Systemic Bank Runs without Aggregate Risk: How a Misallocation of Liquidity May Trigger a Solvency Crisis*

Lukas Altermatt[†] Hugo van Buggenum[‡] Lukas Voellmy[§]

3rd November 2022

Abstract

We develop a general equilibrium model of self-fulfilling bank runs. The key novelty is the way in which the banking system's assets and liabilities are connected. Banks issue loans to entrepreneurs who sell goods to households, which in turn pay for the goods by redeeming bank deposits. The return on bank assets is thus contingent on households being able to withdraw their deposits. In a run, not all households that wish to consume manage to withdraw, since part of banks' cash reserves end up in the hands of households without consumption needs. This lowers revenues of entrepreneurs and bank asset returns, and thereby rationalises the run. Interventions that restrict redemptions in a run can be self-defeating due to their negative effect on demand in goods markets. We show how runs may be prevented with combinations of deposit freezes and redemption penalties as well as with the provision of emergency liquidity.

Keywords: *Fragility, deposit freezes, emergency liquidity.*
JEL codes: E4, E5, G2.

*We thank an anonymous referee as well as Toni Ahnert, Mohammed Aït Lahcen, Aleksander Berentsen, Matt Darst (discussant), Hans Gersbach, Piero Gottardi, Allen Head, Todd Keister, Cyril Monnet, Bruno Sultanum, Alex Vardoulakis, Randall Wright, and seminar participants at the University of Zurich and various conferences for comments that greatly improved the quality of this paper. The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the authors. They do not necessarily reflect the views of the Swiss National Bank (SNB). The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

[†]University of Essex, Wivenhoe Park, Colchester CO43SQ, UK. lu.altermatt@gmail.com
[‡]ETH Zürich, Zürichbergstrasse 18, 8092, Zürich, Switzerland. hvanbuggenum@ethz.ch
[§]Swiss National Bank, Börsenstrasse 15, 8022, Zürich, Switzerland. lukas.voellmy@snb.ch

1 Introduction

Systemic bank runs have long been an important topic in macro-financial economic research, and the financial crisis of 2007-2009 has only highlighted their importance further. However, we believe that the understanding of how systemic runs affect the real economy, how they may trigger solvency crises, and what this implies for the prevention of runs is still incomplete. Empirical observations show that systemic bank runs usually occur simultaneously with downturns in economic activity.[1] While a recession might increase the probability of a bank run, we argue that causality may also go the other way: a systemic bank run leads to a misallocation of liquidity, which hinders economic activity and causes the downturn. The fact that returns on assets held by banks decrease in the downturn may then rationalise the run in the first place. Thus, financial crises and recessions may arise endogenously and reinforce each other.

In this paper, we propose a model that captures this mechanism. The main innovation of our model relative to other bank run papers is that the return on some of the assets held by banks is endogenous and varies with aggregate demand. Specifically, banks make loans to entrepreneurs who sell goods to households, which in turn use bank deposits to purchase these goods. If a bank run occurs, some households who would like to consume early (impatient households) are unable to obtain liquid funds from the bank, which reduces aggregate demand in the goods market. This causes some entrepreneurs to default on their loans, which rationalises the bank run: patient households who wait do not receive the payouts they were promised due to the reduced return on bank assets, making it optimal for them to withdraw early. Importantly, running the bank is rational for patient households even if banks react to bank runs with (full or partial) deposit freezes. Thus, linking the return on bank assets to the conditions in the real economy overturns the result from the literature following Diamond and Dybvig (1983) that full suspension of convertibility prevents belief-driven runs if there are no fundamental shocks.[2,3]

We show that the banking system is prone to runs due to this mechanism unless the supply of assets unaffected by the shortfall in aggregate demand (e.g., government bonds) is high enough. We then show that if banks can use combinations of redemption penalties and deposit freezes, runs can always be stopped once they are observed, and they may also be prevented – however, even if banks observe runs immediately, there are cases where no such policy exists that prevents runs. Finally, we study emergency liquidity provided by a central bank that buys illiquid assets from banks in case of a run. This policy prevents runs if and only if the central bank purchases the assets at or close to face value. If the central bank instead buys these assets at a discount,

[1] See for example Reinhart and Rogoff (2008) or Gorton (1988).

[2] To the best of our knowledge, only two other papers exist that overturn this result, but for different reasons. See our literature review for a discussion of these.

[3] We want to stress here that the basic mechanism in our paper is different from fire sales, even though in both cases, asset prices fall during a run. With fire sales, the fundamental asset value remains unchanged, but banks may need to liquidate assets at a loss because of the run. In contrast, in our paper, the run causes a reduction in the fundamental value of assets. This difference also explains why deposit freezes do not prevent runs in our paper, while they typically do so in a model of fire sales.

some demand shortfall and thus some default by entrepreneurs is inevitable, meaning that running the banks is still rational for patient depositors. In a sense, these results go against Bagehot (1873), who advocated for central banks that lend to solvent, but illiquid banks at a high rate of interest.[4] While Bagehot's policy prescription seems natural, in our model, the banking system's liquidity problem turns into a solvency problem if and only if the central bank charges a sufficiently high penalty on emergency liquidity.

Before we explain the model and our results in more detail, we want to discuss the empirical relevance of our results and the particular mechanism they hinge on. In our model, bank runs are situations where depositors demand physical cash from banks, and if a run occurs, it hinders economic activity in sectors relying on cash transactions. Thus, our model is of particular relevance to cash-intensive economies, i.e. developing economies today, or historical episodes in the U.S. and Western European countries. However, we do believe that our model may also apply to situations where physical cash is not involved, for example, the run on shadow banks in the financial crisis of 2007-2009 or runs on banks in small-open economies, such as in Cyprus in 2012-2013. As long as a run hinders access to liquid assets by agents who would like to consume, our mechanism applies.

While it seems plausible that our mechanism is at play to some extent in many bank run episodes, we want to highlight one where it may have been particularly important, namely the 2001-2002 Argentinian deposit freeze.[5] The deposit freeze was enacted in late 2001 out of fear of bank runs; it restricted cash withdrawals from banks, while deposits could still be used to make payments. Since the Argentinian economy was relatively cash-intensive at the time, the restricted access to cash had negative effects on economic activity and contributed to a sharp drop in GDP. Even with the deposit freeze in place, pressure on banks did not stop, such that stronger measures were enacted at the beginning of 2002, including the forced conversion of dollar deposits to pesos and an exchange rate depreciation. While we do not want to claim that this episode was solely caused by belief-driven bank runs, it seems likely that the crisis was amplified by what we model, as it features all the relevant aspects of our mechanism.

Model summary. Our model builds on Diamond and Dybvig (1983), Lagos and Wright (2005), and Aruoba et al. (2011). As usual in bank run models, households are uncertain about their idiosyncratic liquidity demand, and as usual in New Monetarist models, periods are divided into two subperiods, called DM and CM. If households turn out to be impatient, they want to consume during the DM, but to do so they need money, which pays a (weakly) lower return than illiquid assets. Banks insure households against this liquidity risk by offering an optimal deposit contract in the spirit of Diamond and Dybvig (1983), where households who withdraw late receive weakly higher payouts. Besides holding money to satisfy withdrawals

[4] Bagehot (1873) is about lending while we study the provision of liquidity through asset purchases. However, we show that the results are equivalent for collateralised lending.

[5] See, e.g. Torre et al. (2003) for a detailed account of the episode, including a description of the disruptive effects the crisis had on the Argentinian payment system.

by impatient households, banks invest some of the deposits in illiquid assets that cannot be liquidated prematurely, namely government bonds and loans to entrepreneurs. Goods in the DM are produced by entrepreneurs using their own labour and capital as inputs, as in Aruoba et al. (2011). Since entrepreneurs have no funds of their own, they need loans in order to purchase capital.

We proceed by first discussing run-free steady states, and then we study whether the steady-state equilibria we find are prone to unexpected runs. We show that three steady-state banking equilibria may exist in this economy: a full liquidity-insurance equilibrium (FLI) where households are fully insured against liquidity risk; a zero-lower bound equilibrium (ZLB) where all assets pay the same return as money, such that banks cannot offer liquidity insurance but capital investment and therefore DM consumption is higher than in an economy without banks; and a partial liquidity-insurance equilibrium (PLI), which is an intermediate case. FLI (ZLB) equilibria are more likely to exist if the share of impatient households and the supply of government bonds is high (low), but all three equilibria may coexist for some parameter values.

Turning to the analysis of runs, we show that even if banks react to a run with full or partial deposit freezes, an unexpected bank run is always a Nash equilibrium if the economy is in a ZLB or PLI equilibrium. If the economy is in an FLI equilibrium, a run is a Nash equilibrium if the real supply of government bonds is not too large. To reiterate, running is rational for patient households even with deposit freezes because if other patient households run the bank, some impatient households end up with less money than expected in the DM, which leads some entrepreneurs to default and thus lowers the value of the banks' assets – the run triggers a solvency crisis, which in turn rationalises the run. If banks can impose any combination of redemption penalties and deposit freezes once they observe the run, runs can be prevented if the policy satisfies three criteria: (i) it must be suboptimal for patient depositors to continue running once the policy is implemented; (ii) it must be optimal for impatient depositors to withdraw early once the policy is implemented; (iii) given that the run is stopped once the policy is implemented, it must be suboptimal for patient depositors to run in the first place. It turns out that for certain parameters, no such policy exists, even if banks observe the run immediately.

Finally, we study what happens if the government purchases illiquid assets from banks in case of a run. Note that in our model, banks cannot liquidate illiquid assets early unless the government is willing to purchase them. Thus, in the absence of government intervention, banks implement a full deposit freeze by default once they run out of cash.[6] This provision of emergency liquidity always stops runs, and prevents them if and only if the government buys the assets at or close to face value. In this case, banks can continue paying out the promised amount to impatient depositors, which stabilises aggregate demand, thereby preventing defaults by entrepreneurs and thus giving patient depositors no reason to run in the first place. If the government buys these assets at a sufficiently large discount however, banks will run into a

[6] This is not an assumption, but a result of microfounding the return on illiquid assets. In our model, nobody holds the cash required to purchase these assets before maturity.

solvency crisis unless they reduce payouts to impatient depositors. This leads to a shortfall in demand and some default by entrepreneurs, such that running the bank remains rational for patient depositors ex ante. In ZLB and PLI equilibria, any discount imposed will lead to defaults and fails to prevent runs. In FLI equilibria, runs might still be prevented with positive but small discounts. Importantly, our results on the provision of emergency liquidity are derived under the assumption that the central bank cannot guarantee that the injected liquidity increases real purchasing power. This is because it cannot withdraw money from circulation with lump-sum taxation. In our model, effective stabilisation therefore does not require fiscal backing for a central bank.

Existing literature. The core of our model is based on Aruoba et al. (2011), who introduced a setting in which a neoclassical production technology (with capital and labour as inputs) is used to produce goods sold against fiat money.[7] The model of Aruoba et al. (2011) is itself based on the New Monetarist framework (Kiyotaki and Wright (1989), Lagos and Wright (2005)). Similarly to our paper, Geromichalos and Herrenbrueck (2021) introduce random liquidity needs into a setting based on Aruoba et al. (2011). Rather than focusing on the role of banks as we do in the present paper, Geromichalos and Herrenbrueck (2021) study the role of secondary asset markets (from which we abstract) in satisfying agents' liquidity needs.[8]

Our paper contributes to a literature studying how banks can (or cannot) eliminate panic equilibria, in particular with regard to the role of deposit freezes. A key result of Diamond and Dybvig (1983) is that in the absence of aggregate uncertainty, bank runs can be prevented at no cost if banks fully freeze deposits after a certain number of depositors have withdrawn.[9] This suggests that bank runs are only a difficult problem to overcome in the presence of aggregate uncertainty. Up to this point, two main objections to this result have been raised in the theoretical literature. First, Engineer (1989) showed that in a riskfree Diamond-Dybvig setting with more periods, deposit freezes can give rise to run equilibria where depositors withdraw preemptively out of fear of not being able to access their deposits when needed. Second, Ennis and Keister (2009) showed that Diamond and Dybvig's result hinges crucially on the assumption that banks can commit to fully freeze deposits even if this severely hurts some of their depositors ex post. With limited commitment, deposit freezes are much less effective in preventing runs, even in the absence of aggregate uncertainty.[10] The present paper adds a third, distinct reason why deposit

[7]In New Monetarist terminology, capital is used as an input in DM production. This is in contrast to e.g. Aruoba and Wright (2003) and Lagos and Rocheteau (2008) where capital is used as a production input in the frictionless CM but not in the DM.

[8]In Geromichalos and Herrenbrueck (2021), capital can be sold against cash and thus has 'indirect' liquidity value. In our paper, capital investment by firms backs interest-paying demand deposits issued by banks.

[9]It is well known that full deposit freezes are generally not desirable in the presence of aggregate uncertainty, i.e. if either the investment technology is risky or liquidity needs of depositors are stochastic. Wallace (1990) provides a detailed discussion of full and partial deposit freezes for the case with random liquidity needs and a riskfree investment technology. Matta and Perotti (2021) show that full deposit freezes are generally not optimal when depositors' liquidity needs are known but the investment technology is risky.

[10]Relatedly, even if not explicitly focusing on deposit freezes, Keister and Mitkov (2019) show that it may

freezes may not eliminate panic equilibria in settings without (exogenous) aggregate uncertainty, even if banks can fully commit to any payout policy. Specifically, the present paper highlights a general equilibrium effect of freezing deposits: by restricting access to cash for households that wish to consume, deposit freezes have a negative effect on aggregate demand, which – through its effect on the return to banks' assets – can rationalise the run in the first place.[11]

Our paper is related to a broader literature studying self-fulfilling systemic bank panics in general equilibrium. The previous literature on the topic, starting with the seminal contribution by Gertler and Kiyotaki (2015), has mostly focused on the case where general equilibrium effects of bank runs occur through the effect of runs on asset prices.[12] In these models, banks that are hit by runs liquidate assets, which depresses asset prices and has repercussions on the entire economy and financial system. The present paper does not feature fire sales but instead focuses on a different type of general equilibrium effect of bank runs, namely the misallocation of cash caused by runs. Two closely related papers are Robatto (2019) and Carapella (2015), both of which study general equilibrium effects of runs that are similar to our paper. In both papers, runs reduce the amount of liquid assets (cash or deposits) in the hands of those who wish to consume, which has deflationary effects and thus reduces banks' net worth. Similar to our paper, the feedback effect from impaired liquidity provision by banks to asset prices can lead to self-fulfilling (systemic) run equilibria. Crucially, as both of these papers study endowment economies, they cannot speak to the mutually reinforcing effects of financial crises and real recessions. They also do not study whether deposit freezes may stop or prevent runs.

