. Formally, the sign of the effect of a change in the size of individual deposit d_{-i} is given by the sign of the following expression:

$$\int_{\lambda}^{\overline{n}} u'(c_1)c_1 dn + \int_{\overline{n}}^{1} u'\left(c_1^{run}(n)\right) c_1^{run}(n) dn - \int_{\lambda}^{\overline{n}} \theta^* u'\left(c_2(n)\right) c_2(n) dn, \tag{9}$$

where $c_1 = r_1 d_{-i}$, $c_2(n) = R\frac{1-nr_1}{1-n} d_{-i}$ and $c_1^{run}(n) = \frac{d_{-i}}{n}$. The first two terms in (9)

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

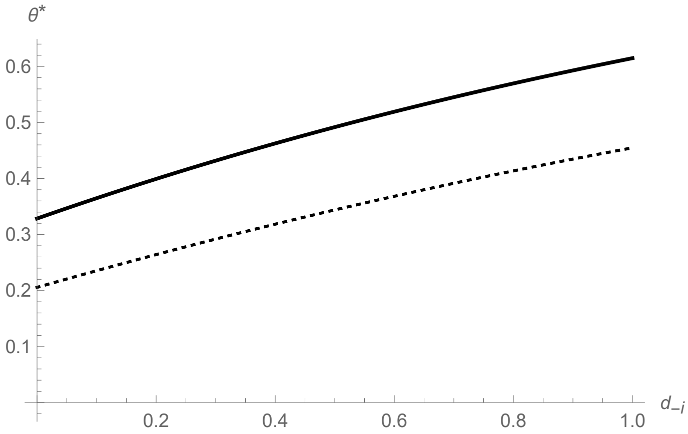

Figure 2: Individual deposit and the run threshold. The figure illustrates the effect of the size of the individual deposit d_{-i} on the run threshold θ^*. Parameters: $u(c) = \frac{(c+f)^{1-\sigma}-f^{1-\sigma}}{1-\sigma}$, $\sigma = 3$, $f = 4$, $R = 5$ and $r_1 = 1.1$ and $r_1 = 1.003$ for the solid and dotted line, respectively.

represent the marginal utilities of date-1 consumption when a run does not occur and depositors receive the promised consumption, and when a run occurs and they receive a pro-rata share of bank available resources, respectively. The third term is instead the expected marginal utility of date-2 consumption. As depositors are risk averse, the overall effect of a rise in d_{-i} depends on their expected level of consumption, in that they value an increase in consumption more when they are poorer. Hence, the overall effect of a rise in d_{-i} depends on their expected level of consumption at date 1 versus date 2. Specifically, while the first two terms of (9) are positive, the third term is negative and could, in principle, dominate them, as utility is concave and date-2 consumption goes to zero when the proportion of early withdrawers approaches $\overline{n}$. However, the assumption that $u(0) = 0$ implies an upper bound on the magnitude of $\lim_{c\to 0} u'(c)c$ that limits this effect. Figure 2 provides a numerical example in which this happens, and therefore the probability of a panic run is increasing in deposits. The comparison between the solid line and the dotted line also confirms Proposition 1 that θ^* is increasing in the deposit rate r_1.

3.2 Deposit rate and saving 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:

$$\max_{r_1} \int_0^{\theta^*(r_1,d_{-i})} u(d_i)d\theta + \int_{\theta^*(r_1,d_{-i})}^1 \left[\lambda u(r_1 d_i) + (1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)\right] d\theta. \quad (10)$$

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 nothing 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:

$$\max_{d_i, c_{0i}} u(c_{0i}) + \int_0^{\theta^*(r_1,d_{-i})} u(d_i)d\theta + \int_{\theta^*(r_1,d_{-i})}^1 \left[\lambda u(r_1 d_i) + (1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d_i\right)\right] d\theta, \quad (11)$$

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 Lemma 1 and Proposition 1, 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 2. *The decentralized equilibrium with panic runs is given by* $r_1 > 1$ *and*

$d > 0$ that solve:

$$\int_{\theta^*(r_1,d)}^{1} \left[u'(r_1 d) - \theta R u' \left(R \frac{1-\lambda r_1}{1-\lambda} d \right) \right] d\theta - \frac{\partial \theta^*(r_1,d)}{\partial r_1} \frac{\Delta}{\lambda d} = 0, \tag{12}$$

$$\begin{aligned} u'(1-d) = & \int_0^{\theta^*(r_1,d)} u'(d) d\theta + \\ & + \int_{\theta^*(r_1,d)}^{1} \left[\lambda r_1 u'(r_1 d) + \theta R(1-\lambda r_1) u' \left(R \frac{1-\lambda r_1}{1-\lambda} d \right) \right] d\theta, \end{aligned} \tag{13}$$

respectively and

$$d_i = d_{-i} = d = D, \tag{14}$$

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* (8) *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 (12), captures improved risk-sharing obtained from the transfer of consumption from late to early depositors. The latter, represented by the second term of (12), 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 that 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 (13). The latter comes from more consumption at dates 1 and 2, as captured by the right-hand side of (13).

