REM WORKING PAPER SERIES

The Welfare Costs of Self-Fulfilling Bank Runs

Elena Mattana and Ettore Panetti

REM Working Paper 017-2017

November 2017

REM – Research in Economics and Mathematics
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

The Welfare Costs of Self-Fulfilling Bank Runs*

Elena Mattana[†] Ettore Panetti[‡]

This draft: November 2017

Abstract

We study the welfare implications of self-fulfilling bank runs and liquidity requirements, in a neoclassical growth model where banks, facing long-lasting possible runs, can choose in any period a run-proof asset portfolio. In this framework, runs distort banks' insurance provision against idiosyncratic liquidity shocks, and liquidity requirements resolve this distortion by forcing a credit tightening. Quantitatively, the welfare costs of self-fulfilling bank runs are equivalent to a constant consumption loss of up to 2.5 percent of U.S. GDP. Depending on fundamentals, liquidity requirements might generate small welfare gains, but also increase the welfare costs by up to 1.8 percent.

Keywords: financial intermediation, bank runs, regulation, welfare

JEL Classification: E21, E44, G01, G20

*This paper previously circulated as "A Dynamic Quantitative Macroeconomic Model of Bank Runs". We would like to thank Aleksander Berentsen, Luca Deidda, Martin Eichenbaum, Roberto Robatto, Pedro Teles, and the seminar participants, among others, at the IFABS Conference 2014, the SUERF/UniCredit & Universities Foundation Workshop, SAET Conference 2015, EEA Mannheim 2015, Vienna Macro Cafe 2015, ASSET Conference 2015 and FIRS Conference 2016 for their valuable comments. Elena Mattana gratefully acknowledges the financial support by FNRS (Fonds de la Recherche Scientifique) Belgium, as part of the Research Project PDR T.0044.13. Ettore Panetti gratefully acknowledges the financial support by FCT (Fundação para a Ciência e a Tecnologia) Portugal, as part of the Strategic Project (Pest-OE/EGE/UI0436/2014). The analyses, opinions and findings of this paper represent the views of the authors, and are not necessarily those of Banco de Portugal or the Eurosystem.

[†]Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4 building 2621, 8210 Aarhus V, Denmark. Email: emattana@econ.au.dk.

[‡]Economics and Research Department, Banco de Portugal, Avenida Almirante Reis 71, 1150-012 Lisboa, Portugal, and CRENoS-University of Sassari, UECE-ISEG and REM. Email: ettore.panetti@bportugal.pt.

1 Introduction

What are the general-equilibrium effects of self-fulfilling bank runs, and of the liquidity requirements needed to offset them? And how large are these effects on welfare? The interest in these questions comes from the observation that self-fulfilling bank runs are not a phenomenon of the past:[1] in fact, they may occur whenever long-term illiquid assets are financed by short-term liquid liabilities, and the providers of short-term funds all lose confidence in the borrower's ability to repay, or are afraid that other lenders are losing their confidence. There exists a wide consensus that the U.S. financial crisis of 2007-2009 can be interpreted as a self-fulfilling bank run of financial intermediaries on other financial intermediaries, and that money market funds and life insurance funds also experienced run-like episodes during the same period (Gorton and Metrick, 2012; Foley-Fisher et al., 2015), leading to a peak-to-trough decline in real per capita GDP of 4.8 percent, with a widespread impact on asset markets, housing markets, government debt and unemployment (Reinhart and Rogoff, 2009, 2014). These numbers justified a massive government intervention,[2] as well as the introduction of new forms of financial regulation, in particular the liquidity requirements of Basel III, with the explicit objective of taming the adverse effects of self-fulfilling bank runs in the future. Yet, these numbers do not allow us to identify the channels through which self-fulfilling bank runs affect the real economy, and their costs in terms of welfare for the whole society. As a consequence, it is not even clear whether it is correct to impute the whole observed drop in GDP to these events, and what the correct government intervention to address them should be. The present work overcomes these limitations, by studying and quantifying the welfare implications of self-fulfilling bank runs and liquidity requirements in a neoclassical growth model with a fully microfounded banking system.

Our model is based on three building blocks, all considered standard workhorses in their own fields. The first one is the neoclassical growth model: an infinite-horizon, general-

[1] During the "National Banking Era", the U.S. economy experienced seven major bank runs, and twenty non-major ones (Jalil, 2015).

[2] In 2008-2009, the U.S. Treasury rescued many financial and non-financial corporations via the "Troubled Asset Relief Program", with an total investment of around US$400 billion. In the same period, the Federal Reserve, through its liquidity facilities, extended credit to the U.S. financial system for around US$1.5 trillion.

equilibrium, dynamic model populated by households and firms. As generally known, this model lacks a proper role for a banking system, as the households provide resources to the firms without any intermediation. This leads to our second building block: the theory of banking of Diamond and Dybvig (1983), where banks provide insurance to their depositors against idiosyncratic liquidity shocks, that force them to consume in an interim period, i.e. before the maturity of their investment. This environment is interesting for two reasons. First, because it rationalizes liquidity and maturity transformation, as a mechanism to decentralize the efficient provision of insurance against idiosyncratic liquidity shocks. Second, because it features financial fragility, as a consequence of multiple equilibria: in fact, this economy exhibits one equilibrium in which only those depositors hit by the idiosyncratic liquidity shocks withdraw in the interim period, and one in which all depositors "run", in the sense that they all withdraw in the interim period, because they expect everyone else to do the same, eventually forcing the banks to go bankrupt. Finally, our third building block is the seminal paper by Cooper and Ross (1998) (and its refinement by Ennis and Keister (2006)) in which this multiplicity of equilibria is solved by assuming that the depositors choose to run in accordance with the realization of an extrinsic "sunspot", completely uncorrelated to the fundamentals of the economy.[3]

More formally, at each point in time the economy is populated by a cohort of one-period-lived risk-averse agents, who are hit by a private idiosyncratic liquidity shock, that makes them willing to consume either in an intermediate period, called "night", or when production takes place, in the following "morning". To hedge against these shocks, the agents get access to an infinitely-lived bank, operating as a social planner, whose objective is to maximize their expected welfare. To this end, the bank pools the idiosyncratic risk of the depositors: it collects their deposits, and offers them a standard deposit contract, stating how much they can withdraw each morning and each night. In turn, the bank finances the contract by investing every period in liquidity, stored for night withdrawals, and in loans to firms, that

[3] Our focus is on one specific role – which we believe is key – played by the banking system in the real economy: the management of liquidity, to hedge against idiosyncratic uncertainty. In this sense, here we do not consider the banks as monitors of investments or producers of information, as in Diamond (1984) or Holmstrom and Tirole (1998) Similarly, in modelling financial crises as sunspot-driven bank runs, we do not take into account crises arising from shocks to the fundamentals of the economy, as in Allen and Gale (1998).

either get repaid the following morning, or can be liquidated at night with a recovery rate smaller than 1. In this environment, in the absence of any other friction, the bank chooses an asset portfolio and deposit contract ensuring perfect intratemporal and intertemporal risk sharing: the deposit contract equalizes morning and night consumption in every period, and the asset portfolio satisfies an Euler equation, i.e. the marginal rate of substitution between consumption in the night of period t and in the morning of period $t+1$ is equal to the marginal rate of transformation of the production technology of the firm to which the bank lends.

The peculiar feature of this economy is the presence of financial fragility: in parallel to the equilibrium described above (that from now on we call "no-run"), the economy also exhibits a "run" equilibrium. In the run equilibrium all depositors withdraw at night regardless of the realization of their idiosyncratic shocks, if they all expect the other depositors to do the same, and they know that the bank does not hold a sufficient amount of liquidity to honor the contract with all of them. Whenever a run equilibrium and a no-run equilibrium coexist, the depositors coordinate between them, and choose to run in accordance with the realization of an extrinsic event, called "sunspot", happening with some exogenous probability. The assumption of an exogenous probability is based on some solid empirical evidence (Foley-Fisher et al., 2015), and allows us to distinguish in an elegant yet parsimonious way among (i) periods of financial stability when runs are not possible, (ii) periods of financial fragility when runs are possible because the probability of the realization of the sunspot is positive, and (iii) periods of crises when they realize.

The bank takes into account the aforementioned equilibrium selection mechanism, and chooses accordingly the asset portfolio and the deposit contract. One possible alternative is a "run-proof" contract, in which the bank holds a sufficient amount of liquidity to pay all night withdrawals, even in the case of a run. To this end, the bank has to satisfy a "run-proof constraint": the amount of liquidity in excess of what it would need to pay night consumption, plus the potential proceeds from loan liquidation, must be sufficient to pay the extra demand of night consumption at a run, by those depositors who are day consumers but withdraw at night anyway. The second alternative is instead that the bank offers a contract with possible runs, according to which it chooses not to completely overcome financial fragility, and let runs happen with some positive probability, in which case it pays the depositors pro-rata as in Allen and Gale (1998).

As there is no closed-form solution for the asset portfolio and deposit contract (and, as a consequence, for the expected welfare) of the two contracts, we rely on a numerical characterization of the banking equilibrium. To disentangle the short- and long-term mechanisms at play, we start our analysis by studying an economy with only one period of financial fragility: the probability of the realization of the sunspot (or, equivalently, the probability of a run) takes a positive value, and then goes back to zero. In order to analyze the bank choice between the run-proof contract and the contract with possible runs, we calculate the welfare costs of both contracts, expressed in terms of units of consumption equivalents, i.e. the constant percentage drop in consumption that would make the no-run equilibrium equivalent to the two contracts, and let the bank pick the one with the lowest value. The welfare costs of the run-proof contract turn out to be independent of the probability of a run, and decreasing in the loan liquidation value, as the higher that is, the slacker the run-proof constraint is, too: in fact, for a loan liquidation value above 0.17, the run-proof contract provides an allocation equivalent to the no-run equilibrium, and the welfare costs are zero. In contrast, the welfare costs of the contract with possible runs are ceteris paribus decreasing in the loan liquidation value, but increasing in the probability of a run. For low values of the loan liquidation value, the welfare costs of the contract with possible runs are lower than those of the run-proof contract, hence the bank chooses to be not run proof. However, the welfare costs of the run-proof contract decrease faster than those of the contract with possible runs. Hence, there exists a unique threshold at which the bank switch to the run-proof contract. Interestingly, as the probability of a run increases and the welfare costs of the run-proof contract remain unchanged, while the welfare costs of the contract with possible runs increase, the threshold at which the bank switches from a contract with possible runs to a run-proof contract decreases, up to the point at which the probability of a run is so high that the bank chooses to be run proof irrespective of it. In other words, the bank independently choose to be run proof only when both the loan liquidation value and the probability of a run are sufficiently high: in our calculation, above 0.11 and 0.014, respectively. The resulting welfare costs of the banking equilibrium with runs, compared to the no-run equilibrium, are the highest whenever the loan liquidation value is the lowest, and the probability of a run is so high that the bank chooses the run-proof contract. This combination brings about welfare costs of 0.23 percentage points in units of consumption equivalents, equal to a constant consumption loss of 0.16 percent of U.S. GDP, at 2014 levels.

We conclude the analysis by characterizing the equilibrium of an economy where financial fragility might be long lasting. To this end, we assume that the economy starts with a positive probability of a run, and that this probability goes to zero in the following period only with some positive probability. Accordingly, the bank solves a "hybrid" dynamic discrete-choice problem, as it chooses in every period some continuous variables (i.e. the asset portfolio and the deposit contract) and one discrete variable, i.e. whether to be run proof. To solve this model, we adapt the procedure developed by Chatterjee and Eyigungor (2012), and introduce a publicly-observable i.i.d. shock to the utility of the depositors in the case of a run, that can be interpreted as a taste shock or an institutional shock during periods of financial fragility. In such a framework, we find long-lasting effects of self-fulfilling bank runs on credit provision and GDP, leading to welfare costs of up to around 3.6 percentage points in units of consumption equivalents, equal to a constant drop in consumption of 2.5 percent of U.S. GDP.

Finally, we use the previous results to evaluate the welfare costs of a regulatory intervention that makes the banking system completely run proof, via two different liquidity requirements: one that forces the bank to be run proof, i.e. so that the liquidation value of its whole asset portfolio is sufficient to pay all depositors in the case of a run, and one that forces it to be "narrow", i.e. so that it is able to serve all depositors with liquidity, and avoid loan liquidation. Our results show that, for intermediate values of the probability of a run and low loan liquidation value, the bank chooses the contract with possible runs in very few periods of financial fragility. Yet, the possible realization of runs in some of them leaves space for some small welfare gains from regulation: in fact, imposing a run-proof liquidity requirement can lead to a drop in welfare costs, with respect to the unregulated equilibrium, of up to 0.22 percentage points in units of consumption equivalents, equal to a constant consumption gain of 0.15 percent of U.S. GDP; similarly, imposing a narrow-banking liquidity requirement can lead to a drop in welfare costs of up to 0.10 percentage points in units of consumption equivalents, equal to a constant consumption gain of 0.07 percent of U.S. GDP.