Further, our contribution is part of a literature that studies the role of banks in providing liquidity insurance – and the fragility of these arrangements – within the New Monetarist framework. In our model, banks issue interest-paying demand deposits, which allows to insure depositors against random liquidity needs in a way that is reminiscent of Berentsen et al. (2007) and of Williamson (2012).[13] As we do in our paper, both Sanches (2018) and Andolfatto et al. (2020) combine the Lagos-Wright and Diamond-Dybvig frameworks. In Sanches (2018), a bank run impairs the use of bank deposits as a means of payment in DM trades. The negative effect on DM activity may persist over and above the period in which the run takes place due to liquidation losses incurred by banks in the run.[14] Andolfatto et al. (2020) study a model where

be privately optimal for banks to refrain from imposing measures that stop a run if they expect to receive a government bailout when hit by a run.

[11] The idea that early withdrawals negatively affect the long-run return on banks' assets is shared by Andolfatto and Nosal (2020) and Kashyap et al. (2020), although for very different reasons than in our paper. In Andolfatto and Nosal (2020), banks' investment technology exhibits increasing returns to scale, such that the average return decreases if banks scale down investment as a result of withdrawals. In Kashyap et al. (2020), early withdrawals negatively affect bank profitability, which in turn reduces bankers' incentive to monitor loans.

[12] See also Liu (2019) and Goldstein et al. (2020) for models of systemic bank panics with fire sales.

[13] In Berentsen et al. (2007), banks intermediate cash between agents with and without consumption needs. In our model, as in Williamson (2012), banks pool agents' cash and assets and issue demand deposits.

[14] The persistent negative effects of runs in Sanches (2018) are due to the fact that liquidation losses incurred in period t affect the deposit contracts banks can offer in period $t+1$. Our model abstracts from such dynamic effects of runs, as the economy reverts to steady state in the period following the run.

banks and markets coexist and examine how the presence of markets affects banks' ability to provide liquidity insurance, a topic that is not the focus of our paper. Gu et al. (2019) highlight the inherent fragility of various aspects of banking in a wide variety of models, including one where banks act as a provider of a means of payment in a Lagos-Wright setting.

Finally, our paper is related to a theoretical literature studying the public sector's role in providing emergency liquidity in a run. The previous literature on the topic (e.g. Gorton and Huang (2006), Martin (2006), Rochet and Vives (2004), Farhi and Tirole (2020)) has largely focused on the trade-off between preventing inefficient runs on the one hand and avoiding the creation of moral hazard on the other hand. Compared to these papers, we highlight banks' role in providing a means of payment and how this affects the way in which emergency liquidity needs to be provided in order to eliminate self-fulfilling panics.

Outline. The rest of this paper is structured as follows: Section 2 presents the environment, Section 3 discusses the planner's solution, and Section 4 discusses the decentralized economy without banks. Section 5 introduces banks and discusses the steady-state banking equilibrium without runs. Section 6 discusses bank runs, Section 7 discusses what the banking system can do to stop or prevent runs, and Section 8 discusses how the central bank may stop or prevent runs. Finally, Section 9 concludes.

2 Environment

The environment is based on Lagos and Wright (2005), Rocheteau and Wright (2005), and in particular on Aruoba et al. (2011). In Section 5 we add banks in the tradition of Diamond and Dybvig (1983). Time is discrete, indexed by $t = 0, 1, 2, ..$, and continues forever. Each period t is divided into two subperiods, called CM and DM.[15] The CM opens at the beginning of each period, and once it closes, the DM opens and remains open until the period ends. Both markets are competitive. There are two types of agents in the economy, a measure one of households and a measure n of entrepreneurs. Households are infinitely-lived, whereas entrepreneurs of generation t are born at the beginning of the CM of period t and live until the end of the CM of period $t + 1$, when they are replaced by a new generation of entrepreneurs. In the CM, a good x can be produced by households at linear disutility l, where one unit of l yields one unit of x. Good x cannot be stored, but it can be converted into capital k by young entrepreneurs one to one in order to produce another nonstorable good q in the DM. This good is produced according to

$$q = f(k, h), \tag{1}$$

where k denotes the amount of capital owned by the entrepreneur, h denotes his labour effort, and $f(k, h)$ satisfies the CRS property. Capital fully depreciates after production. We assume

[15]This terminology follows standard conventions in the literature, whereby each time period is usually divided into two submarkets, a frictionless CM (which may stand for 'centralized market') and a DM (which may stand for 'decentralized market') in which trades need to be settled with fiat money.

that $f_k(k,h) > 0$, $f_{kk}(k,h) < 0$, $f_h(k,h) > 0$, $f_{hh}(k,h) < 0$, $f_{kh} > 0$, and $f(0,h) = f(k,0) = 0$. By inverting $f(k,h)$, we can rewrite it as

$$h = c(q,k), \tag{2}$$

where $c(q,k)$ represents the amount of labour required to produce q units of the consumption good given k. From our assumptions on $f(k,h)$ it follows that $c_q(q,k) > 0$, $c_{qq}(q,k) > 0$, $c_k(q,k) < 0$, $c_{kk}(q,k) > 0$, and $c_{qk}(q,k) < 0$.

During the DM, a fraction θ of households get utility $u(q)$ out of consuming the good q. We call these agents *impatient households*. The remaining fraction $1-\theta$ get no utility from consumption during the DM and are called *patient households*. Each household's type during period t is revealed to them at the beginning of the DM of period t and is private information. The realisation of types is i.i.d. across periods and households. In the CM, all households get utility $U(x)$ from consuming the CM good. The expected lifetime payoff of households is

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left\{ U(x_t) - l_t + \theta u(q_t) \right\}. \tag{3}$$

$U(x)$ and $u(q)$ are both strictly increasing, concave functions satisfying the Inada conditions. Additionally, we impose $u(0) = 0$.

Entrepreneurs get linear disutility from working during the DM and linear utility x from consuming during the CM when they are old. In the interest of symmetry with households, we assume entrepreneurs also discount the next period by a factor β. The utility function of a young entrepreneur at the beginning of the CM is thus given by

$$\mathrm{E}_t \left\{ -h_t + \beta x_{t+1} \right\}. \tag{4}$$

As is standard in the literature, we assume that agents are anonymous during the DM, that there is no record-keeping technology, and that households cannot credibly commit to any future payments.[16] These assumptions rule out unsecured credit, so households need liquid assets in order to purchase consumption goods from entrepreneurs during the DM. We assume that the only liquid asset in the sense that entrepreneurs can recognize it and distinguish it from counterfeited assets is intrinsically worthless fiat money m, which is issued by the government. Thus, households who want to consume in the DM are subject to the constraint

$$p_t q_t \leq m_t, \tag{5}$$

where p_t denotes the price of the DM good in period t. We denote the stock of fiat money at the beginning of period t as M_t and the value of fiat money in terms of good x during the CM of period t as ϕ_t, i.e. m_t units of fiat money buy $\phi_t m_t$ units of good x. This implies that the gross inflation rate is given by $1 + \pi_{t+1} \equiv \frac{\phi_t}{\phi_{t+1}}$.

[16] The assumption that entrepreneurs can commit to repayment while households cannot may be motivated by assuming a record-keeping technology is available in the CM but not during the DM.

Since young entrepreneurs do not have resources of their own, they need to borrow from households in order to invest in capital. We denote by ℓ_t nominal loans extended by households to entrepreneurs. To purchase k_t units of capital in the CM of period t, a young entrepreneur needs to take out a nominal loan of $\ell_t = k_t/\phi_t$. The net nominal interest rate on loans extended in period t is denoted by $\tilde{i}_{\ell,t+1}$, that is, an entrepreneur receiving a nominal loan of ℓ_t in the CM of period t is due to repay $(1+\tilde{i}_{\ell,t+1})\ell_t$ units of fiat money in the CM of $t+1$. We assume entrepreneurs are able to commit to repayment, in the sense that they always repay their loans if they have the funds to do so. If their funds are insufficient to repay the loan in full, they will repay all they have. Notice however that entrepreneurs cannot be forced to work in the DM.[17]

In addition to issuing fiat money, the government also issues nominal bonds B. A household holding an amount b of bonds issued in period t receives $(1+i_{b,t+1})b$ units of fiat money in period $t+1$. Thus, the government budget constraint is given by

$$\phi_t(B_t + M_t) + \Delta_t = \phi_t(M_{t-1} + (1+i_{b,t})B_{t-1}), \tag{6}$$

where Δ_t denote lump-sum taxes (or subsidies if $\Delta_t < 0$) imposed on households in the CM of period t. We assume that the money supply grows at a constant net rate μ, that the government targets a real debt level $\mathcal{B} = \phi_t B_t$, and that lump-sum taxes Δ_t adjust such that the budget constraint holds given these targets.

For future reference, we denote the Fisher rate (i.e. the nominal interest rate that fully compensates for inflation and discounting) as $1 + \iota_{t+1} \equiv \frac{1+\pi_{t+1}}{\beta}$.

3 Planner's Problem

Before turning to market outcomes, we solve the planner's problem to determine the optimal quantities of CM consumption x, DM consumption q, and capital investment k. We denote by x_t and x_t^e the CM consumption levels in period t by households and entrepreneurs respectively. Further, we denote by q_t^e the amount of DM good produced by each entrepreneur, whereas q_t denotes the amount consumed by each (impatient) household. A planner maximises the expected lifetime utility of households and entrepreneurs, giving equal weight to all agents:

$$\max_{\{l_t,h_t,q_t,x_t,x_t^e\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \{U(x_t) - l_t + \theta u(q_t) + n(-h_t + x_t^e)\}$$

$$\begin{aligned} s.t. \quad & l_t = x_t + n x_t^e + n k_t \\ & h_t = c(q_t^e, k_t) \\ & \theta q_t = n q_t^e \end{aligned}$$

[17] The amount of cash that entrepreneurs have in the CM when they are old (and with which they can repay the loan) depends on their labour effort in the preceding DM. Notably, entrepreneurs can always choose not to work at all in the DM, in which case they receive zero revenue and thus will default on their entire loan.

The first constraint says that households' labour effort in the CM must equal consumption of the CM good by households and entrepreneurs plus capital investment. The second constraint is the DM production function, and the third constraint is market clearing in the DM. Inserting the constraints into the objective function, we can reformulate the planner's problem as

$$\max_{\{x_t, x_t^e, k_t, q_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left\{ U(x_t) - nk_t - x_t - nx_t^e + \theta u(q_t) + n \left[-c(q_t \frac{\theta}{n}, k_t) + x_t^e \right] \right\},$$

Defining $\kappa_t \equiv \frac{nk_t}{\theta q_t} = \frac{k_t}{q_t^e}$ as the ratio of capital per DM good produced, we get the following first-order conditions, with an asterisk (*) denoting the (unique) first-best quantities.[18,19]

$$U'(x^*) = 1 \tag{7}$$

$$u'(q^*) = c_q(1, \kappa^*) \tag{8}$$

$$1 = -c_k(1, \kappa^*). \tag{9}$$

Condition (7) pins down households' optimal CM consumption x^*. Impatient households' optimal DM consumption q^* as well as optimal capital investment κ^* are determined jointly by (8) and (9). Condition (8) states that the marginal utility from consuming the DM good must equal the marginal disutility of labour incurred from producing it. Condition (9) states that the marginal benefit from increasing the amount of capital used in the production of the DM good (through reduced hours worked) must equal the disutility of producing more capital.

4 Equilibrium without Banks

Let us now turn to the allocation resulting in a market economy without banks. In this and the next section, we will restrict attention to monetary steady-state equilibria, i.e. equilibria where $\phi_t > 0$ for all t and where all real quantities are constant.[20] In a monetary steady-state, the Fisher rate equals $1 + \iota = \frac{1+\mu}{\beta}$, where μ is the constant money growth rate.

Households. In the CM, households solve[21]

$$\max_{x_t, q_t, \ell_t, b_t} U(x_t) - x_t + \theta u(q_t) + [\beta(1-\theta)\phi_{t+1} - \phi_t]p_t q_t + [\beta\phi_{t+1}(1 + i_{\ell,t+1}) - \phi_t]\ell_t$$
$$+ [\beta\phi_{t+1}(1 + i_{b,t+1}) - \phi_t]b_t, \tag{10}$$

where ℓ_t denotes loans made by households to entrepreneurs, $1 + i_{\ell,t+1}$ denotes the effective nominal interest rate on loans (taking expected defaults into account), and $p_t q_t = m_t$. In (10),

[18] Since the first-best quantities are constant, time subscripts are left out. Note also that any amount of CM consumption by entrepreneurs (x^e) is optimal. This results from the fact that both marginal utility of CM consumption by entrepreneurs as well as the marginal disutility of producing the CM good are equal to 1.

[19] In introducing κ, we have made used of the constant-returns-to-scale property of $f(k, h)$, which is inherited by $c(q, k)$. We thus have $c(q^e, k) = q^e c(1, \kappa)$, $c_q(q^e, k) = c_q(1, \kappa)$, and $c_k(q^e, \kappa) = c_k(1, \kappa)$.

[20] There is always an equilibrium with $\phi_t = 0 \ \forall t$ in this class of models. In this equilibrium, $q = k = 0$ due to the lack of an accepted means of payment to settle DM transactions.