We can substitute (14) and (12) into (13) and obtain an expression summarizing the decentralized equilibrium:

$$u'(1-D) = \int_0^{\theta^*(r_1,D)} u'(D)d\theta + \int_{\theta^*(r_1,D)}^1 u'(r_1D)d\theta - (1-\lambda r_1)\frac{\Delta}{\lambda D}\frac{\partial\theta^*(r_1,D)}{\partial r_1}. \quad (15)$$

The equation above 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.

4 Constrained efficiency and optimal policy

To study the efficiency of the decentralized equilibrium, we characterize a constrained-efficient benchmark. To do so, we consider a social planner who can only offer demand-deposit contracts like 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 (8), 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:

$$\lambda C_1 + (1-\lambda)\frac{C_2}{R} = 1 - C_0. \quad (16)$$

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

$$u(C_0) + \int_0^{\theta^*(r_1,D)} u(C_1^{\text{run}})d\theta + \int_{\theta^*(r_1,D)}^1 \left[\lambda u(C_1) + (1-\lambda)\theta u\left(C_2\right)\right] d\theta. \quad (17)$$

The following lemma characterizes the constrained efficient allocation.

Lemma 2. *The constrained-efficient equilibrium with panic runs is given by* $r_1 > 1$ *and* $D > 0$ *that solve:*

$$\int_{\theta^*(r_1,D)}^{1} \left[u'(r_1 D) - \theta R u'\left(R\frac{1-\lambda r_1}{1-\lambda}D\right)\right] d\theta - \frac{\partial \theta^*(r_1, D)}{\partial r_1}\frac{\Delta}{\lambda D} = 0, \tag{18}$$

$$\begin{aligned} u'(1-D) = & \int_{0}^{\theta^*(r_1,D)} u'(D) d\theta + \\ & + \int_{\theta^*(r_1,D)}^{1} \left[\lambda r_1 u'(r_1 D) + \theta R(1-\lambda r_1) u'\left(R\frac{1-\lambda r_1}{1-\lambda}D\right)\right] d\theta - \frac{\partial \theta^*(r_1, D)}{\partial D}\Delta, \end{aligned} \tag{19}$$

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* (8)*.*

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 economy, 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 (19). In other words, differently from the planner, the decentralized economy exhibits a "saving externality" in the sense that depositors do not internalize the effect of their consumption-saving decisions on the likelihood of panic runs.

To ease the comparison with the decentralized economy, it is useful to substitute (18) into (19) and obtain:

$$\begin{aligned} u'(1-D) = & \int_{0}^{\theta^*(r_1,D)} u'(D) d\theta + \\ & + \int_{\theta^*(r_1,D)}^{1} u'(r_1 D) d\theta - (1-\lambda r_1)\frac{\Delta}{\lambda D}\frac{\partial \theta^*(r_1, D)}{\partial r_1} - \frac{\partial \theta^*(r_1, D)}{\partial D}\Delta. \end{aligned} \tag{20}$$

The following proposition compares the social planner allocation with the decentralized equilibrium. This boils down to the comparison between (20) and (15), as the other equations that pin down the allocation are the same under the social planner as in the decentralized economy.

Proposition 3. *The decentralized equilibrium with panic runs is not constrained efficient. It exhibits an inefficient level of bank liquidity insurance and savings, and, in turn, excessive financial instability.*

By internalizing the effects of savings on the likelihood of panic runs, the social planner chooses a different level of savings than in the decentralized equilibrium. Hence, in the decentralized equilibrium, the level of deposits is constrained inefficient and financial fragility excessive, in that runs are too frequent. 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 constrained inefficient. As mentioned above, this is at odds with the results in Goldstein and Pauzner (2005) and is exclusively due to the saving externality and 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 (12) and (18) are identical. Thus, if depositors saved the constrained efficient amount, banks would provide the constrained efficient level of liquidity insurance.