Yet, some high welfare costs can also arise from financial regulation. From what said above, it is clear that the imposition of a run-proof liquidity requirement brings about an increase in welfare costs, with respect to the unregulated equilibrium, whenever the probability of a run and the loan liquidation value are both so low that the bank would rather choose a contract with possible runs, but is instead forced to be run proof. Quantitatively, this translates into an increase in welfare costs of up to 0.16 percentage points in units of consumption equivalents (equal to a constant consumption loss of 0.11 percent of U.S. GDP) in the model with only one period of financial fragility, and of up to 2.5 percentage points in units of consumption equivalents (equal to a constant consumption loss of 1.8 percent of U.S. GDP) in the model with long-lasting financial fragility. Conversely, the imposition of the narrow-banking liquidity requirement brings about an increase in welfare costs also whenever the loan liquidation value is so high that the bank would be able to sustain the no-run equilibrium, and is instead forced to hold excess liquidity. This translates into an increase in welfare costs of up to 0.19 percentage points in units of consumption equivalents (equal to a constant consumption loss of 0.13 percent of U.S. GDP) in the model with only one period of financial fragility, and of up to 3.5 percentage points (equal to a constant consumption loss of 2.4 percent of U.S. GDP) in the model with long-lasting financial fragility. The marked difference between one-period and long-lasting financial fragility leads us to the conclusion that the highest costs of self-fulfilling bank runs and liquidity requirements originate neither from their immediate impact on the real economy nor from their recovery, but from their persistence.

The present paper contributes to the analysis of the general-equilibrium effects of financial shocks and financial regulation (Van den Heuvel, 2008; Gertler and Kiyotaki, 2015; Mendicino et al., 2017) by offering a fully microfounded assessment of the distortions induced by self-fulfilling bank runs and liquidity requirements. The most recent reference in this literature, by Segura and Suarez (2017), differs from our work in at least two important respects. First, they study the welfare costs of excessive maturity transformation and the welfare gains of regulating debt maturity, by focusing on the banks' liability composition, while here we study the effect of maturity transformation on banks' asset structure, and the welfare effects of liquidity requirements. Second, in their framework banks faces exogenous crisis states, that represent systemic liquidity crises in a reduced form, while we endogenize the emergence of financial fragility, as a consequence of depositors' self-fulfilling expectations and banks' ability to react to them. As far as this last point is concerned, the present paper develops a dynamic general-equilibrium theory that makes reference to the literature on dynamic banking (Bencivenga and Smith, 1991; Qi, 1994; Allen and Gale, 1997; Ennis and Keister, 2003) but with a focus on the welfare distortions of self-fulfilling runs, rather than on growth and intertemporal risk sharing. Finally, from a methodological point of view, our work is related to the most recent literature on the general-equilibrium effects of multiple equilibria (Bocola and Dovis, 2016; Robatto, 2017) and to the application of dynamic discrete-choice models to the quantitative analysis of sovereign default (Chatterjee and Eyigungor, 2012; Muller et al., 2016): in that sense, our work is the first one – to the best of our knowledge – to apply those techniques to models of self-fulfilling bank runs.

2 A Dynamic Model of Banking

Time is infinite and discrete, and every period is divided into two sub-periods, that we label 1 and 2, and call "day" and "night". At every point in time, the economy is populated by a cohort made of a unitary continuum of one-period-lived agents. All agents in the economy are affected by some idiosyncratic uncertainty, that hits them in the form of a preference shock. Being ex-ante equal, every agent draws a type $\theta_t \in \{0, 1\}$, where $0 < \pi < 1$ is the probability of being of type 1, and $1 - \pi$ is the probability of being of type 0. The preference shocks are private information, and are independent and identically distributed across the agents. Therefore, by the law of large numbers, the cross-sectional distribution of the types is equivalent to their probability distribution: π is the fraction of agents who turn out to be of type 1, and $1 - \pi$ is the fraction of agents who turn out to be of type 0. The role of the idiosyncratic shocks is to affect the sub-period when the agents enjoy consumption. This happens according to the utility function:

$$U(c_{2t}, c_{1t+1}, \theta_t) = \theta_t u(c_{2t}) + \beta(1-\theta_t)u(c_{1t+1}), \tag{1}$$

where the parameter β is lower than 1, and represents a discount factor. This functional form states that, depending on the realization of the idiosyncratic shock, each agent either consumes during the night of the same period (when $\theta_t = 1$) or in the day of the following one (when $\theta_t = 0$): in that respect, we talk about "night consumers" and "day consumers", respectively. The felicity function $u(c)$ is increasing, strictly concave, and satisfies the Inada conditions. Moreover, the degree of relative risk aversion is strictly larger than 1.

In order to hedge against the idiosyncratic shocks, the agents have two real technologies. The first is a storage technology, that we call "liquidity", yielding one unit of consumption in every sub-period for each unit invested in the previous one. The second technology is a neoclassical production function $Y_t = F(K_t)$, employed by a large number of competitive firms. We assume that the only input of the production technology is capital, and that capital needs "time to build": the amount invested at time t matures in the morning of $t+1$, yields a return equal to its marginal productivity, and then depreciates at rate d. In what follows, to save on notation, we assume without loss of generality that $d = 1$, and relax this hypothesis in the numerical analysis.

To invest in the two real technologies, the agents get access to a infinitely-lived bank, that operates as a social planner and maximizes the weighted sum of the expected welfare of all cohorts, where the weight of a cohort born at time t is β^t.[4] Moreover, the agents are "isolated", in the sense that they do not interact one with each other, once they enter in a banking relationship.[5] The bank collects the deposits, and performs two tasks: (i) it invests in liquidity on behalf of the depositors; (ii) it provides capital to the firms, in the form of loans. Loans can be liquidated at night (i.e., before being invested), using a "liquidation

[4]Incidentally, there is no generally accepted criterion for the aggregation of different cohorts. Yet, we employ decreasing weights so that the objective function of the planner is bounded and the problem becomes mathematically tractable. See Qi (1994) for a discussion of the hypothesis of an infinitely-lived bank.

[5]In a static environment, Jacklin (1987) shows that, if we relax this hypothesis, the banking equilibrium would be run proof, and would not provide more welfare than autarky.

technology" allowing the bank to recover $r < 1$ units of consumption for each unit liquidated. More formally, the bank solves the following problem:

$$\max_{\{c_{2t}, c_{1t+1}, \ell_t, k_{t+1}, z_t, D_t\}} \sum_{t=0}^{\infty} \beta^t \big[\pi u(c_{2t}) + \beta(1-\pi)u(c_{1t+1})\big], \tag{2}$$

subject to the budget constraints:

$$(1-\pi)c_{1t} + \ell_t + k_{t+1} \leq F(k_t - D_{t-1}) + z_{t-1}, \tag{3}$$

$$\pi c_{2t} + z_t \leq \ell_t + rD_t, \tag{4}$$

$$k_{t+1} \geq D_t, \tag{5}$$

to the incentive-compatibility constraint $c_{2t} \leq c_{1t+1}$, and to the non-negativity constraints $z_t \geq 0$ and $D_t \geq 0$, which must all hold at any point in time.[6] The bank maximizes the expected welfare of the depositors, by choosing an incentive-compatible deposit contract $\{c_{2t}, c_{1t+1}\}$, an asset portfolio of liquidity ℓ_t and loans k_{t+1}, the amount of excess liquidity z_t to roll over to the following period, and the amount of loans to liquidate D_t. The budget constraint in (3) states that the bank employs the production from the unliquidated loans $k_t - D_t$, provided to the production sector in the previous period, and the excess liquidity z_{t-1} rolled over from the previous period, to pay the day consumption c_{1t} (chosen in period $t-1$) of the $(1-\pi)$ day consumers, and the financial investment in liquidity and loans. Then, the bank uses liquidity ℓ_t, plus the return from the loan liquidation rD_t, to pay the night consumption c_{2t} to the π night consumers, and for excess liquidity z_t. To clearly point out the difference between the loans allocated at date t and the actual unliquidated loans (which is going to be important when we introduce bank runs), we explicitly impose, in equation (5), that the amount of loans D_t that the bank can liquidate cannot be larger than the actual total loans k_{t+1}. Finally, since the realization of the idiosyncratic types is private information, by the Revelation Principle the bank needs to impose an incentive-compatibility constraint: day consumption must be at least as large as night consumption, in order to induce truth-telling. In fact, if that was not the case, a day consumer would pretend to be a night consumer,

[6] Because of the Inada conditions, in equilibrium all other variables will be strictly positive.

withdraw from her bank account, and store until the following morning. The definition of the banking equilibrium is as follows:

Definition 1. *Given an initial amount of capital K_0, liquidity z_{-1}, liquidation D_{-1} and day consumption c_{10}, a no-run banking equilibrium is a bank portfolio strategy $\{\ell_t, k_{t+1}, z_t, D_t\}$ and a deposit contract $\{c_{2t}, c_{1t+1}\}$ for every $t = 0, 1, \ldots$ such that, for given prices:*

- *The deposit contract and the bank portfolio strategy solve the banking problem in* (2)*;*
- *The production inputs maximize firm profits;*
- *Resources are exhausted.*

We conclude the description of the environment by summarizing the timing of the events. In every period t: (i) during the morning, production takes place, and the firms produce $F(k_t)$; (ii) at noon, the bank pays the day consumption c_{1t}, and decides the terms of the portfolio strategy $\{\ell_t, k_{t+1}, z_t, D_t\}$ and the banking contract $\{c_{2t}, c_{1t+1}\}$; (iii) at night, the idiosyncratic shocks are revealed, and night consumption c_{2t} takes place according to the contract.

2.1 Discussion of the Assumptions

Several assumptions regarding this environment are worth commenting before the characterization of the equilibrium. The assumption of different life spans between the bank and its depositors is similar to He and Krishnamurthy (2013), and reflects the trade-off between the depositors' short-term incentives to withdraw and the long-term effects of these incentives on credit provision and capital accumulation. Moreover, assuming that the bank operates as a social planner is consistent with the characteristics of a competitive banking system with free entry, where the only banks who operate are those which maximize the expected welfare of their depositors.

The assumption of a loan liquidation value smaller than 1 was first introduced by Cooper and Ross (1998), and is based on some extensive empirical evidence (Altman et al., 2004). As an alternative to an exogenous loan liquidation value, we could introduce a secondary market for bank loans, as in Gertler and Kiyotaki (2015) and Segura and Suarez (2017). However, under this new assumption, we would obtain a different characterization of the banking equilibrium only if the loan liquidation value turned out to be larger than or equal to 1, in which case no run equilibrium would exist.

Three further features of the environment are worth highlighting. First, the economy is fully intermediated, in the sense that the bank channels all household savings into the production sector: in this way, we account for the large extent reached by financial intermediation in the world economy in the last decades (Beck et al., 2010). Second, we assume that the bank cannot suspend deposit convertibility, which would rule out runs, so that we can focus on the role of regulation in achieving the same goal. Third, there exists no deposit insurance. This assumption reflects the aggregate nature of the shock that we want to analyze, and finds its justification in the growing role played in modern financial systems by uninsured bank deposits and the shadow banking system, that offers bank services – and in particular liquidity and maturity transformation – without bank regulation or government assistance.[7]

2.2 The No-Run Banking Equilibrium

We start our analysis with the characterization of the no-run equilibrium, where bank runs are ruled out by assumption, which is the benchmark against which we compare the banking equilibrium with runs of the following section.[8] It is easy to show that the budget constraints all hold with equality, hence we can substitute (4) into (3) and get:

$$(1-\pi)c_{1t} + \pi c_{2t} + z_t - rD_t + k_{t+1} = F(k_t - D_{t-1}) + z_{t-1}. \tag{6}$$

We attach the multipliers $\beta^t \lambda_t$ to (6), $\beta^t \xi_t^k$ to (5), $\beta^t \xi_t^z$ to the non-negativity constraint $z_t \geq 0$, and $\beta^t \xi_t^D$ to the non-negativity constraint $D_t \geq 0$, respectively. The first-order conditions of

[7]In fact, uninsured deposits of FDIC-insured commercial banks and savings institutions grew from around 40 percent of total liabilities in the early Nineties to around 60 percent in 2016; in the same period, according to Bao et al. (2015), also "runnable liabilities" in the U.S. shadow banking system grew from around 40 to 60 percent of GDP. Source: own calculations based on FDIC bank balance-sheet data, and Bao et al. (2015). Runnable liabilities are the sum of uninsured deposits, money market mutual funds, repos, commercial papers, securities lending, federal funds borrowed, variable-rate demand obligations, and funding agreement backed securities.