[21] The households' problem is standard for models based on Lagos and Wright (2005), so we keep the exhibition brief here. For readers interested in the details, we discuss the derivation of the problem in Appendix A.1.

the first two terms capture the utility and opportunity cost from CM consumption, respectively; the third term captures the utility from DM consumption in case the households turns out impatient; and the final three terms capture the portfolio choice regarding m, ℓ, and b, where in each bracket, the first term captures the utility from bringing the respective asset to the CM in $t+1$ (which only happens if the household turns out to be patient for m), and the second term in the bracket captures the cost of acquiring the asset in period t. Defining $\rho_t \equiv \beta\phi_{t+1}p_t \equiv \frac{\phi_t p_t}{1+\iota_{t+1}}$ as the real price of the DM good, the first-order conditions to this problem are given by

$$U'(x_t) = 1 \tag{11}$$

$$u'(q_t) = \rho_t \left(1 + \frac{\iota_{t+1}}{\theta}\right) \tag{12}$$

$$\iota_{t+1} = i_{\ell,t+1} \tag{13}$$

$$\iota_{t+1} = i_{b,t+1}. \tag{14}$$

Note that equations (13) and (14) give the only values of the loan and bond interest rate that allow for interior solutions. At $i_{\ell,t+1} < \iota_{t+1}$ ($i_{\ell,t+1} > \iota_{t+1}$), demand for loans is zero (infinity), and the same holds for bonds. However, since the supply of these assets in a monetary equilibrium is non-zero and finite, the loan and bond markets only clear if the interest rate on these assets equals the Fisher rate, i.e. if equations (13) and (14) hold. At these interest rates, households are willing to hold any amount of ℓ and b.

Entrepreneurs. Entrepreneurs born in the CM of period t choose the loan taken out when young (ℓ_t^e), capital investment when young (k_t), DM labour input (h_t) and CM consumption when old (x_{t+1}^e). Since entrepreneurs do not get utility from consumption when young, they will spend their entire loan to invest in capital; choosing a given loan size thus implies the choice of a given capital stock and vice versa. We first take it as given that entrepreneurs will repay their loan in full, and then show that this will indeed be the case in equilibrium. The entrepreneurs' optimisation problem can then be expressed as:

$$\begin{aligned} \max_{\{k_t,h_t,x_{t+1}^e\}} \quad & \{-h_t + \beta x_{t+1}^e\} \\ \text{s.t.} \quad & x_{t+1}^e = \phi_{t+1}\big[p_t q_t^e - (1+\tilde{i}_{\ell,t+1})\ell_t^e\big] \\ & k_t = \phi_t \ell_t^e \\ & h_t = c(q_t^e, k_t) \end{aligned}$$

The first constraint represents the entrepreneurs' budget constraint, saying that CM consumption when old equals the revenue from selling DM good minus the loan repayment. The second constraint says that the capital stock must be financed with a corresponding nominal loan. The last condition follows from the DM production function. Inserting the constraints into the objective function, we can reformulate this problem as

$$\max_{k_t,q_t^e} \quad \rho_t q_t^e - c(q_t^e, k_t) - \frac{1+\tilde{i}_{\ell,t+1}}{1+\iota_{t+1}} k_t.$$

The first-order conditions with respect to k_t and q_t^e are

$$-c_k(1,\kappa_t) = \frac{1+\tilde{i}_{\ell,t+1}}{1+\iota_{t+1}} \tag{15}$$

$$\rho_t = c_q(1,\kappa_t). \tag{16}$$

Finally, we need to check whether entrepreneurs are willing to work in the DM and repay their loan given their choice of k_t, which is the case if and only if

$$\max_{q_t^e} \quad \rho_t q_t^e - c(q_t^e, k_t) - \frac{1+\tilde{i}_{\ell,t+1}}{1+\iota_{t+1}} k_t \geq 0, \tag{17}$$

i.e. if and only if the maximised return from production is enough to cover the cost of the loan and the disutility of working. If (17) is violated, entrepreneurs are better off not producing at all, defaulting on their entire loan. In Appendix A.2, we verify that condition (17) is indeed slack given the FOCs (15)-(16).[22]

Equilibrium. Note first that there is no default in equilibrium, since under rational expectations, all agents know how large the demand for good q will be in the DM. Households will thus never extend loans to entrepreneurs that are so large that entrepreneurs would not be willing to repay them. Therefore, we have $\tilde{i}_{\ell,t+1} = i_{\ell,t+1}$ in equilibrium, i.e. the expected nominal loan rate equals the contractual rate. Then, market clearing for loans is

$$nk_t = \phi_t \ell_t, \tag{18}$$

while market clearing in the DM is

$$nq_t^e = \theta q_t. \tag{19}$$

Combining conditions (12), (13), (15) and (16), and using the fact that $\tilde{i}_{\ell,t+1} = i_{\ell,t+1}$, we obtain that the equilibrium levels of q and κ in steady state are determined by

$$\frac{u'(q)}{c_q(1,\kappa)} = 1 + \frac{\iota}{\theta} \tag{20}$$

$$-c_k(1,\kappa) = 1. \tag{21}$$

It is easy to see that the Friedman rule ($\iota = 0$) implements the first-best allocation. Away from the Friedman rule, $\frac{\iota}{\theta}$ creates a wedge between the marginal utility of consuming the DM good and the marginal disutility of producing it. The higher inflation and the lower the probability that a household turns out to be impatient are, the further q is below its first-best level. The opportunity cost of holding money ι is compounded by the fact that households do not know ex ante whether they actually want to consume in the DM – the lower θ, the larger the cost resulting from this uncertainty. Comparing (21) with (9) shows that κ is optimal in equilibrium, i.e. a given q will be produced with the socially optimal mix of capital and labour. This results from the fact that the equilibrium real loan rate equals agents' rate of time preference, leading them to coordinate on the efficient level of capital investment to produce a given DM output.

[22] To be more precise, the constraint is *just* slack, with the Lagrange multiplier being zero while the constraint holds at equality. This is due to entrepreneurs making zero profits in equilibrium, which in turn follows from the constant returns to scale property of $f(k,h)$.

5 Banking Equilibrium

To overcome the uncertainty about their consumption preferences, households can form coalitions during the CM which we call banks, à la Diamond and Dybvig (1983). Banking coalitions maximise the expected utility of their participating households, which we call depositors, with each depositor depositing an identical amount in the bank when it is formed.[23] Banks act as price takers, in the sense that they take equilibrium loan and bond rates, the value of money, and DM prices as given. They exist for one period, from CM to the next CM, after which they dissolve and are replaced by a new set of banks.

In this section, we solve for steady-state banking equilibria without runs, which also implies no defaults by entrepreneurs. We then investigate whether the banking equilibria we find are prone to unexpected runs in the subsequent sections.

Banks issue deposits in the CM when they are formed and invest the proceeds in cash (m_t^b), loans (ℓ_t^b) and government bonds (b_t^b), where we use the superscript b to denote portfolio choices made by banks. We denote by $a_t^b \equiv \ell_t^b + b_t^b$ a bank's total holdings of illiquid assets. Since expected loan defaults are zero, the nominal return on loans and bonds needs to be the same in equilibrium, i.e. $i_{\ell,t} = i_{b,t} \equiv i_t \; \forall t$. Thus, when we refer to the interest rate in the economy, this covers both of these rates.[24] Note, however, that this interest rate may differ from the Fisher rate ι_t, which is the opportunity cost of holding cash. A bank formed in the CM of period t promises to pay out d_t^I units of money to impatient depositors in the following DM, and d_{t+1}^P units of money to patient depositors in CM next period.[25] The bank's problem is given by

$$\max_{m_t^b, a_t^b, d_t^I, d_{t+1}^P} \theta u(q_t) + \beta(1-\theta)\phi_{t+1} d_{t+1}^P - \phi(m_t^b + a_t^b)$$

subject to:

$$d_t^I = p_t q_t \tag{22}$$

$$m_t^b \geq \theta d_t^I \qquad (\beta\phi_{t+1}\zeta_t) \qquad \text{liquidity const.} \tag{23}$$

$$m_t^b + a_t^b(1+i_{t+1}) \geq \theta d_t^I + (1-\theta) d_{t+1}^P \qquad (\beta\phi_{t+1}\xi_t) \qquad \text{solvency const.} \tag{24}$$

$$d_{t+1}^P \geq d_t^I \qquad (\beta\phi_{t+1}(1-\theta)\psi_t)) \qquad \text{patient IC const.} \tag{25}$$

$$u(q_t) - u(0) \geq \beta\phi_{t+1} d_{t+1}^P, \qquad \text{impatient IC const.} \tag{26}$$

[23]While banks issue nominal claims, it is immaterial whether deposits are made in cash or in CM goods.

[24]From the previous section, we know that $i_{\ell,t} = i_{b,t} = \iota_t$ must hold in an economy without banks. Households are only willing to hold illiquid assets at these rates outside of banks also in the economy with banks. Banks might be willing to hold them at lower rates, but if $i_{\ell,t} \neq i_{b,t}$, they would strictly prefer to hold whichever asset pays the higher interest rate. Further, no asset can pay a higher interest rate than ι_t, as otherwise demand from households would be infinite.

[25]The assumption that impatient (patient) depositors are only paid out in the DM (CM) is without loss of generality. First, it is straightforward to show that paying out patient depositors in the DM would not increase expected utility, since patient depositors would only hold on to such payouts until the next CM anyway. Second, if the bank had leftover funds to pay out impatient depositors in the CM, depositors would have been (weakly) better off making smaller deposits in the first place. For similar reasons, it is never efficient for the bank to pay out more to impatient depositors than they plan to spend in the DM.

where the terms in brackets denote the Lagrange multipliers on the constraints. Equation (22) states that choosing d^I is equivalent to choosing the DM consumption of impatient households (q_t) as banks take the DM price p_t as given. Equation (23) is a liquidity constraint, which says that the bank must hold enough cash to make payments to early withdrawers. In contrast to individual households, banks can economise on cash holdings by exploiting the law of large numbers. Equation (24) is a solvency constraint, which says that the total value of the banks' assets must be at least as high as the total promises it makes to depositors. Equation (25) is an incentive compatibility constraint for patient depositors, which states that DM payouts cannot exceed CM payouts. Finally, equation (26) denotes the incentive compatibility constraint for impatient depositors, which states that the difference in utility from consuming q_t in the DM instead of nothing should exceed the utility they obtain from withdrawing in the CM. We will ignore this constraint for now and check later whether or not it binds.[26]

Before moving on, we want to highlight the role illiquid asset holdings play for banks. From the constraints (24) and (25), it can be seen that the sole reason banks hold these assets is to satisfy the IC constraint for patient agents. If types were observable, banks could just invest in cash; but since types are private information, banks must hold illiquid assets to make payments to patient households that are large enough to satisfy incentive compatibility.[27]

We refer to Appendix A.3 for a derivation of the optimality conditions of the bank's problem. For the Lagrange multipliers, we get that

$$\zeta_t = \frac{i_{t+1}(1+\iota_{t+1})}{1+i_{t+1}}, \quad \xi_t = \frac{1+\iota_{t+1}}{1+i_{t+1}}, \quad \text{and} \quad \psi_t = \frac{\iota_{t+1}-i_{t+1}}{1+i_{t+1}}, \tag{27}$$

i.e. the liquidity constraint is binding unless the nominal rate is zero, the solvency constraint is always binding (which just means that banks make zero profits) and patient depositors' IC constraint is binding unless the nominal rate equals the Fisher rate. Intuitively, the latter is the case because, if assets pay the Fisher rate, buying additional assets in order to increase CM payouts and relax patient depositors' IC constraint entails no opportunity cost for banks. Further, we get that consumption in the DM is characterised by

$$u'(q_t) = \rho_t \left[1 + \frac{\theta\iota_{t+1}(1+i_{t+1}) + (1-\theta)(\iota_{t+1}-i_{t+1})}{\theta(1+i_{t+1})}\right], \tag{28}$$

and the asset demand schedule of the bank is

$$m_t^b \geq \theta p_t q_t \qquad \text{with equality if } i_{t+1} > 0 \tag{29}$$

$$m_t^b + a_t^b(1+i_{t+1}) \geq p_t q_t \qquad \text{with equality if } i_{t+1} < \iota_{t+1}. \tag{30}$$

Condition (29) implies that banks only hold excess cash if the nominal interest rate is zero. Condition (30) means that promised DM payouts cannot exceed the return the bank earns on its portfolio of cash and loans, which follows from patient depositors' IC constraint.

[26] We assume here that households hold no money outside the bank. This is without loss of generality, as households would never find it optimal to hold money outside of the bank when a bank exists.

[27] Our finding that incentive compatibility for patient households limits banks' ability to provide liquidity insurance is similar to Peck and Setayesh (2022), where banks can only provide optimal liquidity insurance if households deposit a large enough share of their endowment in the bank.

The entrepreneur's problem is unchanged compared to the economy without banks. Thus, the relevant equilibrium conditions are still given by equations (15) and (16).

Market clearing in the loan market is now

$$nk_t = \phi_t(\ell_t + \ell_t^b), \tag{31}$$

and market clearing in the bond market is

$$B_t = b_t + b_t^b. \tag{32}$$

Assuming that all households participate in the banking coalition, DM clearing is

$$nq_t^e = \theta q_t.$$

Combining (15), (16) and (28), and using the fact that $\tilde{i}_{\ell,t+1} = i_{t+1} = i$, we get that q and κ in a monetary steady state satisfy

$$\frac{u'(q)}{c_q(1,\kappa)} = 1 + \frac{\theta\iota(1+i) + (1-\theta)(\iota - i)}{\theta(1+i)} \tag{33}$$

$$-c_k(1,\kappa) = \frac{1+i}{1+\iota}. \tag{34}$$

Taking nominal rates i and ι as given, this system has a unique solution. Condition (34) shows that the capital stock per DM good produced (κ) is uniquely determined through i and ι. Comparing (34) with (9), we get that κ is at its efficient level if $i = \imath$ and is inefficiently high whenever $i < \imath$. Furthermore, from (16), we have that

$$\rho = c_q(1,\kappa) = c(1,\kappa) - \kappa c_k(1,\kappa) \tag{35}$$

in steady state, so that the real price of the DM good (ρ) is uniquely determined through κ. It is easy to see that κ depends negatively on i, while ρ depends positively on i (through κ). From (33) and (35), we can see that changes in i have two opposing effects on DM output q. On the one hand, an increase in i increases ρ which taken by itself has a negative effect on q. On the other hand, the right-hand side of (33) is strictly decreasing in i for any $\theta < 1$, capturing the fact that higher i allows banks to pay higher interest on deposits, thereby decreasing the effective cost of liquidity, which taken by itself has a positive effect on q.