4.1 Optimal policy

We have shown that the decentralized equilibrium features a saving externality. 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:

$$T = \tau D. \tag{21}$$

The consumer's date-0 budget constraint reads:

$$c_{0i} + (1+\tau)\, d_i = 1 + T. \tag{22}$$

Except for the above budget constraints, the economy is the same as described in Section 3. The following lemma characterizes the equilibrium conditions of the economy with taxes.

Lemma 3. *Given a tax on deposit holdings τ, the decentralized equilibrium is characterized by:*

$$\int_{\theta^*(r_1,d)}^{1} \left[u'(r_1 d) - \theta R u'\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] d\theta - \frac{\partial \theta^*(r_1,d)}{\partial r_1}\frac{\Delta}{\lambda d} = 0, \tag{23}$$

$$\begin{aligned}(1+\tau)u'\left(1-d\right) = &\int_{0}^{\theta^*(r_1,d)} u'\left(d\right) d\theta + \\ &+ \int_{\theta^*(r_1,d)}^{1} \left[\lambda r_1 u'(r_1 d) + \theta R(1-\lambda r_1) u'\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] d\theta,\end{aligned} \tag{24}$$

$$d_i = d_{-i} = d = D, \tag{25}$$

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

The tax policy creates a wedge in the intertemporal consumption-savings decision, thereby discouraging or encouraging savings. This can be seen by comparing (24) with (13). Optimal taxation is characterized in the following proposition.

Proposition 4. *The tax on deposit holdings that decentralizes the constrained efficient allocation is:*

$$\tau^{opt} = \frac{\Delta}{u'(1-D)}\frac{\partial \theta^*(r_1, D)}{\partial D}. \tag{26}$$

The optimal wedge is increasing in the marginal effect of deposits on the run probability $\frac{\partial \theta^*(r_1,D)}{\partial D}$ 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. The sign of the saving externality determines whether the optimal policy is a tax on or a subsidy to deposits. As long as the wedge is positive, i.e., $\frac{\partial \theta^*(r_1,D)}{\partial D}$, a benevolent policy-maker should tax deposits to correct over-saving and restore the constrained efficient allocation.

5 Conclusions

Does the size of a bank's deposit base matter for its fragility? If so, do agents correctly internalize this effect when deciding how much to deposit into a bank? We answer these questions in a banking model with endogenous depositors' 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, and that this effect is increasing in a reasonable parameter space. Second, we show that individual depositors do not internalize the effect on financial 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 constrained efficiency of the decentralized equilibrium. Policy-makers should induce individual depositors to internalize the effect of their consumption-saving decisions on financial stability. When the economy features over-saving, a tax on deposits is effective in restoring constrained efficiency.

The saving externality is rooted in the strategic complementarity characterizing depositors' withdrawal decisions. In other words, its existence is linked to the occurrence of panic-driven runs. This suggests that eliminating panic-driven runs should also make the saving externality disappear. The banking literature (Diamond and Dybvig, 1983) has focused on various policy interventions meant to remove the strategic complementarity in depositors' withdrawal decisions, like deposit insurance or a lender of last resort (LOLR) policy. As long as those policies fully guarantee depositors to receive the promised repayments in all possible circumstances, they are effective in restoring the constrained efficient allocation by removing the saving externality.

In more realistic frameworks where the presence of fundamental risk prevents the policy authority from fully guaranteeing depositors to receive the promised repayments, the saving externality instead might not disappear. An example of this case is an intervention designed to provide liquidity only to solvent but illiquid banks, in line with the prescription of Bagehot (1873). Such a policy removes panic-driven runs but does not eliminate the dependence of individual depositors' withdrawal decisions on other depositors' actions, as those determine whether the bank is solvent or not. As a result, the saving externality is still present.[10]

This suggests that there are circumstances in which prudential policies to resolve coordination failures may need to be complemented by other interventions meant to resolve the saving externality. 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 banks, emergency liquidity provision by central banks may also distort savers' incentives, and translate into an excessively large and fragile financial system.

[10]The characterization of this case is available upon request.

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). Financial Intermediaries and Markets. *Econometrica 72*(4), 1023–1061.

Bagehot, W. (1873). *Lombard Street: A Description of the Money Market.* London, UK: Henry S. King & Co.

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). 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.

Deidda, L. G. and E. Panetti (2017). Banks' liquidity management and systemic risk. Banco de Portugal Working Paper No. 2017-13.

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

Egan, M., A. Hortaçsu, and G. Matvos (2017, January). Deposit Competition and Financial Fragility: Evidence from the US Banking Sector. *American Economic Review 107*(1), 169–216.