[8]One way to rationalize the use of the no-run equilibrium as benchmark is to think that a central bank, in full control of the money supply, can create liquidity at (almost) zero costs, and therefore avoid bank runs altogether.

the program read:

$$c_{2t}: \quad u'(c_{2t}) = \lambda_t, \tag{7}$$

$$c_{1t+1}: \quad u'(c_{1t+1}) = \lambda_{t+1}, \tag{8}$$

$$k_{t+1}: \quad \lambda_t = \beta F'(k_{t+1} - D_t)\lambda_{t+1} + \xi_t^k, \tag{9}$$

$$z_t: \quad \lambda_t = \beta \lambda_{t+1} + \xi_t^z, \tag{10}$$

$$D_t: \quad r\lambda_t + \xi_t^D = \beta F'(k_{t+1} - D_t)\lambda_{t+1} + \xi_t^k. \tag{11}$$

In equilibrium, the marginal benefit from increasing the current night consumption c_{2t} and the future day consumption c_{1t+1} must be equal to their shadow values, i.e. the amount λ_t and λ_{t+1} by which such increases tighten the budget constraint in t and $t+1$. Since, at each point in time, the shadow values are the same for both day and night consumption, the bank provides perfect intratemporal risk sharing: the agents who find themselves in the condition of consuming at night receive the exact amount of consumption goods that they would have consumed during the morning of the same day, or $c_{2t} = c_{1t}$ for every t. Furthermore, the bank chooses the amount of loans in portfolio so as to equalize its marginal costs, in terms of a tighter budget constraint at time t, to its marginal benefits, in terms of a slacker budget constraint in $t+1$. Thus, in equilibrium, it allocates its portfolio between liquidity and loans in accordance with an Euler equation, so that the marginal rate of substitution between current night consumption and future day consumption (which is a measure of intertemporal risk sharing) is equal to the marginal rate of transformation of the production technology:

$$u'(c_{2t}) = \beta F'(k_{t+1} - D_t)u'(c_{1t+1}). \tag{12}$$

As the felicity function $u(c)$ is strictly concave, the contract satisfying the Euler equation is incentive-compatible if $\beta F'(k_{t+1} - D_t) \geq 1$. Moreover, as the discount factor β is smaller than 1, this also means that the marginal productivity of capital has to be strictly larger than 1: capital is a technology that provides a higher yield than liquidity, and is the only one that is employed to transfer resources from one period to the following. In other words, providing loans to the production sector always dominates the roll-over of excess liquidity and, in equilibrium, we have that $z_t = 0$ for every t. Yet, liquidity is the only technology employed to finance night consumption: in fact, from the first-order conditions with respect to k_{t+1} and D_t, it is easy to see that $\xi_t^D = (1-r)\lambda_t$, which is strictly positive and implies that $D_t = 0$ by complementary slackness. We summarize our findings in the following proposition:

Proposition 1. *Assume that $\beta F'(k_{t+1}) > 1$ for every t. The no-run banking equilibrium is characterized by:*

$$u'(c_{2t}) = u'(c_{1t}), \tag{13}$$

$$u'(c_{2t}) = \beta F'(k_{t+1}) u'(c_{1t+1}). \tag{14}$$

In equilibrium, the bank does not roll over any excess liquidity and does not liquidate loans, i.e. $z_t = D_t = 0$ for every $t = 0, 1, \ldots$.

The equilibrium allocation characterized in this proposition is observationally equivalent to that of a standard neoclassical growth model. This result highlights the fact that adding a microfounded banking system to a general equilibrium model is a meaningful exercise, that provides more insights than a standard growth model without a banking system, only to the extent that we introduce some financial distortion in the system, too. That is the topic of the incoming sections.

3 Self-Fulfilling Bank Runs

The assumption that the realizations of the idiosyncratic shocks are private information provides a rationale for the existence, in this environment, of a run equilibrium, where all agents withdraw c_{2t} at night, regardless of the actual realization of their idiosyncratic types. The run happens whenever all agents expect that every other agent is going to run, and know that the bank is not able to fulfill its contractual obligations with all of them. Crucially, we assume that the bank "anticipates" the run, in the sense that it modifies its investment strategy ex ante to accommodate for the possibility that the run equilibrium arises ex post. As a consequence, bank runs will affect the deposit contract, and the bank asset portfolio of liquidity and loans.

More formally, assume that, in the case of a run, the bank serves all the agents withdrawing at night according to an "equal service constraint", i.e. such that they get an equal share of the liquidation value of the whole bank portfolio. Then, the budget constraint at a run reads:

$$z_{t+1} + \pi c_{2t} + rD_t = c_{2t}^R. \tag{15}$$

This expression shows that the bank uses the available liquidity ℓ_t (the sum of what is set aside for night consumption πc_{2t} and the excess liquidity z_{t+1}) and the return from loan liquidation (which allows the bank to recover rD_t units of consumption for each D_t units of loans liquidated) to pay an amount of night consumption c_{2t}^R to all agents. Then, a run equilibrium exists if and only if the agents know that the bank does not hold a sufficient amount of liquid resources to pay the night consumption in the case of a run, or $c_{2t} > c_{2t}^R = z_{t+1} + \pi c_{2t} + rD_t$.

Whenever the banking problem exhibits a run equilibrium and a no-run equilibrium simultaneously, the depositors coordinate over which one to select in accordance with the realization of an extrinsic event s_t, called "sunspot". The sunspot takes the values 1 with probability q_t, in which case the agents choose to run, and 0 with probability $(1 - q_t)$, in which case they choose not to run.[9] The bank, in turns, knows the equilibrium-selection mechanism of the depositors, and adjusts ex ante the asset portfolio and banking contract, so as to maximize expected welfare. In doing so, the bank also indirectly affects the amount of consumption c_{2t}^R that the agents receive, in the case of a run. Thus, it can effectively choose whether a run equilibrium exists or not or, in other words, whether to be "run proof".

Before going into the details of the banking problem with runs, it is useful to recap the timing of actions. In every period t: (i) during the morning, production takes place, and the firms produce $F(k_t - D_{t-1})$; (ii) at noon, the bank pays the day consumption c_{1t}, and decides the terms of the portfolio strategy $\{\ell_t, k_{t+1}, z_t, D_t\}$, and the banking contract $\{c_{2t}, c_{2t}^R, c_{1t+1}\}$; (iii) at night, the idiosyncratic shocks are privately revealed, the agents decide whether to

[9] Accordingly, in what follows, we interchangeably refer to q_t as the probability of a run, or of the realization of the sunspot, or that the depositors coordinate over run.

run or not depending on the realization of the sunspot, and the night consumption c_{2t} takes place according to the contract.

To formally describe the banking problem with runs, first define the dummy variable $\mathcal{I}_t$, that takes the values 0 if the bank chooses to be run proof, i.e. if $c_{2t} \leq c_{2t}^R$, and 1 if it chooses a contract with possible runs. Moreover, define:

$$a_t \equiv F(k_t - s_{t-1}\mathcal{I}_{t-1}D_{t-1}) + (1 - s_{t-1}\mathcal{I}_{t-1})\big[z_{t-1} - (1-\pi)c_{1t}\big]. \tag{16}$$

This variable, together with s_{t-1} (i.e. the realization of the sunspot in the previous period), represents the state of the economy at date t, and summarizes the total available resources for investment. It depends on the asset portfolio and banking contract of the previous period, and on whether there has been a run or not: if in $t-1$ a run equilibrium was at the same time possible (i.e. $\mathcal{I}_{t-1} = 1$) and selected (i.e. $s_{t-1} = 1$), the bank would have $k_t - D_t$ resources invested; moreover, the bank would consume the excess liquidity z_{t-1} rolled over from the previous period during the run, but would not pay day consumption c_{1t} to the $(1-\pi)$ late consumers who ran.

Then, making use of the definition of liquidity, $\ell_t = z_{t+1} + \pi c_{2t}$, and of c_{2t}^R from (15), the banking problem can be written recursively as:

$$V(a_t, s_{t-1}) = \max_{\substack{\{c_{2t}, c_{1t+1},\\ k_{t+1}, z_t, D_t\}}} \Big[(1 - q_t\mathcal{I}_t)\big(\pi u(c_{2t}) + \beta(1-\pi)u(c_{1t+1})\big) + q_t\mathcal{I}_t\big[\pi + \beta(1-\pi)\big]u(c_{2t}^R)\Big] +$$

$$+ \beta\mathbb{E}_{s_t}[V(a_{t+1}, s_t)], \tag{17}$$

subject to the budget constraints:

$$\pi c_{2t} + k_{t+1} + z_t \leq a_t, \tag{18}$$

$$k_{t+1} \geq D_t, \tag{19}$$

and to the non-negativity constraints $D_t \geq 0$, $k_{t+1} \geq 0$ and $z_t \geq 0$. If the bank chooses the run-proof contract, its objective function turns into the one of the no-run banking problem. If instead it chooses the contract with possible runs, with probability $(1 - q_t)$ the agents select the no-run equilibrium, and their expected welfare can be written as before; with probability q_t, instead, the agents select the run-equilibrium and all get c_{2t}^R: at a run, π agents are night consumers, and consume right away, and $(1 - \pi)$ agents are day consumers, who store and consume in the morning of the following period. The definition of the equilibrium is as follows:

Definition 2. *Given a sequence of sunspots $\{s_t\}$, and an initial amount of capital K_0, liquidity z_{-1}, liquidation D_{-1} and day consumption c_{10}, a banking equilibrium with runs is a bank portfolio strategy $\{\ell_t,\ k_{t+1},\ z_t,\ D_t\}$, a deposit contract $\{c_{2t}, c_{1t+1}\}$, and a vector of bank's decisions over whether to be run proof or not $\{\mathcal{I}_t\}$ for every $t = 0, 1, \dots$ such that:*

- *The deposit contract, the bank portfolio strategy and the vector of bank's decisions over whether to be run proof or not solve the banking problem in* (17)*;*
- *The production inputs maximize firm profits;*
- *Resources are exhausted.*

In what follows, we study the equilibrium of an economy where the probability of the sunspot q_t can either be zero or a constant positive number q. We first characterize the run-proof contract and the contract with possible runs at time t separately, then analyze the optimal choice of the bank between the two.

3.1 The Run-Proof Contract

When offering a run-proof contract, the bank chooses to hold a sufficient amount of liquidity to pay all depositors in the case of a run, so that $c_{2t} \leq c_{2t}^R$. This condition is equivalent to imposing the run-proof constraint:

$$z_t + rD_t \geq (1 - \pi)c_{2t}, \tag{20}$$

which is a more stringent liquidity requirement than the one of the no-run equilibrium. Intuitively, this is the reason why, to make the contract run proof, the bank is forced to hold more liquidity than the amount necessary to pay night consumption. More formally, the bank solves the no-run banking problem in (2), subject to the liquidity requirement (20), the budget constraints and the non-negativity constraints on D_t, k_{t+1} and z_t. Notice that, despite the fact that the contract does not exhibit runs, the bank must solve for the amount D_t of loans to liquidate, in order to effectively rule them out. Attach the Lagrange multipliers $\beta^t \lambda_t(s_{t-1})$ to the budget constraint, $\{\beta^t \xi_t^D, \beta^t \xi_t^k, \beta^t \xi_t^z\}$ to the non-negativity constraints, and $\beta^t \chi_t$ to the run-proof constraint (20).[10] The first-order conditions of the program read:

$$
\begin{aligned}
c_{2t}: &\quad \pi u'(c_{2t}) = \pi \lambda_t(s_{t-1}) + (1-\pi)\chi_t, &(21)\\
c_{1t+1}: &\quad (1-\pi)u'(c_{1t+1}) = V_c'(a_{t+1}, 0), &(22)\\
k_{t+1}: &\quad \lambda_t(s_{t-1}) = \beta V_k(a_{t+1}, 0) + \xi_t^k, &(23)\\
z_t: &\quad \chi_t + \beta V_z(a_{t+1}, 0) + \xi_t^z = \lambda_t(s_{t-1}), &(24)\\
D_t: &\quad r\chi_t + \xi_t^D = \xi_t^k, &(25)
\end{aligned}
$$

where $V'(a_{t+1}, 0)$ is the derivative of future utility with respect to the control variables conditional on s_t being equal to zero, as the contract is run proof in period t. To characterize the contract, we start by guessing that the run-proof constraint is slack. By complementary slackness, this means that χ_t is equal to zero. Clearly, by (25), this implies that $\xi_t^k = \xi_t^D$. Moreover, from (23) and (24), we obtain:

$$
\beta(F'(k_{t+1}) - 1)u'(c_{1t+1}) = \xi_t^z - \xi_t^D, \tag{26}
$$

where we use the fact that $V_c'(a_{t+1}, 0) = \lambda_{t+1}(0)$ and $V_k'(a_{t+1}, 0) = F'(k_{t+1})\lambda_{t+1}(0)$ by the envelope condition. By the fact that $F'(k_{t+1}) - 1$ is positive, we get that $\xi_t^z > \xi_t^D$. Notice that it must be the case that $D_t > 0$, as it is a cheap way to ensure that the run-proof constraint is satisfied, without further tightening the budget constraint (as the run is off-equilibrium). Therefore, ξ_t^D must be equal to zero by complementary slackness, and ξ_t^z must be strictly

[10] To save on notation, we do not explicitly write that all multipliers depend on the states of the problem, except for the one on the budget constraint, which is crucial to track.

positive, implying that $z_t = 0$. Finally, (21) and (23) give:

$$c_{2t} = c_{1t}, \tag{27}$$

$$u'(c_{2t}) = \beta F'(k_{t+1}) u'(c_{1t+1}), \tag{28}$$

Put differently, whenever the run-proof constraint is slack, the run-proof contract is equivalent to the no-run contract. However, notice that this holds only if the loan liquidation value r is sufficiently large so that $rD_t > (1-\pi)c_{2t}$ holds. Otherwise, $z_t + rD_t = (1-\pi)c_{2t}$ and the run-proof contract is distorted with respect to the no-run equilibrium. To see that, notice that, as the Lagrange multiplier χ_t is strictly positive, by (25) it must be the case that ξ_t^k is also strictly positive in order to have $\xi_t^D \geq 0$. This further implies that the Euler equation is distorted towards the marginal utility of night consumption, or $u'(c_{2t}) > \beta F'(k_{t+1}) u'(c_{1t+1})$. Moreover, from (21), it must also be the case that $c_{2t} < c_{1t}$, by the strict concavity of the felicity function. In other words, in order to be run proof, the bank is forced to tighten credit, and rebalance its asset portfolio towards liquidity, while at the same time lowering the amount of night consumption offered: making the contract run proof comes at the cost of both less intratemporal and intertemporal risk sharing.