Next, we denote by $\mathcal{A}$ the real supply of illiquid assets, which equals the aggregate capital stock plus real government debt:

$$\mathcal{A} \equiv \phi_t a_t = nk + \mathcal{B} = \kappa q\theta + \mathcal{B}. \tag{36}$$

Combining aggregate asset supply (36) with the bank's asset demand schedule (29)-(30), we get that the asset market clears if and only if:[28]

$$\mathcal{A}\begin{cases} \geq (1-\theta)\rho q & \text{if} \quad i = \iota \\ = \frac{1+\iota}{1+i}(1-\theta)\rho q & \text{if} \quad i \in (0,\iota) \\ \leq (1+\iota)(1-\theta)\rho q & \text{if} \quad i = 0 \end{cases} . \tag{37}$$

[28] Recall that all assets in the economy are held by banks whenever $i < \iota$.

Definition 1. *A stationary monetary equilibrium (SME) is given by* $(q, \kappa, \rho, i, \mathcal{A})$ *satisfying* (33)-(37).

The existence proof for SME is given in Appendix A.4:

Proposition 1. *A stationary monetary equilibrium exists.*

In the following, we will distinguish between three equilibrium cases: (i) *full liquidity insurance equilibria (FLI)*, defined as equilibria with $i = \iota$; (ii) *zero lower bound equilibria (ZLB)*, defined as equilibria with $i = 0$ and (iii) *partial liquidity insurance equilibria (PLI)*, defined as equilibria with $i \in (0, \iota)$. Roughly speaking, condition (37) shows that an FLI equilibrium exists if the aggregate asset supply is plentiful, a ZLB equilibrium exists if assets are scarce, and a PLI equilibrium exists if the asset supply is within an intermediate range.

In an FLI equilibrium, we have $\kappa = \kappa^*$, i.e. capital investment is efficient given q, and DM production satisfies $\frac{u'(q)}{\rho} = 1 + \iota$. Thus, the probability of being impatient θ does not affect the consumption level, and therefore, q is larger in an FLI equilibrium than in an economy without banks. Depositors face no liquidity risk in an FLI equilibrium as banks are able to fully insure them against the risk of turning out to be impatient. This is possible because in an FLI equilibrium, banks are saturated with assets, so that asset prices fall to the point where assets pay the Fisher rate. A large $\mathcal{B}$ increases the supply of illiquid assets and thus makes the existence of an FLI equilibrium more likely. A larger θ makes an FLI equilibrium more likely as well, for two reasons. First, it increases activity in the DM, causing entrepreneurs to invest more in capital, thereby increasing the supply of assets. Second, it decreases banks' demand for illiquid assets due to the smaller share of patient households.

In a ZLB equilibrium, we have $\kappa > \kappa^*$, i.e. capital per DM good produced is higher than optimal. Equation (33) reduces to $\frac{u'(q)}{\rho} = 1 + \frac{\iota}{\theta}$, which is the same condition as that in the economy without banks. However, since κ is higher in a ZLB equilibrium than in an economy without banks (where $\kappa = \kappa^*$), the real price of the DM good ρ is lower, and DM production q is higher than in the no-bank equilibrium.[29] A ZLB equilibrium exists if a relatively low supply of illiquid assets coincides with a relatively large demand for illiquid assets by banks, e.g. due to a large share of patient depositors. Banks' demand for illiquid assets then drives the interest rate down to the zero-lower bound, where banks are indifferent between holding illiquid assets or cash. A ZLB equilibrium is more likely to exist if $\mathcal{B}$ and/or θ are low. Finally, PLI equilibria represent an intermediate case between FLI and ZLB equilibria, in which banks are able to eliminate some but not all of the cost associated with uncertain liquidity needs.[30]

[29] Whether welfare is higher in a ZLB equilibrium than in the no-bank equilibrium is unclear. While q may move closer to its efficient level in a ZLB equilibrium, q is also produced with an inefficient mix of capital and labour.

[30] The effects of changes in ι on which equilibrium the economy is in are subtle. An increase in ι reduces DM activity in all three equilibrium cases, which in turn lowers capital investment by entrepreneurs, decreasing the asset supply. At the same time, a lower q also reduces demand for illiquid assets since it becomes easier for banks to satisfy patient depositors' IC because of lower payouts to impatient depositors. It can be shown that the second

Let us now return to the question whether the IC constraint for impatient depositors (equation (26)) is indeed fulfilled in equilibrium, as we have assumed so far. In ZLB and PLI equilibria the IC constraint for patient depositors binds, which makes it straightforward to show that the IC constraint for impatient depositors will be slack. In an FLI equilibrium, the IC constraint for impatient depositors is satisfied as long as

$$\phi a^b \leq \bar{\mathcal{A}} \equiv (1-\theta)u(q), \tag{38}$$

i.e., as long as the banks' holdings of illiquid assets do not exceed the threshold $\bar{\mathcal{A}}$. Thus, for an FLI equilibrium to exist for some parameters, we need the lower bound on the total supply of illiquid assets for which an FLI equilibrium exists (see (37)) to be below the upper bound on the banks' holdings of illiquid assets given in (38). This is the case if

$$\frac{1}{1+\iota}u'(q)q \leq u(q),$$

which is always strictly satisfied given the properties of $u(q)$. From here on, we assume that banks hold all illiquid assets if $\mathcal{A} \leq \bar{\mathcal{A}}$ and that they hold $\bar{\mathcal{A}}$ if $\mathcal{A} > \bar{\mathcal{A}}$.

While it is immaterial for the results we present in the remainder of this paper, we show in an Online Appendix that all three equilibrium cases may coexist for certain functional forms and parametrisations.[31]

6 Bank Runs

We now consider whether a banking equilibrium is prone to self-fulfilling runs. To define what we mean by this, and what other underlying assumptions we are making, this section discusses the necessary preliminaries. We then discuss whether the banking system is prone to runs without (with) the provision of emergency liquidity in Section 7 (Section 8).

A run denotes a situation where all depositors, both patient and impatient, rush to the banks in order to withdraw their deposits in the DM. We treat runs as zero-probability events, which therefore have no effect on the steady-state allocation studied previously. Throughout Sections 6-8, we will use subscript S to denote the steady-state values of variables. Following Diamond and Dybvig (1983) and much of the banking literature, we assume sequential service, i.e. depositors arrive at their bank in random order in the DM and need to be paid out on the spot. Depositors who wish to redeem in the DM form a queue, with each depositor being assigned each place in the queue with identical probability independent of a depositor's true type. Patient depositors run iff they expect that withdrawing in the DM gives them a strictly

effect weakly dominates in FLI equilibria. The same is not true in ZLB and PLI equilibria, where an increase in ι has the additional effect of decreasing κ, which has a negative effect on asset supply.

[31] The online appendix can be found at:

`https://drive.google.com/file/d/1kPs-2q2VPZbgbamQcB0j744-XPn5ZROD/view` higher payout compared to not withdrawing.[32] We will say that the banking system is *fragile* if steady state DM payouts d_S^I are strictly higher than CM payouts in a situation where all depositors run.

In a run, some cash ends up in the hands of running patient depositors, who will not spend it until the next CM. This means that impatient depositors will hold less cash in aggregate than in the no-run steady state. We denote by M^{RE} the total cash paid out to impatient depositors in a run. Without loss of generality, we will take it as given that $0 < M^{RE} \leq \theta d_S^I$, where θd_S^I is the total cash paid out to impatient depositors in the steady state. Further, we denote by χ the share of entrepreneurs that choose to produce in the DM and thus repay their loans in a run.

Market clearing in the DM requires that aggregate real spending on the DM good by impatient depositors equals aggregate real revenues for entrepreneurs. Active entrepreneurs optimally choose their supply given the real price of the DM good, ρ. Denoting $q^e(\rho)$ as the supply by an active entrepreneur, DM clearing in a run therefore requires

$$\frac{1+\pi_S}{1+\pi}\frac{M^{RE}}{\theta d_S^I} = \frac{\chi \rho q^e(\rho)}{\rho_S q^e(\rho_S)}, \tag{39}$$

i.e. the change in aggregate real cash holdings by impatient households relative to the steady state (the LHS) equals the relative change in aggregate real revenues for entrepreneurs (the RHS).

Given that the economy reverts to its steady state after a run, the inflation rate is unaffected as long as no additional money is injected in a run. In this case, a run does not affect the real price of the DM good, and we obtain a particularly simple expression for the share of active entrepreneurs:

Proposition 2. *If the amount of currency in circulation does not change when a run occurs, the real purchasing power of money remains unchanged ($\pi = \pi_S$). Then,*

(i) the real price of DM goods remains at the steady state level ($\rho = \rho_S$), and

(ii) the share of active entrepreneurs equals $\chi = \dfrac{M^{RE}}{\theta d_S^I}$.

Proof. The first part of this proposition follows because runs are assumed to be zero-probability events, and because steady-state inflation is determined through the money growth rate. To see why ρ cannot deviate from its steady state value, recall first that, given k_S and ρ_S, entrepreneurs' payoff-maximizing DM labour effort yields them a payoff of zero. Given that the run does not affect the inflation rate, the real indebtedness of entrepreneurs stays the same, which in turn implies that ρ cannot fall below its steady state value. The reason for the latter is that a decrease in the DM price shifts the payoff schedule downwards, such that any DM price below ρ_S would imply a strictly negative payoff from production to entrepreneurs. A real DM price below ρ_S would thus imply zero production of the DM good, which cannot be consistent with market clearing (39). Next, $\rho > \rho_S$ cannot be consistent with DM clearing (39) either, as it would lead

[32] We assume patient depositors do not redeem in the DM (they 'stay home') in case of indifference.

to an increase in the aggregate supply of DM goods while the aggregate demand for the DM good decreases. Finally, given $\rho = \rho_S$ and $\pi = \pi_S$, item (ii) in Proposition 2 follows immediately from (39). ■

For banks, the effective gross nominal return on loans in a run equals $\chi(1+i_S)$.[33] Notice that the return to an individual bank's portfolio depends on the redemption behaviour of depositors at all banks, making runs inherently systemic events. We want to highlight here that the link between the occurrence of a run and the return on illiquid assets captured by item (ii) in Proposition 2 is the main innovation in our model compared to the existing literature.[34]

Consider now the more general case where a run may affect the price level and hence the real price of the DM good, ρ. Remember that this is relevant only for the analysis in Section 8, where we consider emergency liquidity provision by the government. The share of entrepreneurs who repay their loan then depends both on the cash paid out to impatient depositors relative to the steady state as well as on changes in the capital share caused by the run:[35]

Proposition 3. *Given any real price of the DM good ρ in a run, the share of active entrepreneurs equals*

$$\chi = \min\left\{\frac{\alpha(\rho)}{\alpha(\rho_S)}\frac{M^{RE}}{\theta d_S^I}, 1\right\}, \tag{40}$$

where

$$\alpha(\rho) = \frac{\rho q^e(\rho) - c(q^e(\rho), k_S)}{\rho q^e(\rho)} \tag{41}$$

denotes the capital share, i.e. the share of an active entrepreneur's revenue left after compensating the entrepreneur for his labour disutility of production.

Proof. Active entrepreneurs choose their supply to maximise real profits before repayments, i.e. $\rho q^e - c(q^e, k_S)$. Supply by an active entrepreneur is therefore set according to $q^e = q^e(\rho) \equiv c_q^{-1}(\rho, k_S)$, which implies that $q^e(\rho)$ is a strictly increasing function of ρ. Based on the maximised profits before repayment, entrepreneurs determine whether they produce in the DM and repay their loans, or default and produce nothing. An entrepreneur is willing to produce and repay his loan if and only if

$$\max_{q^e}\{\rho q^e - c(q^e, k_S)\} \equiv \rho q^e(\rho) - c(q^e(\rho), k_S) = \rho\alpha(\rho)q^e(\rho) \geq \beta\frac{1+i_S}{1+\pi}k_S, \tag{42}$$

[33] Since runs and thus defaults are unexpected, $\tilde{i}_\ell = i_\ell = i_S$ still holds.

[34] Note that we do not allow entrepreneurs to renege on their debt contracts. If entrepreneurs could renege, Proposition 2 would not necessarily hold anymore, as entrepreneurs might be willing to sell DM goods at lower prices if their repayment burden is reduced. While this could mitigate the macroeconomic effects of a bank run, it would not help the banking system in stopping or preventing runs, since as long as entrepreneurs repay less than was initially promised, run incentives for patient depositors remain intact.