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., M. Golosov, and A. Tsyvinski (2009). A Theory of Liquidity and Regulation of Financial Intermediation. *Review of Economic Studies 76*, 973–992.

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, May). Banking panics as endogenous disasters and the welfare gains from macroprudential policy. *American Economic Review 110*, 463–469.

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.

Iyer, R. and M. Puri (2012). Understanding Bank Runs: The Importance of Depositor-Bank Relationships and Networks. *American Economic Review 102*(4), 1414–1445.

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). 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.

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

Peck, J. and A. Setayesh (2022, January). Bank Runs and the Optimality of Limited Banking. *Review of Economic Dynamics Forthcoming.*

Stein, J. C. (2012). Monetary Policy as Financial-Stability Regulation. *Quarterly Journal of Economics 127*(1), 57–95.

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 Lemma 1. The proof is done by contradiction. Assume first that depositor i finds it optimal not to 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.[11] □

Proof of Proposition 1. 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:

$$n(\theta, x^*) = \begin{cases} 1 & \text{if } \theta \leq x^* - \varepsilon \\ \lambda + (1-\lambda)\left(\frac{x^* - \theta + \varepsilon}{2\varepsilon}\right) & \text{if } \; x^* - \varepsilon \leq \theta \leq x^* + \varepsilon \\ \lambda & \text{if } \theta \geq x^* + \varepsilon \end{cases} \tag{27}$$

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 depositors 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

[11] This result is based on guessing that $\underline{\theta}(r_1, d_i) < x_{-i}^*$, which is always verified in a symmetric equilibrium. Alternatively, if $x_{-i}^* < \underline{\theta}(r_1, d_i) < x_i^*$, in the interval $(x_{-i}^*, \underline{\theta}(r_1, d_i))$ the depositor i might find it optimal to run if $u(r_1 d_i) > \theta u(r_2 d_i)$ and therefore $x_i^* = \max\{x_{-i}^*, \underline{\theta}(r_1, d_i)\}$. This case would not yield any difference relative to the equilibrium analyzed in Section 3, and its characterization is available upon request.

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 (27) since for any realization of θ, the proportion of depositors running is deterministic. The function $\Delta(x_i, n(\theta))$ is equal to

$$\Delta(x_i, n(\theta)) = \frac{1}{2\epsilon}\int_{x_i-\epsilon}^{x_i+\epsilon} \mathcal{V}(\theta, n(\theta))d\theta, \tag{28}$$

where $\mathcal{V}(\theta, n(\theta))$ is given in (7) and $n(\theta) = n(\theta, x^*)$ as given in (27). 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:

$$f(\theta, r_1, d_{-i}) = \int_{\lambda}^{\frac{1}{r_1}} \left[\theta u\left(R\frac{1-nr_1}{1-n}d_{-i}\right) - u(r_1 d_{-i})\right] dn + \int_{\frac{1}{r_1}}^{1} \left[u(0) - u\left(\frac{d_{-i}}{n}\right)\right] dn, \tag{29}$$

where $\theta(n) = x^*_{-i} + \varepsilon - 2\varepsilon\frac{(n-\lambda)}{1-\lambda}$ from (27). Equation (29) follows from (7) 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 we use the implicit function theorem on (29) and obtain:

$$\frac{\partial\theta^*(r_1, d_{-i})}{\partial r_1} = -\frac{\frac{\partial f(\cdot)}{\partial r_1}}{\frac{\partial f(\cdot)}{\partial \theta^*}} \quad . \tag{30}$$

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

$$\frac{\partial f(\cdot)}{\partial r_1} = -d_{-i} \int_{\lambda}^{\frac{1}{r_1}} \left[u'(r_1 d_{-i}) + \theta^* \frac{nR}{1-n} u' \left(R \frac{1-nr_1}{1-n} d_{-i} \right) \right] dn < 0. \tag{31}$$

Hence, the proposition follows. □

Proof of Proposition 2. Differentiating the bank's objective function in (10) with respect to r_1, we obtain (12). Similarly, differentiating (11) with respect to d yields (13).