Proposition 2. *The run-proof contract is characterized by the following system of equations:*

$$\pi u'(c_{1t}) + (1-\pi)\chi_t = \pi u'(c_{2t}), \tag{29}$$

$$u'(c_{1t}) = \beta F'(k_{t+1}) u'(c_{1t+1}) + \xi_t^k, \tag{30}$$

$$u'(c_{1t}) = \beta u'(c_{1t+1}) + \xi_t^z + \chi_t, \tag{31}$$

$$\xi_t^k = r\chi_t + \xi_t^D. \tag{32}$$

If the loan liquidation value r is sufficiently large, the run-proof contract is equivalent to the no-run equilibrium.

3.2 The Contract with Possible Runs

When choosing a contract with possible runs, the bank takes into account that, depending on the realization of the sunspot, the depositors might all choose to withdraw at night, in which case it offers them an amount of consumption equal to $c_{2t}^R = \pi c_{2t} + z_t + rD_t$. More formally, using the aforementioned definition of c_{2t}^R, the banking problem reads:

$$V(a_t, s_{t-1}) = \max_{\substack{\{c_{2t}, c_{1t+1},\\ k_{t+1}, z_t, D_t\}}} \Big[(1-q)\big(\pi u(c_{2t}) + \beta(1-\pi)u(c_{1t+1})\big) + q\big[\pi + \beta(1-\pi)\big]u(c_{2t}^R)\Big] + \\ + \beta\mathbb{E}_{s_t}[V(a_{t+1}, s_t)], \tag{33}$$

subject to the budget constraints and to the non-negativity constraints. We attach to the constraints the Lagrange multipliers $\beta^t\lambda_t(s_{t-1})$, $\beta^t\xi_t^k$, $\beta^t\xi_t^D$ and $\beta^t\xi_t^z$ respectively, and derive the first-order conditions:

$$c_{2t}: \quad (1-q)u'(c_{2t}) + q[\pi + \beta(1-\pi)]u'(c_{2t}^R) = \lambda_t(s_{t-1}), \tag{34}$$

$$c_{1t+1}: \quad \beta(1-\pi)(1-q)u'(c_{1t+1}) + \beta\mathbb{E}_{s_t}[V_c'(a_{t+1}, s_t)] = 0, \tag{35}$$

$$k_{t+1}: \quad \beta\mathbb{E}_{s_t}[V_k'(a_{t+1}, s_t)] + \xi_t^k = \lambda_t(s_{t-1}), \tag{36}$$

$$z_t: \quad q[\pi + \beta(1-\pi)]u'(c_{2t}^R) + \beta\mathbb{E}_{s_t}[V_z'(a_{t+1}, s_t)] + \xi_t^z = \lambda_t(s_{t-1}), \tag{37}$$

$$D_t: \quad rq[\pi + \beta(1-\pi)]u'(c_{2t}^R) + \beta\mathbb{E}_{s_t}[V_D'(a_{t+1}, s_t)] + \xi_t^D = \xi_t^k. \tag{38}$$

Making use of the Lagrange multiplier $\lambda_t(s_{t-1}) = u'(c_{1t})$ and the envelope conditions:

$$V_c'(a_t, s_{t-1}) = -(1 - s_{t-1})(1-\pi)\lambda_t(s_{t-1}), \tag{39}$$

$$V_k'(a_t, s_{t-1}) = F'(k_t - s_{t-1}D_{t-1})\lambda_t(s_{t-1}), \tag{40}$$

$$V_z'(a_t, s_{t-1}) = (1 - s_{t-1})\lambda_t(s_{t-1}), \tag{41}$$

$$V_D'(a_t, s_{t-1}) = -s_{t-1}F'(k_t - s_{t-1}D_{t-1})\lambda_t(s_{t-1}), \tag{42}$$

we first derive the equilibrium condition:

$$(1-q)u'(c_{2t}) + q\big[\pi + \beta(1-\pi)\big]u'(c_{2t}^R) = u'(c_{1t}), \tag{43}$$

that characterizes the relationship between the night consumption c_{2t} and the current day consumption c_{1t}. The left-hand side of this expression is a linear combination of two positive terms, with $u'(c_{2t}) < u'(c_{2t}^R)$ by the strict concavity of the felicity function and the fact that $c_{2t} > c_{2t}^R$ in the contract with possible runs. Hence, we can characterize the following relation:

Lemma 1. *If β is sufficiently close to 1, the contract with possible runs exhibits $c_{2t} > c_{1t}$.*

In other words, if the depositors do not discount the future too much, the bank reacts to a positive probability of a run by offering a contract with possible runs with more intratemporal risk sharing than the one offered in the no-run equilibrium. The rationale for this result lies in the fact that, with equal service, there are two advantages of night consumption: (i) it provides higher welfare to night consumers, when all depositors coordinate over no-run, and (ii) it increases liquidity, that goes to all depositors, when instead they coordinate over run. This means that, for a given day consumption c_{1t}, the marginal utility of night consumption has to go down, with respect to what happens in the no-run equilibrium. Hence, night consumption has to increase.

To derive the intertemporal condition, we instead make use of the expressions in (35), (36) and (38), together with the corresponding envelope conditions, and get:

$$u'(c_{1t}) = (1-q)\beta F'(k_{t+1})u'(c_{1t+1}) + rq[\pi + \beta(1-\pi)]u'(c_{2t}^R) + \xi_t^D. \tag{44}$$

This is a distorted Euler equation, that equalizes the marginal costs of offering one more unit of loans, in terms of tightening the current budget constraint, to its marginal benefits, in terms of relaxation of the future budget constraint. If the depositors coordinate over no-run (with probability $(1-q)$) more loans make future consumption higher, as in the no-run equilibrium. If they instead coordinate over run (with probability q) the marginal benefit of offering one more unit of loans, in terms of relaxing the future budget constraint *and* holding more resources that can be liquidated (i.e. relaxing the constraint $k_{t+1} \geq D_t$) must be equal to the marginal cost of liquidating that one unit which, in turns, must be equal to the marginal benefits of liquidation. This is made further clear by the equilibrium condition for the optimal choice of D_t:

$$\beta q F'(k_{t+1} - D_t)\lambda_{t+1}(1) + \xi_t^k = rq[\pi + \beta(1-\pi)]u'(c_{2t}^R) + \xi_t^D, \tag{45}$$

where the marginal costs of liquidation, in terms of tightening the future budget constraint (the left-hand side of (45)), is equal to its marginal benefits (the right-hand side of (45)), in terms of extra consumption that the depositors can enjoy, if they coordinate over run. Finally, from the first-order condition with respect to excess liquidity z_t, we derive the equilibrium condition:

$$u'(c_{1t}) = q\big[\pi + \beta(1-\pi)\big]u'(c_{2t}^R) + \beta(1-q)u'(c_{1t+1}) + \xi_t^z, \tag{46}$$

where the marginal costs of holding excess liquidity, in terms of tightening the current budget constraint (the left-hand side of (46)), is equal to its marginal benefits (the right-hand side of (46)), in terms of the extra consumption that the depositors can enjoy, if they coordinate over run, and further relaxation of the budget constraint, if they coordinate over no-run.

4 Numerical Analysis

The last step missing in the characterization of the banking equilibrium with runs is the bank choice between the run-proof contract and the contract with possible runs. In order to compare them, we need to calculate the expected welfare of the depositors in both cases, which is impossible in closed-form solution. Hence, we rely on a numerical analysis of the equilibrium. To this end, we first calibrate the parameters of the model to the U.S. economy, and then run two separate experiments: first, we study how the model behaves in the presence of short-term financial fragility, i.e. when the probability of a run q is positive in only one period; then, we extend the analysis to the case of multi-period financial fragility, i.e. when q is positive in a series of consecutive periods before going back to zero. In both experiments, we evaluate the evolution of the economy on impact, as well as on its trajectory back to the steady state.

4.1 Parameter Calibration

We choose a standard Cobb-Douglas production function $Y_t = K_t^{\alpha}$, with $\alpha = 0.45$.[11] As mentioned before, we relax the hypothesis of full depreciation of capital from one period to the following, and find the depreciation rate d from the investment equation in steady state, i.e.:

$$1 = 1 - d + \frac{I_t}{K_t}. \tag{47}$$

With the investment-to-capital ratio I_t/K_t approximately equal to 0.076, as it is common in the literature, we get $d = 0.076$.

As far as the felicity function is concerned, we use a CRRA formulation. From the Euler equation of the no-run banking equilibrium in (12), we calibrate the value of the discount factor β:

$$1 = \beta(F'_K + 1 - \delta) = \beta\left(\alpha\frac{Y}{K} + 1 - d\right). \tag{48}$$

With the output-to-capital ratio Y/K approximately equal to 0.30, and with $d = 0.076$, we obtain that β is approximately equal to 0.94.

The last parameter left to calibrate is the probability π of the realization of the idiosyncratic shock θ_t, that makes the agents willing to consume at night. To this end, we take the equilibrium budget constraint of the no-run banking problem in (4), and divide both sides by GDP Y_t to get:

$$\frac{\ell_t}{Y_t} = \frac{\pi c_{2t}}{Y_t} = \frac{\pi c_{1t}}{Y_t} = \pi\frac{c_{1t}}{K_t}\frac{K_t}{Y_t}, \tag{49}$$

where we use the fact that, in the no-run banking equilibrium, $c_{2t} = c_{1t}$. From the resource constraint, we know that it must be the case that $c_{1t} = Y_t - I_t$. Hence, dividing both sides by K_t and plugging the result into (49), we obtain:

$$\pi = \frac{\frac{\ell_t}{Y_t}\frac{Y_t}{K_t}}{\frac{Y_t}{K_t} - \frac{I_t}{K_t}}. \tag{50}$$

[11] In all exercises, the assumption that $\beta F'(k_{t+1}) > 1$ is satisfied.

We set the ratio of liquid assets to GDP to 0.0146, equal to the average value for the U.S. financial businesses in the period 1990-2010, that we derive from the U.S. financial accounts.[12] From here, π comes out to be approximately equal to 0.02, which is close to the value in Gertler and Kiyotaki (2015) ($\pi = 0.03$).

Finally, as far as the realization of the sunspot q and the loan liquidation value r are concerned, we compare different solutions for q in the interval $[0, 0.06]$ and r in the interval $[0, 0.25]$. Our choice for the upper bound of the loan recovery rate is inconsequential for the characterization of the banking equilibrium, and is mostly consistent with the observed recovery rates of financial institutions hit by self-fulfilling runs during the 2007-2009 U.S. financial crisis.[13]

4.2 One-Period Runs

In order to understand the underlying mechanisms regulating the banking equilibrium, we start our analysis with the characterization of the two contracts, when the probability of a run is positive only at date $t = 0$, and then goes back to zero in the following periods.

Figure 1 shows an example economy, where the probability of a run is $q = 0.02$ and the loan liquidation value is high ($r = 0.25$). In accordance with the characterization of the previous section, in this case the loan liquidation value is so high that the bank is able to sustain a run-proof contract equivalent to the no-run equilibrium, hence the economy is unaffected by a positive probability of a run. With a contract with possible runs, instead, the top left panel of Figure 1 shows that credit would decrease (i) in anticipation of a possible run even if not realized (in the example, by around 8.7 percent), and (ii) if a run is actually realized (by around 15.3 percent). In the former case, the bank tightens credit in order to increase night consumption, and provide more intratemporal risk sharing against the possible realization of

[12] Liquid assets are the sum of checkable deposits and currency held by the U.S. financial businesses. Financial businesses include: finance companies; securities brokers and dealers; money market mutual funds; real estate investment trusts; insurance companies and pension funds; government-sponsored enterprises. Source: Flow of Funds of the United States. The assumption of a constant liquidity-to-GDP ratio is supported by the empirical observation that the time series exhibits no statistically significant trend, and a low standard deviation (0.007) around its mean.

[13] For example, according to some recent estimates, the average creditor of Lehman Brothers is supposed to recover around 18 percent of the face value of her claim (Hardy, 2013). Arguably, this is a pertinent example, as Lehman Brothers indeed suffered a self-fulfilling run, and was not covered by deposit insurance, as the bank in our environment.