[35] One implication of the result in Proposition 3 is that changes in inflation have no direct effect on entrepreneurs' incentive to default. The reason is that inflation acts as a double-edged sword: a reduction in the value of cash reduces both the real purchasing power of the impatient depositors and the real debt burden of the entrepreneurs. Notice that inflation may in principle still affect defaults indirectly through its effect on ρ and hence on α.

i.e. if and only if the maximized real profits before repayment weakly exceed the real debt burden. If the relationship imposed by (42) holds with strict inequality, entrepreneurs strictly prefer to be active in the DM and thus repay their loans, whereas if it holds with equality, entrepreneurs are indifferent between repayment and default. Exploiting that (42) holds with equality in the steady state, we find

$$\frac{\rho\alpha(\rho)q^e(\rho)}{\rho_S\alpha(\rho_S)q^e(\rho_S)} \geq \frac{1+\pi_S}{1+\pi}, \quad \text{with equality if } \chi < 1, \tag{43}$$

i.e. to incentivize entrepreneurs to produce in the DM, their real revenue left after compensating them for their labour disutility cannot decrease by more than the reduction in the real debt burden caused by an increase in inflation. Combining (39) and (43) then leads to the result in Proposition 3. ■

Whether the banking system is fragile ex ante depends on the banks' and the government's ex post reaction to runs. We start in Section 7 with the banks' potential ex post reactions by focusing on two measures: deposit freezes and penalties on early redemptions. These are arguably the most widely used measures to stop or prevent runs that do not rely on a public backstop such as a lender of last resort or deposit insurance. In line with the bulk of the bank run literature, we assume the banking system reacts to runs as a single, consolidated entity.[36] The speed at which the queue is served is the same at all banks, i.e. whenever a bank has served some fraction of queuing depositors in the DM, all other banks will have served the same fraction of their queue. In Section 8, we widen the scope of our analysis by also considering public provision of emergency liquidity, where the government stands ready to purchase illiquid assets from the banks, possibly at a discount.

We assume for the remainder of the paper that all banks and the government realise simultaneously that a run is underway after a fraction $\lambda \in [0, \theta]$ of depositors have withdrawn in the DM, at which point banks may jointly impose a deposit freeze and/or penalties on redemptions, or the government starts to inject emergency liquidity.[37] Since the share of impatient depositors is known to equal θ, banks always know that a run is going on if more than a fraction θ of depositors wish to redeem in the DM. $\lambda = 0$ means that the run is immediately spotted, i.e. after a measure zero of patient depositors have withdrawn.

[36] One may imagine that banks jointly commit to reacting in a certain way if they realize that a run is underway. Ennis and Keister (2009, 2010) interpret the concerted action by banks as the result of a centralized banking authority stepping in once a systemic run has started, where the banking authority could be regarded as a reduced-form representation of the banking system together with the relevant regulatory agencies. Importantly, unlike the banking authority in Ennis and Keister (2009, 2010), the consolidated banking system does not suffer from limited commitment in our model.

[37] We do not model explicitly how banks realise that a run is underway. As has been shown by Andolfatto et al. (2017) and Cavalcanti and Monteiro (2016), mechanisms exist which may help banks to learn about a bank run. We consider the case of $\lambda = 0$ to model the effects of such mechanisms in a reduced-form way. Note also that, as long as the government does not provide emergency liquidity, banks will impose a full deposit freeze by default once they run out of cash in the DM since assets cannot be liquidated prematurely.

In case banks impose a partial freeze or penalties on redemptions after realising that a run is underway, depositors who were not among the first λ to show up can choose whether they still want to redeem in the DM or whether they want to leave the queue and be paid out in the CM instead. If patient depositors leave the queue once the measures are in place, we will say that these measures *stop runs.* Finally, if the banking system is not fragile given that banks impose deposit freezes and / or redemption penalties and / or the government provides emergency liquidity after observing a run (off the equilibrium path), we will say that these measures *prevent runs.*

7 Deposit Freezes and Redemption Penalties

Suppose that, upon realising that a run is underway, banks can freeze any fraction $1-\eta \in (0,1]$ of deposits. If deposits are (partially) frozen, depositors can only redeem a fraction η of their deposit in the DM, thus receiving a DM payout of ηd_S^I, while the remaining part of the deposit is locked in until the CM. In the CM, depositors are paid out pro-rata, i.e. if depositors who did not withdraw anything in the DM receive some amount d in the CM, then depositors who redeemed a fraction η of their deposit in the DM receive $(1-\eta)d$ in the CM. The standard full deposit freeze studied by Diamond and Dybvig (1983) and others corresponds to $\eta = 0$ and $\lambda = \theta$.

Recall that at the point in time when banks impose a deposit freeze, a fraction λ of depositors have already withdrawn their deposit. Denoting ω as the share of depositors that successfully redeem in the DM after a partial freeze with $\eta > 0$ has been imposed, the banks' liquidity constraint implies

$$\omega \leq \overline{\omega}(\eta) \equiv \min\left\{1-\lambda, \frac{m_S^b - \lambda d_S^I}{\eta d_S^I}\right\}. \tag{44}$$

For future reference, we define

$$\overline{\eta} \equiv \min\left\{\frac{m_S^b - \lambda d_S^I}{(1-\lambda)d_S^I}, 1\right\} \quad \text{and} \quad \overline{\overline{\eta}} \equiv \min\left\{\frac{\overline{\eta}}{\theta}, 1\right\}, \tag{45}$$

so that $\overline{\omega}(\overline{\eta}) = 1-\lambda$ and $\overline{\omega}\left(\overline{\overline{\eta}}\right) = \theta(1-\lambda)$. That is, after realizing that a run is underway, banks' cash reserves are just sufficient to convert a fraction $\overline{\eta}$ of all remaining deposits into cash in the DM; and cash reserves are just sufficient to convert a fraction $\overline{\overline{\eta}} > \overline{\eta}$ of all remaining deposits held by *impatient* depositors into cash.[38]

In addition to partially freezing deposits, banks may impose a penalty (or haircut) on early redemptions, which we denote by σ. Depositors who redeem a fraction η of their deposit in the DM then forgo a fraction $\sigma \in [0, 1-\eta]$ of their deposit. That is, if depositors who do not redeem in the DM receive some amount d in the CM, those redeeming a fraction η of their deposit receive $(1-\eta-\sigma)d$ in the CM.[39] Note that we express banks' payout policy in terms of (σ, η) for notational convenience. An equivalent formulation would be to say that, after realizing a run is

[38] Since $m_S^b < d_S^I$ holds in all banking equilibria, we have $\overline{\eta} < 1$ as well as $\frac{\partial \overline{\eta}}{\partial \lambda} < 0$ and $\frac{\partial \overline{\overline{\eta}}}{\partial \lambda} \leq 0$.

[39] The term 'partial suspension' has sometimes been used in the literature to describe the case with $\sigma = 1-\eta$.

underway, banks pay some amount (d^{DM}, d^{CM}) to depositors who withdraw in the DM, where d^{DM} is paid out in the DM and d^{CM} in the DM, and banks then distribute the CM revenue not pledged to depositors who withdrew in the DM equally among depositors who did not withdraw in the DM.

We denote by d_R^P the CM payouts to depositors who did not withdraw in the DM after banks imposed a given combination of deposit freezes and redemption penalties. The banks' solvency constraint in a run then writes

$$m_S^b + (1+i_S)(b_S^b + \chi \ell_S^b) = \lambda d_S^I + \omega(\eta d_S^I + (1-\eta-\sigma)d_R^P) + (1-\lambda-\omega)d_R^P, \tag{46}$$

so that we can express CM payouts as:

$$d_R^P(\chi, \omega; \eta, \sigma) = \frac{[m_S^b - (\lambda + \eta\omega)d_S^I] + (1+i_S)(b_S^b + \chi \ell_S^b)}{1 - \lambda - \omega(\eta + \sigma)}. \tag{47}$$

The numerator in (47) equals the total resources of a bank in the CM given that a run occurred in the DM; the first term in the numerator equals left-over cash not paid out in the DM and the second term equals the proceeds from a bank's portfolio of government bonds and loans. The denominator is the measure of outstanding deposits in the CM.

7.1 Deposit Freezes

In this subsection we will focus on pure deposit freezes, setting $\sigma = 0$. Findings in the previous literature suggest that deposit freezes are effective in preventing runs in economies without (extrinsic) aggregate uncertainty and with full commitment. We show that this is not the case once general equilibrium effects of deposit freezes are taken into account.

We start by noting that a partial deposit freeze with $\eta \in (0,1)$ cannot stop a run that has already started, in the following sense: if redeeming the entire deposit is patient depositors' (strictly) best response, then so is redeeming any fraction $\eta > 0$ of the deposit. To see this, note that redeeming the allowed fraction η is patient depositors' best response iff $\eta d_S^I + (1-\eta)d_R^P(\cdot) > d_R^P(\cdot)$, which is equivalent to $d_S^I > d_R^P(\cdot)$.

To understand whether the banking system is fragile, we need to determine how many entrepreneurs default in a run. Consider deposit freezes with $\eta \geq \overline{\eta}$, in which case banks pay out their entire cash holdings to redeeming depositors in a run. In this case, a share λ of depositors manages to withdraw in full in a run, a fraction $\overline{\omega}(\eta)$ manages to withdraw a fraction η of their deposit, while the remaining fraction $1 - \lambda - \overline{\omega}(\eta)$ of depositors cannot redeem anything in the DM. By a law of large numbers, a fraction θ of depositors who manage to withdraw in the DM will be impatient, which implies that the total cash paid out to impatient depositors equals θm_S^b. By Proposition 2, we then have

$$\chi = \frac{m_S^b}{d_S^I} \in (0,1) \tag{48}$$

We find it useful to explicitly distinguish between deposit freezes, where funds are 'frozen' but not lost, and penalties on redemptions.

for $\eta \geq \overline{\eta}$.[40] Combining (47) and (48), we get that the banking system is fragile under any deposit freeze satisfying $\eta \geq \overline{\eta}$ iff

$$d_S^I > d_R^P\left(\frac{m_S^b}{d_S^I}, \overline{\omega}(\eta); \eta, 0\right) \quad \Leftrightarrow \quad d_S^I > m_S^b + (1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right). \tag{49}$$

In words, the banking system is fragile under deposit freezes whenever DM payouts are higher than the value of banks' portfolios after taking into account loan defaults caused by a run. This leads us to the following result:

Proposition 4. *For any $\lambda \in [0, \theta]$ and any deposit freeze $\eta \in [0, 1)$, the banking system is fragile in all zero lower bound (ZLB) and partial liquidity insurance (PLI) equilibria, as well as in full liquidity insurance (FLI) equilibria where the real aggregate asset supply satisfies $\mathcal{A}_S < \min\{\bar{\mathcal{A}}, \bar{\mathcal{A}}^F\}$, where $\bar{\mathcal{A}}^F \equiv (1-\theta)q_S(\rho_S + \kappa^*\theta)$.*

Proof. Consider first ZLB and PLI equilibria. In these equilibria, we have $d_S^I = d_S^P = m_S^b + (1 + i_S)(b_S^b + \ell_S^b)$. Since $m_S^b/d_S^I < 1$, it follows immediately from condition (49) that the banking system is fragile in ZLB and PLI equilibria. Consider next FLI equilibria. In FLI equilibria, we have $\theta d_S^I = m_S^b$. Substituting this and $d_S^I = p_S q_S$ into condition (49) and rearranging yields the condition $(1-\theta)p_S q_S > (1+\imath)(b_S^b + \theta \ell_S^b)$. Substituting $\rho \equiv \frac{\phi}{1+\iota}p$, $\phi \ell^b = \theta q \kappa$, $\mathcal{A} \equiv \phi(b^b + \ell^b)$ and using the fact that $\kappa = \kappa^*$ in an FLI equilibrium yields that, assuming banks hold all assets in the economy (which is the case if $\mathcal{A}_S \leq \bar{\mathcal{A}}$), they are fragile iff $\mathcal{A}_S \leq \bar{\mathcal{A}}^F$.[41] ∎

The fragility of the banking system in ZLB and PLI equilibria is closely related to the fact that the incentive constraint for patient depositors binds in these equilibria. Even slight losses on banks' investments cause CM payouts to fall below DM payouts, making it optimal for patient depositors to run. In FLI equilibria, patient depositors receive strictly higher payouts than impatient depositors in the steady state, such that banks have a buffer to absorb a certain amount of losses on their loans. Losses caused by a run on banks' asset portfolios are decreasing in the share of government bonds in banks' portfolios and increasing in the share of impatient depositors; keeping all else the same, the minimum buffer sufficient to prevent runs, $\bar{\mathcal{A}}^F$, is thus decreasing in $\frac{\mathcal{B}}{q}$ and increasing in θ.

A remarkable result of Proposition 4 is that as long as banks' buffer to absorb losses is sufficiently small, deposit freezes can neither prevent runs nor stop them from starting, independent of how quickly banks can react to runs and independent of the fraction of deposits they freeze. Even if banks can impose a (partial) deposit freeze immediately after a run has started, the fact

[40] If banks instead impose a stricter freeze with $\eta < \overline{\eta}$, not all cash will be paid out in the DM in a run. The total cash paid out to impatient depositors in a run will then be strictly below θm_S^b, so that χ will be lower than in (48). Since the incentive to run for patient depositors is increasing in the share of defaulting entrepreneurs $(1-\chi)$, a freeze with $\eta < \overline{\eta}$ unambiguously exacerbates incentives to run compared to a freeze with $\eta \geq \overline{\eta}$, which allows us to focus on $\eta \geq \overline{\eta}$ without loss.

[41] If $\mathcal{A} > \bar{\mathcal{A}}$ then not all assets in the economy are held by banks. Whether banks are fragile depends then additionally on how illiquid assets are allocated between households and banks; the larger the share of government bonds (loans) in banks' portfolios is, the less (more) likely it is that banks are fragile.

that impatient depositors can withdraw less cash due to the freeze makes running optimal for patient depositors. This latter point is the major difference between our paper and other papers in the literature, where the final return on illiquid assets is given exogenously and thus can be protected by freezing deposits, as this prevents banks from early liquidation of illiquid assets. In our model, the return on illiquid assets is tied to the conditions in the real economy. While a (partial) freeze saves some of the banks' assets for late withdrawals, it also implies that even less cash ends up in the hands of impatient depositors who would like to consume early and thus hurts aggregate demand, which in turn lowers the return on illiquid assets and increases the incentive for patient depositors to run.

The reason that the exact values of λ and η have no effect on whether deposit freezes can prevent runs is that DM activity and loan defaults in a run depend only on the total amount of cash paid out to impatient depositors. This payout equals θm_S^b independent of λ and η. However, different values of λ and η do affect the distribution of cash among impatient depositors in a run. Keeping all else the same, a higher λ makes payouts to impatient depositors in a run more unequal since a larger share of depositors manages to redeem their deposit in full in a run, implying that less cash will be left for those impatient depositors not among the first λ to arrive. Similarly, keeping λ fixed, ex post payouts in a run will become more unequal as banks increase η: while those (relatively) early in line can withdraw a larger amount, a larger share of depositors cannot withdraw anything in the DM since banks run out of cash before all depositors can be served.