To prove that $r_1 > 1$, evaluate (12) at $r_1 = 1$ using $d_i = d_{-i} = d = D$. This leads to:

$$\lambda \int_{\underline{\theta}}^{1} \left[u'(d)d - \theta R d u'(Rd) \right], \tag{32}$$

since $\theta^* \to \underline{\theta}$ when $r_1 = 1$, and $\Delta = 0$ by definition of $\underline{\theta}$ in (6). This expression is positive because relative risk aversion is larger than 1 for $c > 0$ and $\bar{c} < I$. To see that, notice that $u'(d)d - \theta R d u'(Rd) > u'(d)d - Rdu'(Rd)$ and $u'(c)c$ is decreasing in c. This follows directly from $-u''(c)c/u'(c) > 1$. Notice that the solution is an interior because for given d, the equilibrium r_1 must be consistent with runs not always occurring, i.e., with $\theta^* < \overline{\theta}$. Choosing r_1 such that $\theta^* \to \overline{\theta} \to 1$ would imply that depositors obtain $u(d)$, which is even lower than the utility that they could obtain by setting $r_1 = 1$. The equilibrium size of deposit d is also an interior solution for any r_1 since by choosing $d = 0$ depositors would accrue $u(1)$, which is lower than what they could obtain by accessing liquidity insurance provided by bank deposits. Thus, the proposition follows. □

Proof of Lemma 2. The two conditions in the lemma are obtained by simply differentiating (17) with respect to r_1 and D. The proof of $r_1 > 1$ is analogous to that of Proposition 2. □

Proof of Proposition 3. The proof follows directly from the comparison of (15) and (20). When evaluating (15) at the optimal level of investment solving (20), (15) is positive since the two first-order conditions only differ for the term $\frac{\partial \theta^*}{\partial D}\Delta$. This implies that in the decentralized allocation, the level of aggregate deposits D is different than that chosen by the planner. The results about the excessively high level of financial fragility follow directly from the fact that the planner chooses the level of aggregate deposits to limit runs, as panic runs are inefficient and reduce the economy expected aggregate welfare in (17). 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 take r_1 as the solution to (12). 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 3. The derivation follows the steps of the proof of Proposition 2. The tax only affects the consumer's problem. For a general run threshold $\tilde{\theta}$, the problem becomes:

$$\max_{d}\ u\left[1-(1+\tau)d+T\right]+\int_{0}^{\tilde{\theta}(r_1,d)} u\left(d\right)d\theta+ \\ +\int_{\tilde{\theta}(r_1,d)}^{1}\left[\lambda u(r_1 d)+(1-\lambda)\theta u\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta. \tag{33}$$

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

$$(1+\tau)\,u'\left(1-d\right)=\int_{0}^{\tilde{\theta}(r_1,d)} u'\left(d\right)d\theta+ \\ +\int_{\tilde{\theta}(r_1,d)}^{1}\left[\lambda r_1 u'(r_1 d)+\theta R(1-\lambda r_1)u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta. \tag{34}$$

Hence, the lemma follows. □

Proof of Proposition 4. Constrained efficiency is determined by Lemma 2. By substitution, we find that the expression in the lemma makes the decentralized equilibrium identical to the constrained efficient one. □

B Alternative specification: Individual deposits as a function of the deposit rate

In this appendix, we prove the robustness of our result concerning the existence of the saving externality, by considering an alternative specification. Specifically, we study an economy in which the individual saving decisions depend on the deposit rate set by the bank. To achieve this, we modify the timing of the actions at date 0 as follows: First, the bank sets r_1; then, for given r_1, each depositor i chooses the amount of deposits d_i.

As in the main text, the model is solved by backward induction. The date-1 decision is as in the main text, and the same applies to the choice of d. Hence, the run threshold is still given by (8) and $d_i = d_{-i} = d$ still solves (13). However, it is evident that the bank has the possibility to affect the equilibrium amount of deposits via the choice of the deposit rate r_1. This choice is taken by the bank in the previous stage as the solution to:

$$\max_{r_1} \; EU = \int_0^1 u\left(1-d\right) d\theta + \\ + \int_0^{\theta^*} u\left(d\right) d\theta + \int_{\theta^*}^1 \left[\lambda u\left(r_1 d\right) + \left(1-\lambda\right) u\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] d\theta. \tag{35}$$

The first-order condition with respect to r_1 can be expressed in a compact formulation as:

$$FOC_{r_1} = \frac{\partial EU}{\partial r_1} + \frac{\partial EU}{\partial d}\frac{dd}{dr_1}. \tag{36}$$