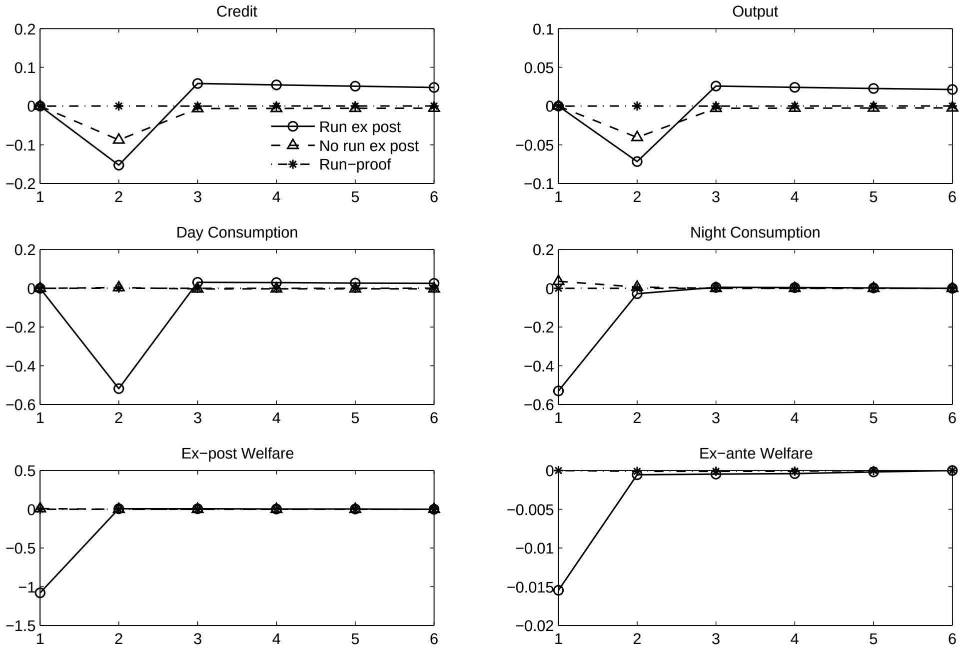

Figure 1: Impulse responses of the run-proof contract and of the contract with possible runs, with $q = 0.02$ and $r = 0.25$.

a run. In the latter case there is a further credit tightening to finance the consumption of the depositors running. However, in contrast to the results of any static bank-run model, in case of a run the bank does not liquidate all the loans in portfolio, but optimally addresses the trade-off between its short-term incentives, that urges it to provide the highest possible amount of consumption, and the provision of credit for the long-term growth of the real economy. Interestingly, the disruption in credit provision following the realization of a run leads to a drop in day and night consumption, which in turns allows a rebound in credit provision, that then slowly decays back to its steady state. The bottom left panel of Figure 2 shows that the drop in consumption leads to massive ex-post welfare costs if a run is realized, and negligible ones if it is not realized. Therefore, in expectations, the drop in welfare costs is small, of around 2 percent of the maximum ex-post costs.

Figure 2 reports the evolution of a similar economy, but where the loan liquidation value is lower ($r = 0.05$). In this case, the run-proof contract is distorted with respect to the no-run

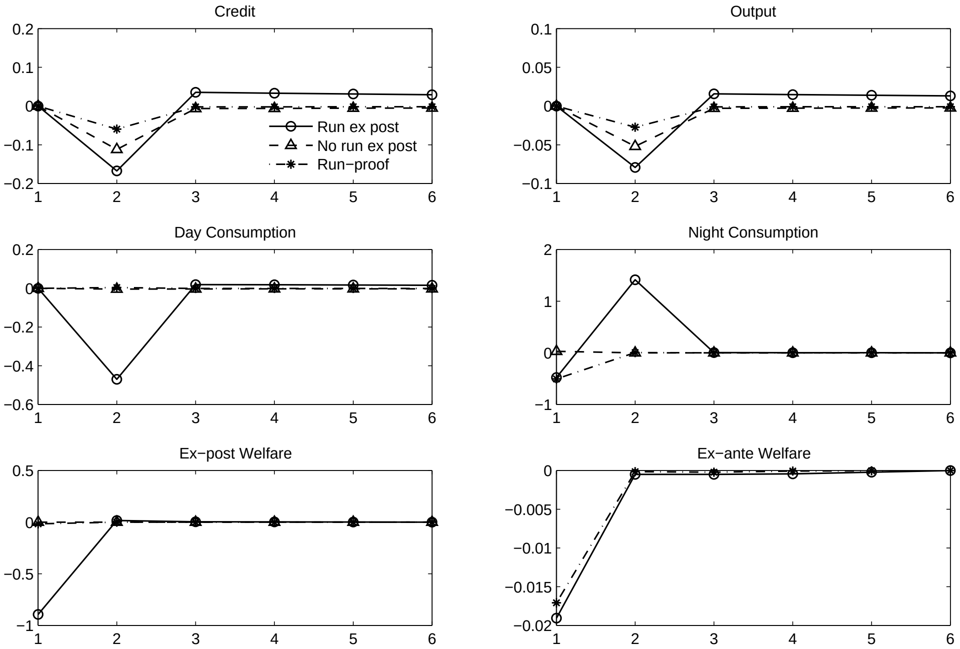

Figure 2: Impulse responses of the run-proof contract and of the contract with possible runs, with $q = 0.02$ and $r = 0.05$.

equilibrium, as the bank, in order to satisfy the run-proof constraint in the cheapest possible way, lowers the amount of night consumption and increases excess liquidity z_t. Moreover, a binding run-proof constraint distorts the Euler equation, and triggers a credit tightening of around 6 percent. All this tightens the bank budget constraint in $t + 1$, and leads to a drop in future day consumption, that is slowly reabsorbed. The resulting drop in expected welfare is sizeable, and can be higher than the one of the contract with possible runs.

More generally, Figure 3 shows that, in the contract with possible runs, the higher the probability of a run q, the higher the distortions induced by a run, hence the lower consumption and loans. In contrast, for given probability of a run q, the higher the loan liquidation value r, the lower the distortions induced by a run, hence the higher consumption and credit. Moreover, for given loan liquidation value r, an increase in the probability of a run q, by tightening credit, also tightens the amount that can be liquidated, and forces the bank to hold an increasing amount of excess liquidity. Finally, for given probability of a run q, an increase

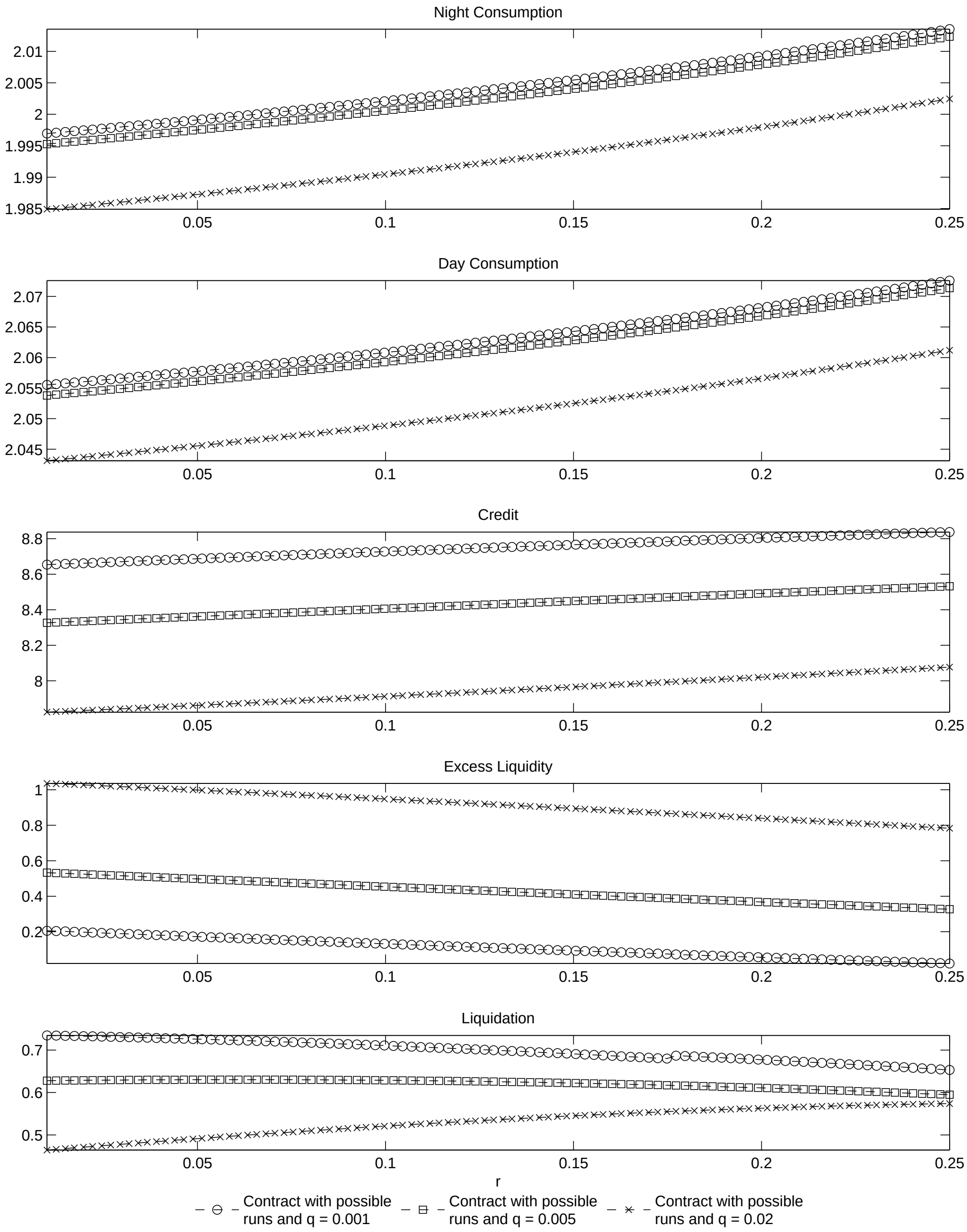

Figure 3: The impact of a positive probability of a run on the deposit contract and bank portfolio in the contract with possible runs, as a function of the loan liquidation value r (on the x-axis).

in the loan liquidation value r has the effect of lowering the need to hold excess liquidity, but has a non-trivial positive effect on the amount of loans that the bank liquidates. To see that, rearrange the equilibrium condition (45) by making use of the first-order condition with respect to k_{t+1}, and derive an expression for the Lagrange multiplier on the non-negativity constraint of D_t:

$$\xi_t^D = u'(c_{1t}) - \beta(1-q)F'(k_{t+1})u'(c_{1t+1}) - q\Big[\pi + \beta(1-\pi)\Big]ru'(c_{2t}^R). \tag{51}$$

From this expression, it is clear that a change in the loan liquidation value r has an effect on liquidation through two channels: on one side, a higher loan liquidation value lowers the distortions induced by a run (in particular, loans increase), and increases the marginal cost of liquidation, in terms of forgone return (the second part of the right-hand side of (51)); on the other, a higher loan liquidation value increases the direct advantage to rely on liquidation to finance night consumption in the case of a run (the third part of the right-hand side of (51)). The bottom panel of Figure 3 indeed shows that the first effect dominates when the probability of a run q is small, and the second dominates when q is high.

4.2.1 The Banking Equilibrium with Runs

With the two contracts in hand, we conclude the analysis of the banking equilibrium with runs by characterizing the choice between the run-proof contract and the contract with possible runs. To this end, we calculate the welfare costs of both contracts with respect to the no-run equilibrium, in terms of the constant percentage drop in consumption that would make the no-run equilibrium equivalent to each of them, and let the bank choose the contract with the lowest value.

As the preceding analysis of the equilibrium conditions suggests, the welfare costs of the run-proof contract are independent of the probability of a run q (as runs are completely ruled out), and decreasing in the loan liquidation value r, as the higher that is, the slacker the run-proof constraint also is. Additionally, as the theory predicts, for a loan liquidation value sufficiently high the run-proof contract is able to sustain an allocation equivalent to the no-run equilibrium, and the welfare costs go to zero. Figure 4 shows that the welfare

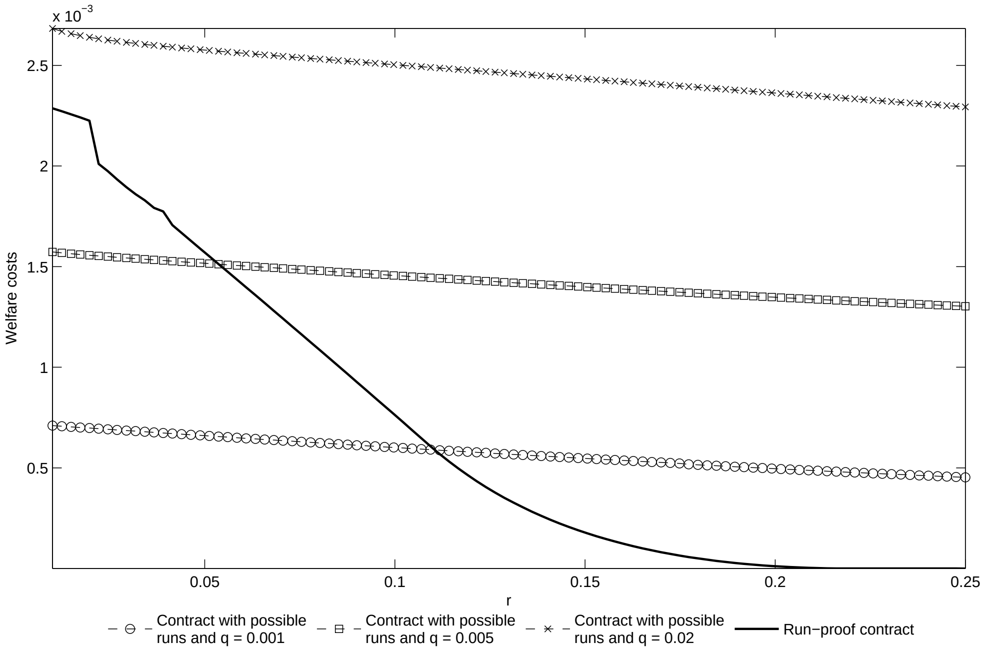

Figure 4: The welfare costs of the run-proof contract and of the contract with possible runs, for different values of the loan liquidation value r (on the x-axis).

costs of the contract with possible runs are instead increasing in the probability of a run q and, ceteris paribus, decreasing in the loan liquidation value r. In particular, for low values of both the loan liquidation value r and the probability of a run q, the welfare costs of the contract with possible runs are lower than those of the run-proof contract. The welfare costs of the run-proof contract decrease faster with r than those of the contract with possible runs. Hence, for every value of r, there exists a unique threshold q^* above which the bank prefers the run-proof contract. As the probability of a run q increases and the welfare costs of the run-proof contract remain unchanged, the welfare costs of the contract with possible runs increase. Therefore, the threshold q^* is decreasing in the loan liquidation value r, up to the point at which the bank chooses to be run proof regardless of the probability of a run q: put differently, q^* tends to zero as the loan liquidation value r increases.