Assuming that preventing runs with pure deposit freezes is not feasible, how should banks set η in order to minimize ex post welfare losses caused by a run? Given that aggregate DM production and loan defaults are the same for all $\eta \geq \overline{\eta}$, and given that DM preferences are strictly concave, it is straightforward that the best banks can do is to distribute all the cash as evenly as possible to the impatient depositors who did not manage to withdraw their deposit in full. This leads to the following proposition, which we state without separate proof:

Proposition 5. *With pure deposit freezes, the ex post welfare loss of a run is minimised if banks impose a deposit freeze with $\eta = \overline{\eta}$ once they realize that a run is underway.*

7.2 Penalties on Redemptions

We now turn to the case where, in addition to freezing part of deposits, banks can impose penalties on early redemptions after noticing that a run is underway. To streamline the exposition, we will take it as given that banks set $\eta \in [\overline{\eta}, \overline{\overline{\eta}}]$.[42]

A given mix of deposit freezes and penalties on redemptions is said to *stop* a run if a patient depositor (who is not among the first λ of depositors in the queue) is better off not withdrawing

[42]Footnote 40 discusses why ignoring $\eta < \bar{\eta}$ is without loss; for $\eta > \overline{\overline{\eta}}$, banks' cash reserves would not be sufficient to pay out all impatient depositors in the DM, even if patient depositors stop running after banks impose the partial freeze. Banks could then always adjust (η, σ) in such a way that CM payouts and hence run incentives are not affected but that the available cash is distributed more evenly among impatient depositors in the event of a run. See also Footnote 45 below.

in the DM even if (hypothetically) all other depositors continue running after banks impose (η, σ).[43] If all depositors continue to redeem in the DM after banks impose (η, σ), we have that $\omega = \overline{\omega}(\eta)$ and $\chi = m_S^b / d_S^I$ (for the latter, see the discussion in the previous subsection). From (47), we obtain that CM payouts in such a situation equal

$$d_R^P \left(\frac{m_S^b}{d_S^I}, \overline{\omega}(\eta); \eta, \sigma \right) \equiv \underline{d}_R^P(\eta, \sigma). \tag{50}$$

Banks' reaction to a run thus stops the run iff

$$\eta d_S^I + (1 - \eta - \sigma) \underline{d}_R^P(\eta, \sigma) \leq \underline{d}_R^P(\eta, \sigma) \text{ with } \eta \in [0, 1] \text{ and } \sigma \in [0, 1 - \eta]. \tag{51}$$

Proposition 6. *There always exists a policy (η, σ) with $\eta > \overline{\eta}$ that stops a run.*

This proposition shows that banks can always stop a run, even without completely freezing deposits (which trivially stops the run). Specifically, they can do so by setting η low enough and σ high enough. We show in Appendix A.5.1 that the condition in (51) can be reformulated as a lower bound on the redemption penalty, $\sigma \geq \underline{\sigma}(\eta)$, and that there exists a value $\tilde{\eta}^{max} > \overline{\eta}$ such that $\underline{\sigma}(\eta) \in [0, 1 - \eta]$ for all $\eta \in [\overline{\eta}, \tilde{\eta}^{max}]$.

Banks' reaction to runs *prevents* a run if the prospect of banks imposing (η, σ) after realizing that a run is underway eliminates patient depositors' incentives to redeem d_S^I in the first place. While stopping a run ex post is necessary to prevent a run ex ante, it is not sufficient. Different to pure deposit freezes, stopping and preventing runs are thus not equivalent with redemption penalties.

Suppose banks' reaction to runs satisfies condition (51), such that it stops the run. Suppose also for the moment that all impatient depositors redeem after banks impose (η, σ). We then have that $\omega = \theta(1 - \lambda)$ and $M^{RE} = \theta \left[\lambda + (1 - \lambda)\eta \right] d_S^I$. By Proposition 2, the share of non-defaulting entrepreneurs then equals

$$\chi = \lambda + (1 - \lambda)\eta \tag{52}$$

and, by (47), CM payouts are given by

$$d_R^P\big(\lambda + (1 - \lambda)\eta, \theta(1 - \lambda); \eta, \sigma\big) \equiv \overline{d}_R^P(\eta, \sigma). \tag{53}$$

Of course, impatient depositors must be willing to redeem in the DM (and incur the redemption penalty) rather than to wait until the CM, which requires[44]

$$u(\eta q_S) + \beta \phi_+ (1 - \eta - \sigma) \overline{d}_R^P(\eta, \sigma) \geq \beta \phi_+ \overline{d}_R^P(\eta, \sigma) \text{ , with } \eta \in [0, 1] \text{ and } \sigma \in [0, 1 - \eta], \tag{54}$$

[43]Since withdrawal decisions of patient depositors are strategic complements, this is the same as saying that not withdrawing ηd_S^I in the DM must be the dominant strategy for patient depositors.

[44]Since defaults are decreasing (and hence CM payouts are increasing) in the number of impatient depositors who redeem in the DM, withdrawal decisions of impatient depositors in the DM are strategic substitutes. Condition (53) is thus the same as saying that withdrawing in the DM after banks have imposed the redemption penalty must be the dominant strategy for impatient depositors.

where ϕ_+ denotes the value of money next period. In Appendix A.5.2, we show that the condition in (54) can be rewritten as an upper bound on the redemption penalty, $\sigma \leq \overline{\sigma}(\eta)$. Finally, given that banks' reaction to runs satisfies conditions (51) and (54), patient depositors' incentive to run in the first place will be eliminated iff

$$d_S^I \leq \overline{d}_R^P(\eta, \sigma) \text{ with } \eta \in [0, 1] \text{ and } \sigma \in [0, 1 - \eta]. \tag{55}$$

In Appendix A.5.3, we show that the condition in (55) can be rewritten as a lower bound on the redemption penalty, $\sigma \geq \hat{\underline{\sigma}}(\eta)$. Notice that the lower bounds on the redemption penalty resulting from conditions (51) and (55) are distinct. The first says that the redemption penalty must be high enough to deter patient depositors from running after the penalty has been imposed; the latter says that the redemption penalty incurred by impatient depositors must be high enough to deter patient depositors from running in the first place.

In sum, in order to prevent runs, banks' reaction to runs, (η, σ), must satisfy conditions (51), (54) and (55).[45]

Proposition 7. *Even with $\lambda = 0$, there may be no policy (η, σ) that prevents a run.*

We prove this proposition by providing in Appendix A.7 an example where there is no (η, σ) preventing runs even if $\lambda = 0$. However, in many cases, as long as λ is low enough, banks can prevent runs by setting (η, σ) appropriately; we provide an example for this case below. Note in particular that the redemption penalty σ has a redistributive function, in the sense that it redistributes funds from (impatient) depositors who redeem in the DM to (patient) depositors who redeem in the CM. If (η, σ) is such that defaults caused by a run are kept sufficiently low and the redistribution towards patient depositors implemented by the redemption penalty in the event of a run is sufficiently large, patient depositors' incentives to participate in the run in the first place will be eliminated.

Minimizing Welfare Losses Caused by Runs

As with pure deposit freezes, we can ask how banks should set (η, σ) such as to minimize ex post losses caused by runs, assuming preventing runs is not feasible. Note first that if banks do not stop the run, the situation is equivalent to the one with pure deposit freezes. Independent of the exact value of (η, σ), the total cash paid out to impatient depositors in a run then equals θm_S^b, which in turn pins down DM activity and loan defaults. However, with penalties on redemptions,

[45] In conditions (55) and (54), we have implicitly required that banks' reaction to runs should be such that all impatient depositors are able and willing to redeem in the DM. This is without loss of generality, as none of the constraints that need to be fulfilled in order to prevent runs could be relaxed by setting (η, σ) in such a way that some impatient depositors cannot or do not want to redeem in the DM. If (η, σ) is such that the run is stopped and the remaining impatient depositors do not receive uniform payouts, then banks could always change (η, σ) in such a way as to distribute the same aggregate payout to impatient depositors uniformly among them. This would increase the payoff for impatient depositors (as a result of strictly concave DM preferences) and would leave run incentives for patient depositors unaffected since CM payouts depend only on the aggregate payout to impatient depositors.

banks can do better by deterring patient depositors from withdrawing once the run has been detected. This allows to increase the aggregate amount of cash paid out to impatient depositors.

The fact that we restrict attention to $\eta \leq \overline{\overline{\eta}}$ means that all impatient depositors who are not among the first λ of depositors to arrive in a run receive identical DM payouts whenever (η, σ) is such that patient depositors are deterred from withdrawing once the run is detected. Minimizing welfare losses caused by a run is then equivalent to maximizing DM activity in a run, which is achieved by maximizing the cash paid out to impatient depositors (i.e. maximizing η) subject to the relevant constraints (51) and (54). Specifically, DM activity in a run will be maximized by setting η to the highest level within $(\overline{\eta}, \overline{\overline{\eta}}]$ consistent with stopping the run while ensuring that impatient depositors are willing to withdraw. We denote this level with $\eta^{\max}$.

Proposition 8. *Suppose $\lambda < \theta$. Let $\eta^{\max} \equiv \min\{\overline{\overline{\eta}}, \tilde{\eta}^{max}, \hat{\eta}^{max}\}$, where $\tilde{\eta}^{max}$ is the unique value of η solving $\underline{\sigma}(\eta) = 1 - \eta$, and $\hat{\eta}^{max}$ is the unique strictly positive value of η solving $\underline{\sigma}(\eta) = \overline{\sigma}(\eta)$. Then $\eta^{max} \in (\overline{\eta}, 1)$ and the ex post welfare loss of a run is minimized if banks set $\eta = \eta^{\max}$ and $\sigma \in [\underline{\sigma}(\eta^{\max}), \overline{\sigma}(\eta^{\max})]$.*

We refer to Appendix A.6 for the proof and the derivations related to Proposition 8. Intuitively, the maximum amount that can be paid out to impatient depositors in the DM in case of a run can be constrained for three different reasons. It can be constrained by the fact that banks have limited cash left once they notice that a run is underway (in which case $\eta^{\max} = \overline{\overline{\eta}}$), it can be constrained by the fact that patient depositors must be deterred from continuing running while the redemption penalty cannot exceed the fraction of frozen deposits (in which case $\eta^{\max} = \tilde{\eta}^{max}$), or it can be constrained because patient depositors must be deterred from continuing running while impatient depositors must still find it attractive to withdraw (in which case $\eta^{\max} = \hat{\eta}^{max}$).

Example where Runs can be Prevented

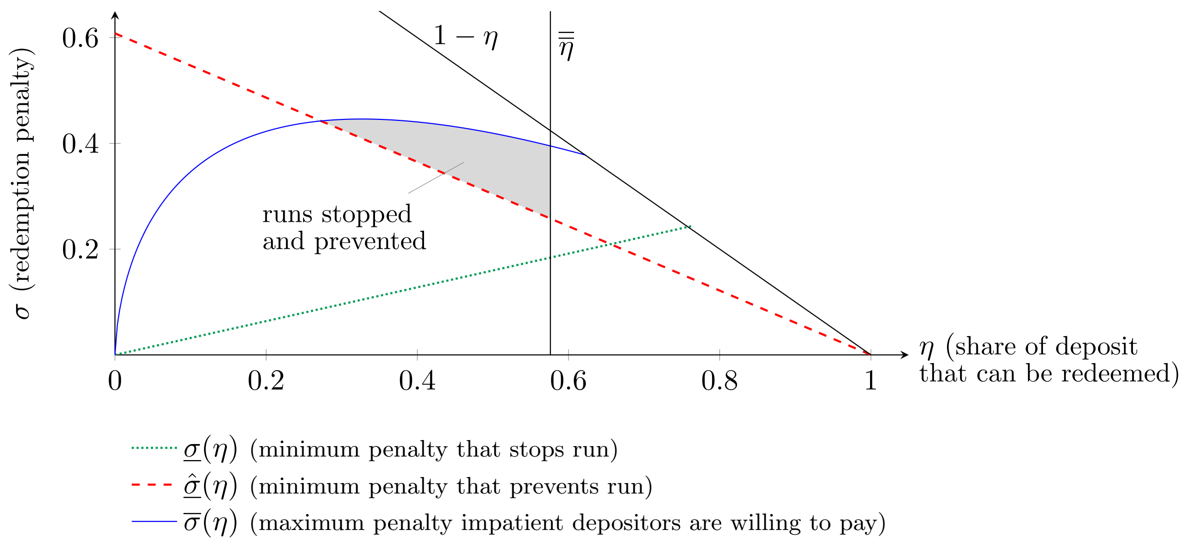

Figure 1: Example where runs can be prevented.

Figure 1 provides a numerical example illustrating how (η, σ) needs be set in order to both stop and prevent runs. We assume that the DM utility function equals $u(q) = q^{\nu}$ and that DM production follows a Cobb-Douglas technology, i.e.

$$f(k, h) = k^{\alpha} h^{1-\alpha}. \tag{56}$$

The parameters used for this example are summarized in Table 1. They are such that the economy is in a zero lower bound steady state equilibrium.[46] The grey area in Figure 1 depicts the set of (σ, η) that both stop and prevent runs. The fact that $\underline{\sigma}(\eta)$ is strictly increasing in η while $\underline{\hat{\sigma}}(\eta)$ is strictly decreasing in η is not specific to the example at hand; as we show in Appendix A.5, this is always the case if banks are fragile under pure deposit freezes. Higher DM payouts η increase the incentive for patient depositors to continue running, which is why the redemption penalty required to stop the run increases in η. Furthermore, given that the redemption penalty stops runs, higher DM payouts to impatient depositors lead to more DM activity and less defaults, thereby decreasing patient depositors' incentives to participate in the run in the first place. For this reason, the minimum redemption penalty necessary to prevent a run is decreasing in η. This is a characteristic property of our model – increasing payouts to impatient depositors can lower patient depositors' incentives to run due to the tight connection between banks' liquidity provision and the return to banks' assets.