Using (13), we know that:

$$\frac{\partial EU}{\partial d} = 0 + \frac{\partial \theta^*}{\partial d}\Delta, \tag{37}$$

with Δ as specified in Proposition 2. It follows that the condition $FOC_{r_1} = 0$ can be expressed as follows:

$$\int_{\theta^*}^1 \left[u'(r_1 d) d\theta - R\theta u'\left(R\frac{1-\lambda r_1}{1-\lambda} d\right)\right] = \left(\frac{\partial \theta^*}{\partial r_1} + \frac{\partial \theta^*}{\partial d}\frac{dd}{dr_1}\right)\frac{\Delta}{\lambda d}. \tag{38}$$

Moving on to the analysis of the constrained efficient allocation, we follow the same steps as before but notice that the planner internalizes the fact that $\frac{\partial \theta^*}{\partial d}$ is different from zero, so $\frac{\partial EU}{\partial d}=0$. This gives the following first-order condition with respect to r_1:

$$\int_{\theta^*}^{1} R\theta u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)d\theta=\int_{\theta^*}^{1} u'\left(r_1 d\right)d\theta-\frac{\partial\theta^*}{\partial r_1}\frac{\Delta}{\lambda d}, \tag{39}$$

while the first-order condition with respect to d is equal to:

$$u'\left(1-d\right)=\int_{0}^{\theta^*} u'\left(d\right)d\theta+\int_{\theta^*}^{1}\left[\lambda r_1 u'\left(r_1 d\right)+R\theta\left(1-\lambda r_1\right)u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta-\frac{\partial\theta^*}{\partial d}\Delta. \tag{40}$$

Substituting (36) into (13), we obtain:

$$u'\left(1-d\right)=\int_{0}^{\theta^*} u'\left(d\right)d\theta+\int_{\theta^*}^{1} u'\left(r_1 d\right)d\theta-\left(1-\lambda r_1\right)\left(\frac{\partial\theta^*}{\partial r_1}+\frac{\partial\theta^*}{\partial d}\frac{dd}{dr_1}\right)\frac{\Delta}{\lambda d}. \tag{41}$$

Doing the same for the planner yields instead:

$$u'\left(1-d\right)=\int_{0}^{\theta^*} u'\left(d\right)d\theta+\int_{\theta^*}^{1} u'\left(r_1 d\right)d\theta-\left(1-\lambda r_1\right)\frac{\partial\theta^*}{\partial r_1}\frac{\Delta}{\lambda d}-\frac{\partial\theta^*}{\partial d}\Delta. \tag{42}$$

Comparing (41) and (42), they differ because of the term:

$$\frac{dd}{dr_1}-\frac{\lambda}{1-\lambda r_1}d, \tag{43}$$

We can compute $\frac{dd}{dr_1}$ using the implicit function theorem as follows:

$$\frac{dd}{dr_1}=-\frac{\frac{\partial FOC_d}{\partial r_1}}{SOC_d}, \tag{44}$$

where:

$$\begin{aligned}\frac{\partial FOC_d}{\partial r_1}=&\frac{\partial\theta^*}{\partial r_1}\left[\lambda r_1 u'\left(r_1 d\right)d\theta+R\theta^*\left(1-\lambda r_1\right)u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)-u'\left(d\right)\right]+\\&-\lambda\int_{\theta^*}^{1}\left[u'\left(r_1 d\right)d\theta-R\theta u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta+\\&-\lambda d\int_{\theta^*}^{1}\left[r_1 u''\left(r_1 d\right)d\theta-R^2\theta\frac{\left(1-\lambda r_1\right)}{1-\lambda}u''\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right]d\theta,\end{aligned} \tag{45}$$

and:

$$SOC_d = -u''(1-d) + \frac{\partial \theta^*}{\partial d}\left[\lambda r_1 u'(r_1 d)\, d\theta + R\theta^*(1-\lambda r_1)\, u'\left(R\frac{1-\lambda r_1}{1-\lambda}d\right) - u'(d)\right] + \\ -\int_0^{\theta^*} u''(d)\, d\theta + \int_{\theta^*}^1 \left[\lambda r_1^2 u''(r_1 d)\, d\theta + R^2\theta\frac{(1-\lambda r_1)^2}{1-\lambda}u''\left(R\frac{1-\lambda r_1}{1-\lambda}d\right)\right] d\theta. \tag{46}$$

The expression in (44) is different from zero. In other words, the saving externality is present also when the bank anticipates that its choice of the deposit rate affects depositors' consumption-saving decisions. The intuition behind this result is that the bank tries to trade off setting a deposit rate that induces depositors to choose the constrained efficient amount of deposits, with the one that achieves the constrained efficient level of liquidity insurance. However, the bank cannot achieve both objectives at the same time with a single instrument, so the equilibrium allocation is still constrained inefficient.