Figure 5 provides the graphical representation of the contractual choice in the banking equilibrium with one-period runs: there exists a small region of the parameter space where

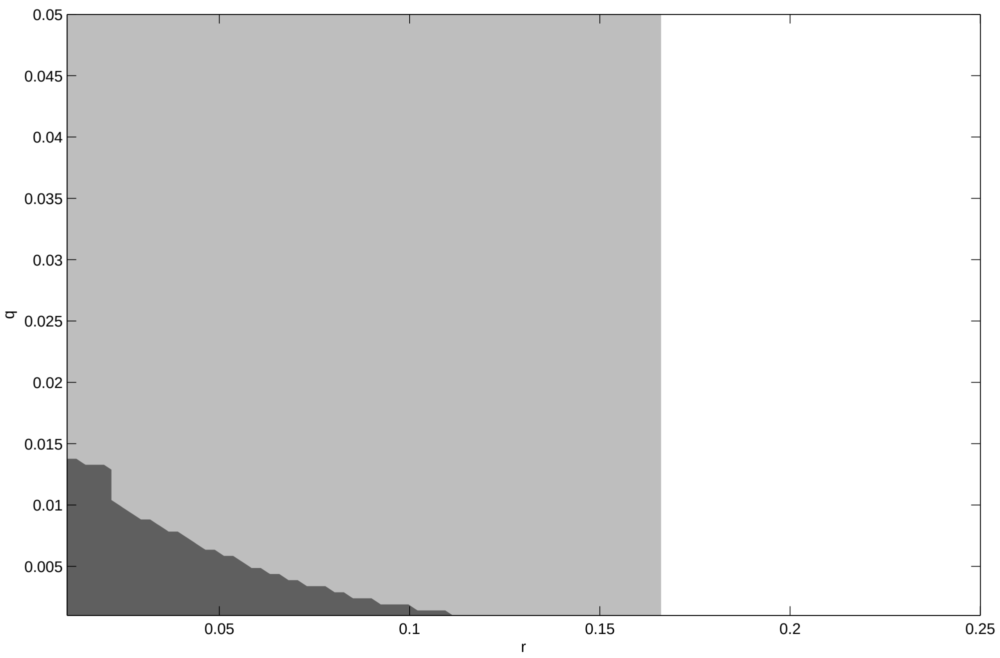

Figure 5: The contractual choice in the banking equilibrium with one-period runs, for different values of the loan liquidation value r (on the x-axis) and of the probability of a run q (on the y-axis). The dark grey area corresponds to the parameter space where the bank chooses the contract with possible runs, the light grey area to the distorted run-proof contract, and the white area to the no-run equilibrium.

the bank prefers a contract with possible runs to a run-proof contract. Moreover, for a loan liquidation value higher than around 0.17, the bank is able to offer a run-proof contract equivalent to the no-run equilibrium, and a positive probability of a run affects neither the portfolio allocation nor the banking contract. In between, instead, the bank chooses the run-proof contract, but its allocation is distorted with respect to the no-run equilibrium. This result concludes the characterization of the banking equilibrium with one-period runs, and allows us to quantify the distortion that this type of financial fragility brings about. In fact, the welfare costs of the banking equilibrium with runs, resulting from its comparison to the no-run equilibrium, come out to be the highest whenever the loan liquidation value is the lowest, and the probability of a run is so high that the bank chooses the run-proof contract. This combination of parameters is associated with welfare costs of around 0.23 percentage points

Table 1: The threshold q^*, comparative statics

	$r=0.01$	$r=0.02$	$r=0.05$	$r=0.1$	$r=0.2$	$r=0.3$
$\sigma=2;\ \pi=0.02$	0.0134	0.0129	0.0054	0.0015	0.0000	0.0000
$\sigma=2;\ \pi=0.04$	0.0173	0.0167	0.0100	0.0019	0.0000	0.0000
$\sigma=3;\ \pi=0.02$	0.0119	0.0112	0.0042	0.0000	0.0000	0.0000

in units of consumption equivalents, roughly equivalent to a constant yearly consumption loss of US$27 Billion per year, or 0.16 percent of U.S. GDP, at 2014 levels.

4.2.2 Robustness Checks

We conclude the characterization of the banking equilibrium with runs by checking the robustness of the results to changes in some parameters, namely the probability of the realization of the idiosyncratic shock and the coefficient of relative risk aversion. As far as the first one is concerned, we follow an alternative calibration strategy. Telyukova and Visschers (2013), using CEX data, estimate the standard deviation of the idiosyncratic component of the household consumption of "cash goods", defined as goods purchased only with liquid assets. We calibrate π so as to match their most conservative estimate (st.dev=0.169), and find a value of around 0.04, in the same order of magnitude of our original choice. The second row of Table 1 shows that our results are qualitatively unchanged, and quantitatively robust: in equilibrium, the bank still chooses the run-proof contract for values of the probability of a run q higher than a threshold q^*, which is, as before, decreasing in the loan liquidation value r. Similarly, in the last row of Table 1, we report the thresholds q^* for a coefficient of relative risk aversion equal to 3, and find our conclusions qualitatively unaltered: for given relative risk aversion, q^* is relatively low and decreasing in the loan liquidation value. Ceteris paribus, the higher relative risk aversion is, the higher the distortion of the contract with possible runs with respect to the no-run equilibrium, and the lower the threshold q^* at which the bank prefers to switch to a run-proof contract. This is a consequence of the fact that the more risk averse the depositors are, the less they tolerate any difference in their ex-post consumption profiles, hence the more the bank distorts the economy when the probability of a run is positive.

4.2.3 The Welfare Costs of Liquidity Requirements

Based on the previous results, we now assess the effect on welfare of a regulatory intervention aimed at making the economy immune from financial fragility. To this end, we focus on the welfare costs of imposing a run-proof liquidity requirement, equal to the run-proof constraint (20), regardless of the loan liquidation value r and of the probability of a run q. The main results of the banking equilibrium with runs are that, for given loan liquidation value r, there exists a threshold q^* for the probability of a run, below which the bank prefers a contract with possible runs to a run-proof contract, and that this threshold is decreasing with the loan liquidation value r itself. Therefore, imposing a run-proof liquidity requirement can generate no positive welfare gain with respect to the banking equilibrium with runs. Conversely, a run-proof liquidity requirement can generate welfare costs, from forcing the bank to be run proof in that region of the parameter space where it would rather let runs happen with some positive probability. Thus, the highest distortion of imposing a run-proof liquidity requirement arises when the probability of a run q and the loan liquidation value r are both at their lowest levels. That leads to a increase in welfare costs, with respect to the banking equilibrium with runs, of around 0.16 percent in units of consumption equivalents in the baseline calibration, or of 0.20 percent with the higher probability of the idiosyncratic shock ($\pi = 0.04$), or of 0.15 percent with a coefficient of relative risk aversion equal to 3. These numbers roughly amount to a yearly costs (in terms of 2014 consumption expenditure) of between US$18 and 23 Billion per year, or of between 0.11 and 0.14 percent of U.S. GDP, at 2014 levels.

4.2.4 Equilibrium with Narrow Banking

In the previous sections, we characterized the banking equilibrium with runs in an economy where a bank could provide consumption in the case of a run both with liquidity and by liquidating loans. Yet, in the recent years there has been a revamped interest in the concept of "narrow banking", according to which all deposits should be backed by extremely liquid resources, such as currency or central bank reserves, in order to guarantee the stability of the system itself. In fact, this idea was first proposed in the late Thirties, as part of the so-called "Chicago Plan", with the intended aim to gain a better control of the credit cycle, reduce private debt, and eliminate bank runs (Fisher, 1936).

To analyze the welfare costs of narrow banking, we characterize the run-proof contract in an environment where the bank has to satisfy the "narrow-banking" constraint:

$$z_t \geq (1-\pi)c_{2t}. \tag{52}$$

This expression states that, in equilibrium, the bank has to be run proof while at the same time avoiding loan liquidation, by holding sufficient excess liquidity to cover at any point in time for the demand of night consumption of the $(1-\pi)$ day consumers who might run. More formally, the bank solves the banking problem in (17) (with $D_t = 0$ for every t) subject to the budget constraints, the non-negativity constraint $z_t \geq 0$, and the narrow-banking constraint (52). Attach to these constraints the Lagrange multipliers $\beta\lambda_t(s_{t-1})$, $\beta\xi_t^z$ and $\beta\chi_t$, respectively. The first-order conditions of the program read:

$$c_{2t}: \quad \pi u'(c_{2t}) = \pi\lambda_t(s_{t-1}) + (1-\pi)\chi_t, \tag{53}$$

$$c_{1t+1}: \quad u'(c_{1t+1}) = \lambda_{t+1}(s_t), \tag{54}$$

$$k_{t+1}: \quad \lambda_t(s_{t-1}) = \beta F'(k_{t+1})\lambda_{t+1}(s_t), \tag{55}$$

$$z_t: \quad \chi_t + \beta\lambda_{t+1}(s_t) + \xi_t^z = \lambda_t(s_{t-1}). \tag{56}$$

Clearly, as in equilibrium c_{2t} has to be positive in order to satisfy the Inada conditions, the narrow-banking constraint ensures that also excess liquidity z_t has to be positive in equilibrium, therefore ξ_t^z is equal to zero, by complementary slackness. This, together with the first-order conditions with respect to k_{t+1} and z_t, allows us to characterize the equilibrium value of the Lagrange multiplier $\chi_t = \beta(F'(k_{t+1}) - 1)\lambda_{t+1}(s_t)$, which is strictly larger than zero, as $F'(k_{t+1}) > 1$ and $\lambda_{t+1}(s_t)$ has to be positive in equilibrium to satisfy the Inada conditions. Hence, the narrow-banking constraint is binding in equilibrium: the bank holds excess liquidity in an amount which is exactly sufficient to pay night consumption to the day consumers who might run. This marks a crucial difference between the equilibrium with narrow banking and the banking equilibrium with runs of the previous sections: there is no

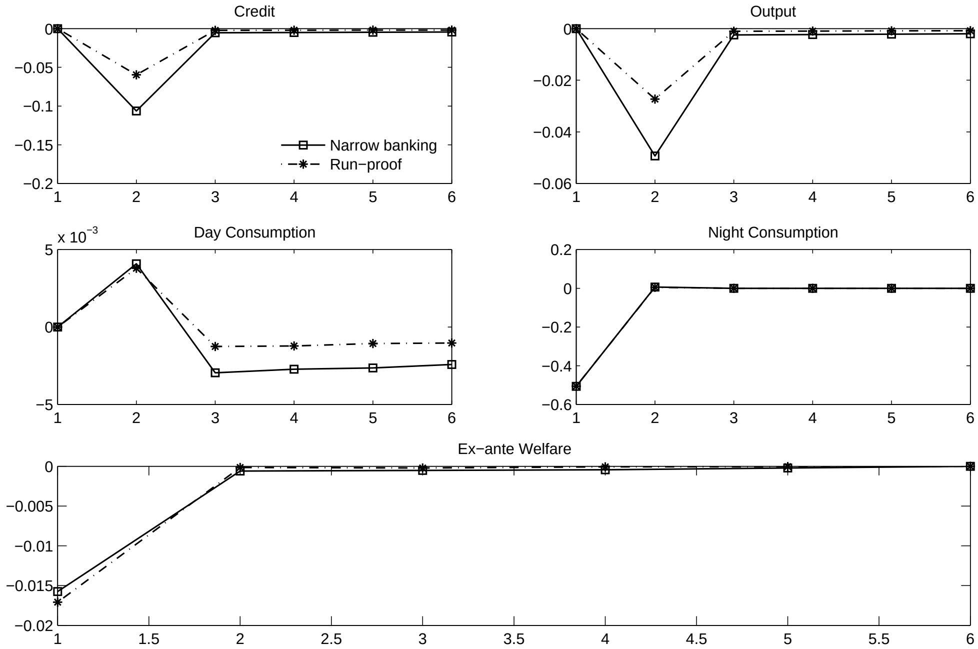

Figure 6: Impulse responses of the banking equilibrium with narrow banking and of the run-proof contract, with $q = .02$ and $r = .05$.

way in which a bank that satisfies the narrow-banking constraint can sustain an allocation equivalent to the no-run equilibrium. Eventually, the excess liquidity that the bank is forced to hold in period t is rolled over to period $t + 1$, relaxing the future budget constraint. The distorted Euler equation, characterizing the equilibrium amount of intertemporal risk sharing, accounts for that:

$$\pi u'(c_{2t}) + (1 - \pi)\beta u'(c_{1t+1}) = \beta F'(k_{t+1})u'(c_{1t+1}). \tag{57}$$

This expression makes clear that, in equilibrium, the bank chooses a portfolio allocation such that the marginal benefits of holding liquidity, in terms of current consumption and relaxation of the future budget constraint through excess liquidity, is equal to the marginal benefits of issuing loans. Moreover, the equilibrium allocation satisfies the incentive compatibility constraint: as $\beta F'(k_{t+1}) > 1$, the distorted Euler equation implies that $\pi u'(c_{2t}) + (1 - \pi)\beta u'(c_{1t+1}) > u'(c_{1t+1})$, hence $c_{2t} < c_{1t+1}$, by concavity of the felicity function.