Table 1: Parameter values for Figure 1.

α	ν	θ	n	λ	ι	$\mathcal{B}$	q_S	ρ_S	κ_S
0.6	0.65	0.6	0.6	0.45	0.05	0	0.037	1.904	1.199

8 Emergency Liquidity

We now turn to the role of policy in stopping and preventing runs. For expositional convenience, we assume the government provides emergency liquidity to the banking system by purchasing assets from banks. We show in Appendix A.11 how the intervention can be reinterpreted as the provision of collateralized loans to banks without changing any of the results.

We assume that the government cannot increase its real indebtedness, which restricts the ability to provide liquidity in real terms.[47] Specifically, we have

$$\frac{M + (1 + i_S)B}{M_S + (1 + i_S)B_S} \leq \frac{1 + \pi}{1 + \pi_S}, \tag{57}$$

where $M + (1 + i_S)B$ denotes the government's nominal liabilities at the beginning of the next CM, just after entrepreneurs have repaid loans, and $1+\pi = \phi_t/\phi_{t+1}$ is the inflation rate given the

[46] We show in the Online Appendix that with a Cobb-Douglas production function, $\mathcal{B} = 0$ and $\theta < \frac{1}{1+\alpha}$, the economy is in a ZLB equilibrium.

[47] This constraint can be motivated in various ways. For instance, it may capture the central bank's inability to raise taxes, or it may capture political constraints such as a limit on real government debt.

government's provision of emergency liquidity. The condition imposes that nominal liabilities cannot grow faster than inflation, implying that the real value of the government's liabilities is bounded from above.[48] It implies that the provision of emergency liquidity will be inflationary whenever it leads to an increase in nominal government liabilities beyond the point at which the assets purchased by the government mature.

To provide emergency liquidity, the government stands ready to convert government bonds with a gross face value (i.e. principal plus interest) of one dollar into $\delta_b \leq 1$ dollars of cash. Similarly, a loan with a promised gross repayment of one dollar can be converted into $\delta_\ell \leq 1$ dollars of cash. Here $\delta_b < 1$ and $\delta_\ell < 1$ can be interpreted as discounts. For simplicity, we assume throughout this section that $\lambda = \theta$, i.e. the government provides emergency liquidity after a measure θ of depositors has withdrawn. Furthermore, we assume $\delta_\ell < \delta_b$, as bonds are safer assets than loans. Based on the discounts and the banks' asset holdings, the total amount of liquidity a bank can access once the run is detected is

$$m_S^b - \theta d_S^I + \delta_b(1+i_S)b_S^b + \delta_\ell(1+i_S)\ell_S^b. \tag{58}$$

Following the results established in Section 7, we assume that once banks notice a run is underway, they charge a redemption penalty that reflects the cost of obtaining emergency liquidity:[49]

$$\eta = 1 - \sigma = \min\left\{\frac{m_S^b - \theta d_S^I + \delta_b(1+i_S)b_S^b + \delta_\ell(1+i_S)\ell_S^b}{(1-\theta)d_S^I}, 1\right\}. \tag{59}$$

In the following, we limit our attention to cases where σ is sufficiently small so that impatient depositors still withdraw in the DM even if they are subject to the redemption penalty. Regardless of whether the run continues or stops after a measure θ of depositors have withdrawn, the aggregate amount of cash held by impatient depositors then equals

$$M^{RE} = \theta\left[\theta + \eta(1-\theta)\right] d_S^I. \tag{60}$$

8.1 Inflation and Defaults with Emergency Liquidity

Let us first consider the implications of the provision of emergency liquidity for inflation. These depend on the type and quantity of assets purchased by the government. Let τ_b and τ_ℓ denote the fraction of bonds and loans that banks sell to the government, and let $\mathcal{I} \in \{0, 1\}$ be an indicator variable that takes a value of one when the run continues after a measure θ of depositors have

[48] The reason why condition (57) is written as an inequality is that the government can always increase inflation by injecting additional money in the CM. However, once money has been injected in the DM, the government cannot withdraw it from circulation in the CM by increasing lump-sum taxation. Note also that the government may use part of the money injected in the DM to purchase government bonds from banks, such that B may be lower than B_S.

[49] If the bank charges a redemption penalty (59), depositors who redeem internalize the cost of obtaining emergency liquidity. If the run continues, all remaining $1-\theta$ depositors receive the same payout. Note that this is the lowest redemption penalty (i.e. the highest η) that may potentially stop the run. If banks set a higher η, then a run cannot be stopped once it has started since the depositors in the end of the queue receive nothing in the DM and also nothing in the CM (the bank has sold off all its assets before all depositors are served).

withdrawn. If $m_S^b - \theta d_S^I < \eta(1-\theta)[\theta + \mathcal{I}(1-\theta)]d_S^I$, the bank needs to sell assets in order to meet withdrawals, and the asset sales are such that

$$\eta(1-\theta)\left[\theta + \mathcal{I}(1-\theta)\right] d_S^I - (m_S^b - \theta d_S^I) = \tau_b \delta_b (1+i_S) b_S^b + \tau_\ell \delta_\ell (1+i_S) \ell_S^b. \tag{61}$$

The LHS in equation (61) equals the additional money needed by the bank to pay out withdrawing depositors, and the RHS equals the money obtained from selling assets to the government. Nominal liabilities of the government at the beginning of the next CM just after the active entrepreneurs have repaid their loans then become

$$M_S + [\tau_b \delta_b (1+i_S) b_S^b + \tau_\ell \delta_\ell (1+i_S) \ell_S^b] - \tau_\ell \chi (1+i_S) \ell_S^b + (1-\tau_b)(1+i_S) b_S^b + (1+i_S)(B_S - b_S^b). \tag{62}$$

The first two terms in (62) are the total money in circulation after the injection of emergency liquidity, the third term is the money withdrawn from circulation in the beginning of the CM through loan repayments by entrepreneurs on the loans purchased by the government, and the last two terms are government bonds held by banks and households, respectively. Combining (62) with (57), we obtain[50]

$$\frac{1+\pi}{1+\pi_S} = \max\left\{1 + \frac{(1+i_S)(\delta_\ell - \chi)\tau_\ell \ell_S^b - (1+i_S)(1-\delta_b)\tau_b b_S^b}{M_S + (1+i_S)B_S}, 1\right\}. \tag{63}$$

It is immediate from (63) that the provision of emergency liquidity may only be inflationary if $\chi < \delta_\ell$, i.e. if loan defaults exceed the discount imposed on asset purchases. Put differently, emergency liquidity can only be inflationary if the government makes losses from its intervention. As long as there are no losses, the amount of money withdrawn from circulation when assets mature is at least as large as the amount injected in the run, such that the provision of emergency liquidity does then not increase the long-run amount of nominal government liabilities in circulation.

Our next result shows that the government never makes losses on its asset purchases, implying that the provision of emergency liquidity has no effect on inflation:

Proposition 9. *With the provision of emergency liquidity, we have (i) $\delta_\ell \leq \chi$, with strict inequality when $\delta_\ell < 1$, and (ii) $\pi = \pi_S$.*

The proof of Proposition 9 can be found in Appendix A.8. The intuition is as follows. In the steady-state, only part of entrepreneurs' DM revenue goes towards loan repayments, since entrepreneurs keep part of the revenue as compensation for their labor effort. As long as the share of entrepreneurs' revenue going to loan repayments (the capital share) remains constant, a given decrease in aggregate DM spending thus translates into a less than 1:1 decrease in loan repayments. This in turn means that a shortfall in DM spending caused by a given discount on

[50] In (63) we take it as given that the government wants to stabilise prices. In particular, the government can always prevent its intervention from being deflationary by printing money in the CM, whereas inflation cannot be prevented because of the upper bound on real government liabilities.

asset purchases translates into a smaller increase in defaults, such that the government incurs no losses from its asset purchases. The government's intervention can thus only cause losses if it is accompanied by a decrease in the capital share. As it turns out, such a decrease in the capital share is inconsistent with equilibrium. To see this, note that a decrease in the capital share needs to go together with a corresponding increase in inflation reducing entrepreneurs' real debt burden. Otherwise, entrepreneurs' real debt burden reduces by less than the revenue left after compensating them for their labour effort, implying that all entrepreneurs are better off not producing in the DM. However, for similar reasons as those described above, a given decrease in the capital share will not translate into sufficiently large losses for the government that create the amount of inflation required to reduce entrepreneurs' debt burden enough to allow for the decreasing capital share.

Interestingly, the implications of the provision of emergency liquidity on the price level are independent of how loan contracts are denoted:

Proposition 10. *$\pi = \pi_S$ also holds in case of a run if entrepreneurs' gross repayment is fixed in real terms.*

The proof for Proposition 10 is in Appendix A.9. The intuition is that with real loan contracts being bought by the government, inflation acts as a double-edged sword on the government's balance sheet: on the one hand, inflation reduces the real purchasing power of impatient depositors. Due to a fixed real repayment burden, this inevitably leads to default by entrepreneurs and hence to lower repayments to the government. On the other hand, inflation increases the nominal repayment by active entrepreneurs. Keeping the capital share α constant, these two forces cancel each other out. Thus, the effects of emergency liquidity provision are the same with nominal and real debt contracts.

Finally, we obtain from (60) and from our results in Propositions 2 and 9 that the share of active entrepreneurs in a run with provision of emergency liquidity equals

$$\chi = \frac{M^{RE}}{\theta d_S^I} = \theta + \eta(1 - \theta). \tag{64}$$

A run therefore leads to defaults if and only if $\eta = 1 - \sigma < 1$, i.e. if and only if banks charge a penalty on DM redemptions. According to (59), whether or not banks charge a redemption penalty depends on the discounts (δ_ℓ, δ_b) that the government imposes on its asset purchases. In particular, we have

$$\eta < 1 \iff m_S^b + (1 + i_S)(\delta_\ell \ell_S^b + \delta_b b_S^b) < d_S^I. \tag{65}$$

Recall that $d_S^I = d_S^P = m_S^b + (1 + i_S)(\ell_S^b + b_S^b)$ in ZLB and PLI equilibria and that $d_S^I \leq m_S^b + (1 + i_S)(\ell_S^b + b_S^b)$ in FLI equilibria, with strict inequality except for a knife-edge case. The following result then follows immediately:

Corollary 1. *With the provision of emergency liquidity, a run leads to defaults in ZLB and PLI equilibria if and only if $\delta_\ell < 1$, i.e. if and only if the government imposes a discount on its asset* *purchases. In FLI equilibria, a run leads to defaults if and only if discounts are sufficiently high and/or banks hold few illiquid assets.*

Intuitively, a discount on asset purchases implies that the bank may not be able to satisfy its steady-state promises, especially in cases where steady state DM payouts are relatively high compared to CM payouts. It thus uses the redemption penalty to pass on the costs of liquidating bonds and loans to depositors. In turn, the penalty leaves the impatient depositors who withdraw after the run has been detected with less cash compared to the steady state. This causes a drop in aggregate demand, which leads some entrepreneurs to default on their loans. If no discounts are imposed however, banks are still solvent even with steady-state promises, so they do not impose redemption penalties, and in turn all impatient depositors can obtain d_S^I in the DM even in a run. Given this and since there are no inflationary implications associated with providing emergency liquidity, the real economy is unaffected by the bank run.

8.2 Preventing Runs with Emergency Liquidity

Let us now examine whether the provision of emergency liquidity can stop a run after it has been detected, and whether the prospect of emergency liquidity provision prevents runs from starting in the first place. Without loss of generality, we suppose $\tau_b = \tau_\ell = \theta + (1-\theta)\mathcal{I}$, i.e. the fraction of bonds and loans sold to the government equals the fraction of remaining depositors that continue to run once the run is detected. The bank's CM payout to non-withdrawing depositors after a run occurred in the DM is then given by

$$\frac{m_S^b - \theta d_S^I + (1+i_S)(b_S^b + \chi \ell_S^b)}{1-\theta}.$$

Note that the discounts (δ_ℓ, δ_b) imposed on asset sales have no (direct) effect on CM payouts since the redemption penalty ensures that depositors who withdraw in the DM internalize the cost of liquidating assets. A run stops once it has been detected if and only if

$$\eta d_S^I \leq \frac{m_S^b - \theta d_S^I + (1+i_S)(b_S^b + \chi \ell_S^b)}{1-\theta}, \tag{66}$$

i.e. if and only if patient depositors are better off not withdrawing in the DM once the redemption penalty is imposed. Combining (66) with (59) yields the following result:

Proposition 11. *Runs can always be stopped with the provision of emergency liquidity, independent of the discounts imposed on asset purchases.*

Proof. Suppose first that the discounts (δ_ℓ, δ_b) are such that banks do not need to impose a redemption penalty, i.e. $\eta = 1$. By (64), this implies $\chi = 1$. Condition (66) then reduces to $d_S^I \leq m^b + (1+i_S)(b_S^b + \ell_S^b)$, which is always fulfilled since $d_S^I \leq d_S^P$, i.e. since steady state DM payouts cannot exceed CM payouts. Suppose next that the discounts (δ_ℓ, δ_b) are such that banks need to impose a redemption penalty, i.e. $\eta < 1$. Substituting for η, condition (66) then becomes $\delta_b b_S^b + \delta_\ell \ell_S^b \leq b_S^b + \chi \ell_S^b$, which is always fulfilled since $\delta_b \leq 1$ and $\delta_\ell \leq \chi$ (for the latter, see Proposition 9). ■

The reason why the provision of emergency liquidity always stops runs is closely related to our previous result that $\delta_\ell \leq \chi$, i.e. the drop in aggregate demand is always smaller than the discount on asset purchases imposed by the government. This ensures not only that the government does not make losses on its asset purchases but also that the reduction in the value of illiquid assets held by banks is always smaller than the redemption penalty, such that withdrawing late is more lucrative for patient depositors than running the bank.