The plots in Figure 6 compare the asset portfolio and the deposit contract of a typical equilibrium with narrow banking to the corresponding run-proof contract. It is clear that forcing the bank to be run proof *and* to avoid liquidation exacerbates its reaction to financial fragility, as it leads to a credit tightening almost double the size of that triggered by a run-proof contract (6 vs 11 percent). As before, we can also calculate the welfare costs of imposing on the economy a narrow-banking liquidity requirement equivalent to the narrow-banking constraint (52), irrespective of the probability of a run q and of the loan liquidation value r. To this end, we again compare the welfare costs of the banking equilibrium with runs to the welfare costs of the banking equilibrium with narrow banking. Differently from the case of imposing a run-proof liquidity requirement, the highest increase in welfare costs, compared to the banking equilibrium with runs, now arise whenever the loan liquidation value r is so high that the bank would be able to sustain a run-proof contract equivalent to the no-run equilibrium, while, with narrow banking, it is forced to distort the equilibrium allocation. For our main parameter choice, this generates an increase in welfare costs of around 0.19 percent in units of consumption equivalents, equivalent to a drop in real consumption of around US$22 billion per year, or 0.13 percent of U.S. GDP, at 2014 levels.

4.3 Multi-Period Runs

In this final section, we complete our analysis with the characterization of the banking equilibrium when the probability of a run can stay positive for more than one period. To this end, we assume that, as in the model of the previous section, the economy starts with a period of financial fragility, i.e. with a positive probability of a run $q_0 = q > 0$, and make two key extensions. First, in every subsequent period $t > 0$, we introduce a positive probability $P(q_{t+1} = 0 \mid q_t = q)$ that the probability of a run q_{t+1} goes to zero, i.e. that the economy exits financial fragility and starts its transition back to the no-run steady state, and a complementary probability $P(q_{t+1} = q \mid q_t = q) = 1 - P(q_{t+1} = 0 \mid q_t = q)$ that it stays positive at $q_{t+1} = q$. We calibrate this probability to 0.3125, to match the average duration of a financial crisis according to Reinhart and Rogoff (2014). Second, we rewrite the per-period utility function as follows:

$$W(c_{2t}, c_{1t+1}) = (1 - q\mathcal{I}_t)\big(\pi u(c_{2t}) + \beta(1-\pi)u(c_{1t+1})\big) + q\mathcal{I}_t\big[\pi + \beta(1-\pi)\big]\big[u(c_{2t}^R) + \epsilon_t\big]. \tag{58}$$

In it, ϵ_t is a publicly observable shock to the utility of the depositors in the case of a run, and is i.i.d. with a continuous distribution function $F(\epsilon_t)$. Adding this type of shock is common practice in the sovereign default literature after the contribution of Chatterjee and Eyigungor (2012), as it ensures the continuity of the value function and the convergence of the solution algorithm, and can be interpreted as a shock to depositors' tastes or an institutional shock to the economy during a financial crisis.

The bank observes both q_t and ϵ_t before choosing the asset portfolio and deposit contract, and forms expectations on the future duration of financial fragility and the utility that the depositors can enjoy at a run. The value function of the banking problem incorporates these expectations: with probability $P(q_{t+1} = 0 \mid q_t = q)$ the economy will exit financial fragility in the following period, and with probability $1 - P(q_{t+1} = 0 \mid q_t = q)$ financial fragility will continue. More formally, the value function reads:

$$\begin{aligned} V(a_t, s_{t-1} \mid q_t = q) = \max_{\{c_{2t}, c_{1t+1}, k_{t+1}, z_t, D_t\}} \; & W(c_{2t}, c_{1t+1}) + \\ & + \beta\Big[P(q_{t+1} = 0 \mid q_t = q)V(a_{t+1}, s_t \mid q_{t+1} = 0) + \\ & + (1 - P(q_{t+1} = 0 \mid q_t = q))\mathbb{E}_\epsilon[V(a_{t+1}, s_t \mid q_{t+1} = q)]\Big], \end{aligned} \tag{59}$$

where $\mathbb{E}_\epsilon[\cdot]$ is the expectation with respect to the shock ϵ_{t+1}, and the value function when $q_t = 0$ is simply the no-run value function:

$$V(a_t, s_{t-1} \mid q_t = 0) = \max_{\{c_{2t}, c_{1t+1}, k_{t+1}, z_t, D_t\}} W(c_{2t}, c_{1t+1}) + \beta V(a_{t+1}, 0 \mid q_{t+1} = 0). \tag{60}$$

Recall that, in every period of financial fragility, the bank takes a discrete choice between the run-proof contract and the contract with possible runs. The dummy variable $\mathcal{I}_t$ takes the values 0 if the bank chooses to be run proof at date t and 1 if it chooses a contract with possible runs. Consequently, the bank chooses the run-proof contract at date t if:

$$V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 0) \geq V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 1) + q_t\big[\pi + \beta(1-\pi)\big]\epsilon_t. \tag{61}$$

We can denote $\tilde{\epsilon_t} = q_t\big[\pi + \beta(1-\pi)\big]\epsilon_t$, and rearrange the previous expression in:

$$\tilde{\epsilon_t} \leq V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 0) - V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 1) \equiv V_t^*. \tag{62}$$

This implies that the probability of the bank choosing the run-proof contract at any period t is $\tilde{F}(V_t^*)$, and that we can rewrite the expected value function as:

$$\begin{aligned}\mathbb{E}_\epsilon[V(a_t, s_{t-1} \mid q_t = q)] =& \tilde{F}(V_t^*)\, V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 0) + \\ &+ (1 - \tilde{F}(V_t^*)) V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 1). \end{aligned} \tag{63}$$

The introduction of the shock ϵ_t is equivalent to introducing randomization in the problem. In the discrete choice literature, the distribution of the shock is either normal distribution or the generalized extreme value (GEV) distribution (Arcidiacono and Ellickson, 2011). We choose the latter because this distribution allows us to write a closed-form expression for the probability at date t of choosing the run-proof contract, as a function of the two alternative value functions (McFadden, 1978):

$$\tilde{F}(V_t^*) = \frac{1}{1 + \exp\Big(V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 0) - V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 1)\Big)}. \tag{64}$$

To sum up, in order to characterize the banking equilibrium with runs, we need to calculate the expected value functions of the no run equilibrium in (60) as well as of the run equilibrium in (63) at every point in time. The former is straightforward, and we solve it by a standard value function iteration. As far as the latter is concerned, we instead iterate over a grid of the three-dimensional state space $\{k_t, c_{1t}, z_{t-1}\}$, in order to calculate a_t according to the algorithm described in Appendix A. Then, with the expected value functions in hand, we simulate the economy: first, starting with financial fragility in the first period (i.e. $q_0 = q > 0$), we use the probability $P(q_{t+1} = 0 \mid q_t = q)$ to draw the duration of financial fragility (i.e. the number of periods with $q_t = q$); second, in every period of financial fragility, we use q to draw the occurrence of a run. For different combinations of the probability of a run q and loan liquidation value r, we simulate the economy 10,000 times over 40 periods. We solve for the banking equilibrium with runs and for the two regulated equilibria, where the bank is forced to satisfy either the run-proof constraint (20) or the narrow-banking constraint (52) in every period.

Figure 7 shows the average evolution of the same example economy of Figure 2, where the probability of a run and the loan liquidation value are $q = 0.02$ and $r = 0.05$, respectively. In this case, the economy suffers a credit tightening on impact of around 6 percentage points, leading to a GDP drop on impact of almost 3 percentage points. In addition to that, financial fragility is long-lasting (its average half-life is of around 4 periods), and this leads to an equally long-lasting low-credit recovery, before the economy exits financial fragility and starts its transition path towards the steady state. Financial fragility also leads to a drop in both night and day consumption, implying a welfare drop on impact of around 9 percentage points, and again a long-lasting recovery.

Table 2 reports the share of periods of financial fragility in which the bank independently chooses to be run proof, and shows that it is weakly increasing in both the probability of a run q and the loan liquidation value r. Moreover, these numbers confirm and generalize the result of the previous section: similarly to the economy with one-period runs, there exists a threshold probability of a run q^* at which the bank independently chooses to be fully run proof, and this threshold is decreasing in the loan liquidation value r.

Table 3 reports the welfare costs of the banking equilibrium with runs and of the two regulated equilibria. As in the previous section, we calculate the welfare cost in terms of the permanent increase in consumption needed to provide to the depositors the same present discounted value of welfare of the no-run banking equilibrium. An increase in the probability of a run q affects the welfare costs of the banking equilibrium with runs in two ways: on one hand, runs are more likely to happen, and the resulting credit tightening and loan liquidation increase the welfare costs; on the other hand, as shown in Table 2, the bank is also more

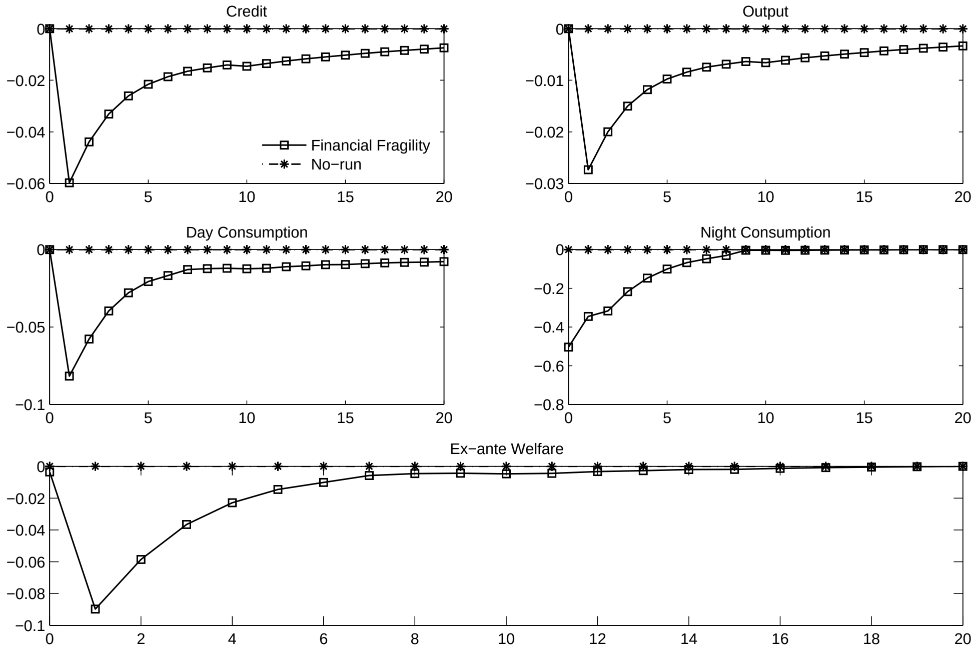

Figure 7: Impulse responses of the banking equilibrium with multi-period runs, with $q = 0.02$ and $r = 0.05$.

likely to be run proof. As long as the bank chooses the contract with possible runs with some positive – albeit small – probability, which is true for low values of the probability of a run q and loan liquidation value r, the rare realization of a run increases the welfare costs, and more so the higher q and r are. For some combinations of parameters, this can result in banking equilibria with higher welfare costs than the corresponding regulated equilibria. From the results in Table 3, we see that the welfare costs of the banking equilibrium with runs are weakly increasing in both the probability of a run q and the loan liquidation value r, go to zero for a sufficiently high loan liquidation value, and reach their peak at around 3.6 percentage points in units of consumption equivalents. This number is equivalent to a constant yearly consumption loss of around US$415 Billion, or around 2.5 percent of U.S. GDP, at 2014 levels.

Finally, we provide a definitive answer to our original question: what are the welfare costs of a liquidity requirement aimed at making the economy immune from financial fragility?

Table 2: Share of run-proof periods in the banking equilibrium with multi-period runs (% of number of periods with positive q)

	$r = 0.01$	$r = 0.03$	$r = 0.05$	$r = 0.07$	$r = 0.1$	$r = 0.25$
$q = 0.06$	99.5719	99.8428	99.8428	99.9115	99.9766	100.000
$q = 0.04$	99.5286	99.7353	99.8085	99.9115	99.9766	100.000
$q = 0.02$	78.2107	85.3622	99.8428	95.3887	99.9766	100.000
$q = 0.01$	75.8462	76.4005	93.7073	88.6384	99.8373	100.000
$q = 0.005$	55.7366	60.1632	90.7725	77.5986	99.9012	100.000
$q = 0.001$	3.3396	24.4957	28.2333	28.9919	74.6019	100.000

Again, we define the welfare costs as the distortion, in terms of extra welfare costs that would arise in the banking equilibrium with runs, if we imposed either the run-proof liquidity requirement (20) or the narrow-banking liquidity requirement (52) in every period, regardless of the probability of a run q and of the loan liquidation value r. As suggested by the theory, the welfare costs of a run-proof liquidity requirement are independent of the probability of a run q and decreasing in the loan liquidation value r, with an upper bound at around 3.4 percentage points; those of a narrow-banking liquidity requirement are instead independent of both the probability of a run q and of the loan liquidation value r and slightly higher, at almost 3.5 percentage points.

Table 3 shows that, differently from the case of one-period runs, for some combinations of parameters the imposition of a run-proof or a narrow-banking liquidity requirement can indeed bring about some welfare gains, with respect to the banking equilibrium with runs: that happens because, for intermediate values of the probability of a run q, the bank does not find convenient to be completely run-proof, but runs still realizes (albeit in extremely rare occasions). Quantitatively, this means that imposing a run-proof or a narrow-banking liquidity requirement can generate welfare gains of up to around 0.22 and 0.10 percentage points in units of consumption equivalents, respectively. Conversely, for most combinations of parameters, the imposition of a run-proof or a narrow-banking liquidity requirement leads to an increase in welfare costs. In fact, the run-proof liquidity requirement forces the bank to be run proof in that region of the parameter space where it would rather let runs happen with some positive probability. Thus, the highest increase in welfare costs, with respect to the banking equilibrium with runs, arises when the probability of a run q and the loan liquidation value r are both at

Table 3: The welfare costs of the banking equilibrium with multi-period runs (% of consumption equivalents)

	$r = 0.01$	$r = 0.03$	$r = 0.05$	$r = 0.07$	$r = 0.10$	$r = 0.25$
$q = 0.06$	3.4287	2.6482	1.9411	1.3082	0.3660	0.0000
$q = 0.04$	3.4105	2.6527	1.9370	1.2051	0.3651	0.0000
$q = 0.02$	3.6033	2.9533	1.9271	1.5027	0.3642	0.0000
$q = 0.01$	3.2267	2.6471	1.9857	1.4874	0.3697	0.0000
$q = 0.005$	2.5117	2.1895	1.8861	1.3748	0.3646	0.0000
$q = 0.001$	0.8285	1.1564	0.9988	0.7584	0.3739	0.0000
Run proof	3.3761	2.6275	1.9201	1.2944	0.3616	0.0000
Narrow banking	3.4973	3.4973	3.4973	3.4973	3.4973	3.4973

their lowest level, and is of around 2.5 percentage points in units of consumption equivalents. This number amounts to a cost (in terms of 2014 real consumption expenditure) of almost US$300 Billion per year, or of around 1.8 percent of U.S. GDP, at 2014 levels. Imposing the tighter narrow-banking liquidity requirement, instead, would increase the welfare costs whenever the loan liquidation value r is so high that the bank is able to implement the no-run equilibrium, and the regulation forces it to be extremely liquid. Therefore, the corresponding maximum increase in welfare costs, with respect to the banking equilibrium with runs, is again of around 3.5 percentage points in units of consumption equivalents, corresponding to a constant yearly drop in consumption of around US$405 Billion, or 2.4 percent of U.S. GDP, at 2014 levels.

5 Concluding Remarks

The present paper contributes to the literature on the economics of banking and financial fragility by studying the welfare implications of self-fulfilling bank runs, and of the liquidity requirements needed to offset them, in a neoclassical growth model with a fully microfounded banking system and multiple equilibria. The first takeaway of the analysis is that self-fulfilling runs might lead to a credit tightening, and a resulting drop in GDP, due to the bank anticipation of them *as well as* to their actual realization. These effects can be very persistent, but the banking system can avoid them by independently choosing to be run proof, at the cost of lower intratemporal and intertemporal risk sharing. The banks' choice of whether to be run proof crucially depends on the fundamentals of the economy, and highlights that self-fulfilling runs are not an inevitable by-product of all competitive banking systems: indeed, it is competition among banks that provides the correct incentives for them to avoid risky investment strategies, that might harm depositors' savings and give rise to financial fragility.

Our second takeaway is that self-fulfilling bank runs are costly. In fact, their potential welfare costs are neither as low as the welfare costs of business cycles (Lucas, 1987) or inflation (Lucas, 2000), nor as high as those of rare macroeconomic disasters (Barro, 2009) or long-run risk (Epstein et al., 2014), but are comparable to some recent estimates of the welfare costs of large unemployment shocks, like the Great Depression (Chatterjee and Corbae, 2007).

Our third takeaway is that there exists some space for liquidity requirements to make the banking system more resilient to self-fulfilling runs. However, the increase in welfare costs that might arise whenever the banks are forced to hold unnecessary high amounts of liquidity might be substantial, and higher than some recent estimates of the welfare costs of capital and liquidity requirements (Van den Heuvel, 2008, 2016). In particular, our results highlight how important the microfoundations of the banking system are, in order to properly quantify these costs. Moreover, they confirm the danger of imposing a 100-percent liquidity requirement that forces the banks to be narrow, as that would bring about the highest welfare costs exactly when the banking system could avoid self-fulfilling runs at zero costs. Moreover, any other feature of the economy that is not part of the present environment, and that would aggravate the consequences of a run, would force the banks, because of perfect competition, to be *more* run proof, thus lowering the welfare costs of imposing a run-proof liquidity requirement. In contrast, the welfare costs of narrow banking would be unaffected, thus further highlighting the dangerousness of such a regulatory intervention.

Finally, it is worth noticing that in our work the depositors' behavior is consistent with the features of the environment, but still depends on a shift of their expectations that is left unexplained. In other words, a dynamic theory of the joint evolution of depositors' expectations and the probability of a run is missing. One way to address this point would be via the introduction of a "global game" among the depositors, like in the static environment of Morris and Shin (1998) and Goldstein and Pauzner (2005). We leave the analysis of this issue for future research.

References

Allen, F. and D. Gale (1997, June). Financial Markets, Intermediaries, and Intertemporal Smoothing. *Journal of Political Economy 105*(3), 523–546.

Allen, F. and D. Gale (1998, August). Optimal Financial Crises. *Journal of Finance 53*(4), 1245–1284.

Altman, E., A. Resti, and A. Sironi (2004, July). Default Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence. *Economic Notes 33*(2), 183–208.

Arcidiacono, P. and P. B. Ellickson (2011, September). Practical methods for Estimation of Dynamic Discrete Choice Models. *The Annual Review of Economics 3*, 363–394.

Bao, J., J. David, and S. Han (2015, September). The Runnables. FEDS Notes 2015-09-03. Board of Governors of the Federal Reserve System.

Barro, R. J. (2009, March). Rare Disasters, Asset Prices, and Welfare Costs. *American Economic Review 99*(1), 243–264.

Beck, T., A. Demirguc-Kunt, and R. Levine (2010). Financial Institutions and Markets Across Countries and Over Time: The Updated Financial Development and Structure Database. *World Bank Economic Review 24*(1), 77–92.

Bencivenga, V. R. and B. D. Smith (1991, April). Financial Intermediation and Endogenous Growth. *Review of Economic Studies 58*(2), 195–209.

Bocola, L. and A. Dovis (2016, September). Self-Fulfilling Debt Crises: A Quantitative Analysis. mimeo.

Chatterjee, S. and D. Corbae (2007, September). On the aggregate welfare cost of Great Depression unemployment. *Journal of Monetary Economics 54*(6), 1529–1544.

Chatterjee, S. and B. Eyigungor (2012). Maturity, Indebtedness, and Default Risk. *American Economic Review 102*(6), 2674–2699.

Cooper, R. and T. W. Ross (1998, February). Bank runs: Liquidity costs and investment distortions. *Journal of Monetary Economics 41*(1), 27–38.

Diamond, D. W. (1984, July). Financial Intermediation and Delegated Monitoring. *Review of Economic Studies 51*(3), 393–414.

Diamond, D. W. and P. H. Dybvig (1983, June). Bank Runs, Deposit Insurance, and Liquidity. *Journal of Political Economy 91*(3), 401–419.

Ennis, H. M. and T. Keister (2003, April). Economic Growth, Liquidity, and Bank Runs. *Journal of Economic Theory 109*(2), 220–245.

Ennis, H. M. and T. Keister (2006, March). Bank runs and investment decisions revisited. *Journal of Monetary Economics 53*(2), 217–232.

Epstein, L. G., E. Farhi, and T. Strzalecki (2014, September). How Much Would You Pay to Resolve Long-Run Risk? *American Economic Review 104*(9), 2680–97.

Fisher, I. (1936, April-June). 100% Money and the Public Debt. *Economic Forum*, 406–420.

Foley-Fisher, N. C., B. Narajabad, and S. H. Verani (2015, March). Self-fulfilling Runs: Evidence from the U.S. Life Insurance Industry. Federal Reserve Board Finance and Economics Discussion Series No. 2015-032.

Gertler, M. and N. Kiyotaki (2015). Banking, Liquidity and Bank Runs in an Infinite Horizon Economy. *American Economic Review 105*(7), 2011–2043.

Goldstein, I. and A. Pauzner (2005, June). Demand-Deposit Contracts and the Probability of Bank Runs. *Journal of Finance 60*(3), 1293–1327.

Gorton, G. B. and A. Metrick (2012, January). Getting up to Speed on the Financial Crisis: A One-Weekend-Reader's Guide. NBER Working Paper No. 17778.

Hardy, D. C. (2013, July). Bank Resolution Costs, Depositor Preference, and Asset Encumbrance. IMF Working Paper No. 13/172.

He, Z. and A. Krishnamurthy (2013). Intermediary Asset Pricing. *American Economic Review 103*(2), 732–770.

Holmstrom, B. and J. Tirole (1998, February). Private and Public Supply of Liquidity. *Journal of Political Economy 106*(1), 1–40.

Jacklin, C. J. (1987). Demand Deposits, Trading Restrictions and Risk Sharing. In E. C. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*, pp. 26–47. University of Minnesota Press.

Jalil, A. J. (2015). A New History of Banking Panics in the United States, 1825–1929: Construction and Implications. *American Economic Journal: Macroeconomics 7*(3), 295–330.

Lucas, R. E. (1987). *Models of Business Cycles*. Wiley-Blackwell.

Lucas, R. E. (2000, March). Inflation and Welfare. *Econometrica 68*(2), 247–274.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models*, Volume I, Chapter 25, pp. 75–96. London, UK: North-Holland.

Mendicino, C., K. Nikolov, J. Suarez, and D. Supera (2017). Optimal Dynamic Capital Requirements. *Journal of Money, Credit and Banking forthcoming*.

Morris, S. and H. S. Shin (1998, June). Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks. *American Economic Review 88*(3), 587–597.

Muller, A., K. Storesletten, and F. Zilibotti (2016, September). Sovereign Debt and Structural Reforms. mimeo.

Qi, J. (1994, Summer). Bank Liquidity and Stability in an Overlapping Generation Model. *The Review of Financial Studies 7*(2), 389–417.

Reinhart, C. M. and K. S. Rogoff (2009, January). The Aftermath of Financial Crises. NBER Working Paper No. 14656.

Reinhart, C. M. and K. S. Rogoff (2014, May). Recovery from Financial Crises: Evidence from 100 Episodes. *American Economic Review 104*(5), 50–55.

Robatto, R. (2017, March). Financial Crises and Systemic Bank Runs in a Dynamic Model of Banking. mimeo.

Segura, A. and J. Suarez (2017, October). How Excessive Is Banks' Maturity Transformation? *Review of Financial Studies 30*(10), 3528–3580.

Telyukova, I. A. and L. Visschers (2013, November). Precautionary Money Demand in a Business-Cycle Model. *Journal of Monetary Economics 60*(8), 900–916.

Van den Heuvel, S. J. (2008, March). The Welfare Costs of Bank Capital Requirements. *Journal of Monetary Economics 55*(2), 298–320.

Van den Heuvel, S. J. (2016, April). The Welfare Effects of Bank Liquidity and Capital Requirements. mimeo.

Appendices

A Details of the Numerical Algorithm

In order to characterize the banking equilibrium with runs, we need to calculate, for every possible combination of loan liquidation value r and probability of a run q, the expected value function of the no-run equilibrium in (60) as well as of the banking equilibrium with runs in (63). The former is straightforward, and we solve it by a standard value function iteration. To calculate the expected value function of the banking equilibrium with runs, instead, we need the value functions specific to the two alternative contracts that the bank can offer: the value function of the run-proof contract, $V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 0)$, and of the contract with possible runs, $V(a_t, s_{t-1} \mid q_t = q, \mathcal{I}_t = 1)$. To get to these, we take the following steps, for every parameter combination:

(i) build a 20x20x20 grid for the state space $\{k_t,\ c_{1t},\ z_{t-1}\}$;[14]

(ii) for every grid point, make an initial guess $\mathbb{E}V^0$ for the expected value function of the banking equilibrium with runs;

(iii) start iterating: solve for the run-proof contract and the contract with possible runs and, using $\mathbb{E}V^0$, write down the value functions specific to the two alternative contracts;

(iv) using equation (63), find $\mathbb{E}V^1$;

(v) repeat from step (iii) and iterate until convergence.

[14] For some combinations of the parameters, we increased the number of grid points to 30x30x30. However, the heavy curse of dimensionality, and the lack of any significant difference in the results, led us to keep a sparse grid for most of our simulations.