Finally, the provision of emergency liquidity prevents a run if and only if

$$d_S^I \leq \frac{m_S^b - \theta d_S^I + (1+i_S)(b_S^b + \chi \ell_S^b)}{1-\theta}, \tag{67}$$

i.e. if and only if patient depositors are better off not withdrawing their entire deposit in the DM (before a redemption penalty is imposed) given that emergency liquidity will be provided in a run. Combining condition (67) with (64) yields the following result:

Proposition 12. *In ZLB and PLI equilibria, emergency liquidity prevents runs if and only if $\delta_\ell = 1$, i.e. if and only if the government imposes no discount on its asset purchases. In FLI equilibria where the aggregate real asset supply satisfies $\mathcal{A}_S \leq \bar{\mathcal{A}}$,[51] emergency liquidity prevents runs if and only if*

$$(\delta_\ell - 1)\theta q_S \kappa^* + (\delta_b - 1)\mathcal{B} \geq -\left(1 + \frac{q_S}{\theta \kappa^*}\right)(\mathcal{A}_S - (1-\theta)\rho_S q_S), \tag{68}$$

i.e. if and only if the discounts (δ_ℓ, δ_b) are small and/or the aggregate asset supply $\mathcal{A}_S$ is large.

The proof of Proposition 12 is given in Appendix A.10. Note that, by condition (37), we have $\mathcal{A}_S \geq (1-\theta)\rho_S q_S$ in FLI equilibria, with strict inequality except for a knife-edge case. The right-hand side of condition (68) is thus negative, meaning that in FLI equilibria, it is possible to impose strictly positive discounts on asset purchases without making the banking system fragile, as long as the discounts are not too high.

Intuitively, if discounts are imposed, some impatient depositors may end up with less cash compared to the steady state; this in turn negatively affects aggregate demand, which triggers default by some entrepreneurs. These anticipated losses on the loans held by banks may make it rational for patient depositors to run, unless the bank holds a sufficiently large amount of assets to absorb such losses. Due to the discounts, a run can therefore start because of self-fulfilling reasons, just as in our model without emergency liquidity. If no discounts are imposed however, aggregate demand is stabilised and default by entrepreneurs is prevented, which in turn eliminates the incentive to run for patient depositors in the first place.

9 Conclusion

In this paper, we developed a model of the macroeconomy where entrepreneurs borrow in order to produce goods that impatient households want to consume. We showed that introducing

[51] If $\mathcal{A} > \bar{\mathcal{A}}$ then not all assets in the economy are held by banks. As in Proposition 4, whether banks are fragile depends then additionally on how illiquid assets are allocated between households and banks, without fundamentally changing the result of Proposition 12; for given discounts, banks are then more (less) likely to be fragile if they hold a large share of outstanding loans (government bonds).

banks into such an economy improves outcomes by either insuring households against liquidity risk, i.e. the cost of carrying real balances they may end up not needing, or by lending to entrepreneurs at lower rates than households would, thereby increasing aggregate supply. We called the first situation a full liquidity-insurance equilibrium (FLI) and the second situation a zero lower bound equilibrium (ZLB). A combination of the two is also a possible outcome, which we call a partial liquidity-insurance equilibrium (PLI). These three equilibrium cases may coexist for a range of parameters, particularly when banks are able to invest in government bonds in addition to loans to entrepreneurs.

We showed that all three equilibria are prone to runs even if the banks know the share of impatient households in the economy and fully freeze deposits once the expected number of households has withdrawn their deposits. The reason for this is that a bank run creates a misallocation of liquidity, which in turn reduces aggregate demand in the goods market. This demand shortfall then leads to defaults on some loans by entrepreneurs, and the anticipation of these defaults rationalises the run by patient depositors, which causes the misallocation of liquidity.

The banking system may be able to prevent runs by using a combination of deposit freezes and redemption penalties, but even if banks observe runs immediately, there may be no such policy that prevents runs. If the government provides emergency liquidity by purchasing illiquid assets from banks, runs can be prevented if and only if the government is willing to purchase these assets at face value (or close to it if the economy is in an FLI equilibrium). This result can be interpreted as a violation of Bagehot (1873)'s rule: if emergency liquidity is provided without a discount (i.e. at market rates), the liquidity crisis is contained and banks remain solvent. If emergency liquidity is provided at a discount (i.e. at a penalty rate), banks become insolvent unless they pass on the discount to depositors – but doing so causes a shortfall in aggregate demand, which in turn reduces the return on bank assets; as a result, running the banks remains rational for patient depositors.

Bibliography

Andolfatto, D., Berentsen, A., and Martin, F. (2020). Money, banking and financial markets. *Review of Economic Studies*, 87(5):2049–2086.

Andolfatto, D. and Nosal, E. (2020). Shadow bank runs. *FRB Atlanta Working Paper No. 2020-14*.

Andolfatto, D., Nosal, E., and Sultanum, B. (2017). Preventing bank runs. *Theoretical Economics*, 12(3):1003–1028.

Aruoba, S. B., Waller, C., and Wright, R. (2011). Money and capital. *Journal of Monetary Economics*, 58:98–116.

Aruoba, S. B. and Wright, R. (2003). Search, money, and capital: A neoclassical dichotomy.

Journal of Money, Credit and Banking, 35(6):1085–1105. Part 2: Recent Developments in Monetary Economics.

Bagehot, W. (1873). *Lombard Street.* Henry S. King & Co.

Berentsen, A., Camera, G., and Waller, C. (2007). Money, credit and banking. *Journal of Economic Theory*, 135 (1):171–195.

Carapella, F. (2015). Banking panics and deflation in dynamic general equilibrium. *Finance and Economics Discussion Series 2015-018. Board of Governors of the Federal Reserve System.*

Cavalcanti, R. and Monteiro, P. K. (2016). Enriching information to prevent bank runs. *Economic Theory*, 62:477–494.

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

Engineer, M. (1989). Bank runs and the suspension of deposit convertibility. *Journal of Monetary Economics*, 24:443–454.

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

Ennis, H. M. and Keister, T. (2010). Banking panics and policy responses. *Journal of Monetary Economics*, 57(4):404–419.

Farhi, E. and Tirole, J. (2020). Shadow banking and the four pillars of traditional financial intermediation. *The Review of Economic Studies*, 88(6):2622–2653.

Geromichalos, A. and Herrenbrueck, L. (2021). The liquidity-augmented model of macroeconomic aggregates: A new monetarist dsge approach. *Review of Economic Dynamics.*

Gertler, M. and Kiyotaki, N. (2015). Banking, liquidity, and bank runs in an infinite horizon economy. *American Economic Review*, 105(7):2011–2043.

Goldstein, I., Kopytov, A., Shen, L., and Xiang, H. (2020). Bank heterogeneity and financial stability. *NBER Working Paper 27376.*

Gorton, G. (1988). Banking panics and business cycles. *Oxford Economic Papers*, 40(4):751–781.

Gorton, G. and Huang, L. (2006). Bank panics and the endogeneity of central banking. *Journal of Monetary Economics*, 53(7):1613–1629.

Gu, C., Monnet, C., Nosal, E., and Wright, R. (2019). On the instability of banking and financial intermediation. *University of Missouri WP1901.*

Kashyap, A. K., Tsomocos, D. P., and Vardoulakis, A. P. (2020). Optimal bank regulation in the presence of credit and run risk. *NBER Working Paper 26689.*

Keister, T. and Mitkov, Y. (2019). Bailouts, bail-ins and banking crises. *CRC TR 224 Discussion Paper No. 91.*

Kiyotaki, N. and Wright, R. (1989). On money as a medium of exchange. *Journal of Political Economy*, 97 (4):927–954.

Lagos, R. and Rocheteau, G. (2008). Money and capital as competing media of exchange. *Journal of Economic Theory*, 142:247 – 258.

Lagos, R. and Wright, R. (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113 (3):463–484.

Liu, X. (2019). Diversification and systemic bank runs. *mimeo.*

Martin, A. (2006). Liquidity provision vs. deposit insurance: preventing bank panics without moral hazard. *Economic Theory*, 28(1):197–211.

Matta, R. and Perotti, E. C. (2021). Pay, stay, or delay? how to settle a run. *mimeo.*

Peck, J. and Setayesh, A. (2022). Bank runs and the optimality of limited banking. *Review of Economic Dynamics*, forthcoming.

Reinhart, C. M. and Rogoff, K. S. (2008). Is the 2007 us sub-prime financial crisis so different? an international historical comparison. *American Economic Review: Papers & Proceedings*, 98(2):339–344.

Robatto, R. (2019). Systemic bank panics, liquidity risk, and monetary policy. *Review of Economic Dynamics*, 34:20–42.

Rochet, J. and Vives, X. (2004). Coordination failures and the lender of last resort: Was bagehot right after all? *Journal of the European Economic Association*, 2(6):1116–1147.

Rocheteau, G. and Wright, R. (2005). Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium. *Econometrica*, 73(1):175–202.

Sanches, D. (2018). Banking panics and output dynamics. *Review of Economic Dynamics*, 29:148–171.

Torre, A. D. L., Yeyati, E. L., Schmukler, S. L., Ades, A., and Kaminsky, G. (2003). Living and dying with hard pegs: The rise and fall of argentina's currency board. *Economia*, 3(2):43–107.

Wallace, N. (1990). A banking model in which partial suspension is best. *Federal Reserve Bank of St. Louis Quarterly Review*, 14(4):11–23.

Williamson, S. (2012). Liquidity, monetary policy, and the financial crisis: A new monetarist approach. *American Economic Review*, 102 (6):2570–2605.

Appendix A Additional Derivations and Proofs

A.1 Deriving the Households' Problem in the Economy without Banks

Each household chooses hours worked in the CM (l_t), consumption of the CM good (x_t^h), the portfolio of cash, loans and government bonds carried out of the CM (m_t, ℓ_t, b_t) as well as DM consumption in case of turning out impatient (q_t).

We take it as given that households do not carry more cash than what they need to finance q in case they turn out to be impatient, that is, we will take it as given that $q_t = m_t p_t$. It is a standard result in the literature that households will never choose to hold excess cash if $\iota > 0$, which we take as given for the moment.[52] Furthermore, we denote by $i_{\ell,t+1}$ the effective net nominal return on loans. Since (some) entrepreneurs may (partly of fully) default on their loan, $i_{\ell,t+1}$ may be lower than the contractual interest rate $\tilde{i}_{\ell,t+1}$. We assume that each household holds a fully diversified portfolio of loans, such that all households receive the same effective interest rate on their loan portfolio. Denoting $V_t(m, \ell, b)$ as the CM value function in period t,[53] we have:

$$V_t(m,\ell,b) = \max_{\{x_t^h, l_t, m_t, \ell_t, b_t\}} \Big\{ U(x_t^h) - l_t + \theta \left[u(q_t) + \beta V_{t+1}(0, \ell_t, b_t)\right] + (1-\theta)\beta V_{t+1}(m_t, \ell_t, b_t) \Big\}$$

$$\begin{aligned} s.t. \quad & l_t = x_t^h + \Delta_t + \phi_t\big[(m_t - m) + (\ell_t - (1+i_{\ell,t})\ell) + (b_t - (1+i_{b,t})b)\big], \\ & p_t q_t = m_t. \end{aligned}$$

The first constraint is the household's flow budget constraint and the second constraint is the liquidity constraint on DM consumption. Inserting the constraints into the objective function, we can rewrite the value function as:

$$\begin{aligned} V_t(m,\ell,b) = {} & \phi_t \left[m + (1+i_{\ell,t})\ell + (1+i_{b,t})b\right] - \Delta_t + \max_{\{x_t^h, m_t, \ell_t, b_t\}} \Big\{ U(x_t^h) - x_t^h - \phi_t(m_t + \ell_t + b_t) \\ & + \theta u\left(\frac{m_t}{p_t}\right) + \beta\big[\theta V_{t+1}(0, \ell_t, b_t) + (1-\theta) V_{t+1}(m_t, \ell_t, b_t)\big] \Big\} \end{aligned}$$

By the usual envelope result, quasi-linear preferences in the CM imply that CM value functions are linear in asset holdings:

$$\frac{\partial V_t(m,\ell,b)}{\partial m} = \phi_t, \quad \frac{\partial V_t(m,\ell,b)}{\partial \ell} = \phi_t(1+i_{\ell,t}), \quad \frac{\partial V_t(m,\ell,b)}{\partial b} = \phi_t(1+i_{b,t}).$$

Moving these one period forward and plugging them back into the above equation allows to write the CM problem as in (10), where we use q as a choice variable instead of m for simplicity.

A.2 Confirming that the Entrepreneur's Constraint to Work in the DM Is Slack

Since f is homogeneous of degree one, c is also homogeneous of degree one. Hence, we can rewrite the entrepreneur' objective function as

$$q_t^e \left[\rho_t - c\left(1, \frac{k_t}{q_t^e}\right) - \frac{1+\tilde{i}_{\ell,t+1}}{1+\iota_{t+1}} \frac{k_t}{q_t^e}\right].$$

[52] Even if $\iota = 0$, carrying zero excess cash is usually optimal from the point of view of an individual household, although not necessarily uniquely so.

[53] That is, $V_t(m, \ell, b)$ is the maximum attainable continuation utility when starting with a portfolio (m, ℓ, b) of cash, loans and bonds in the CM of period t.

Furthermore, the fact that c is homogeneous of degree one implies that the entrepreneurs' FOC for q_t^e (16) can be rewritten as

$$\rho_t = c\left(1, \frac{k_t}{q_t^e}\right) - \frac{k_t}{q_t^e} c_k \left(1, \frac{k_t}{q_t^e}\right)$$

Using this, together with the FOC for k_t (15), in the rewritten objective function implies that

$$q_t^e \left[\rho_t - c\left(1, \frac{k_t}{q_t^e}\right) - \frac{1+\tilde{i}_{\ell,t+1}}{1+\iota_{t+1}} \frac{k_t}{q_t^e}\right] = q_t^e \left[\rho_t - c\left(1, \frac{k_t}{q_t^e}\right) + \frac{k_t}{q_t^e} c_k \left(1, \frac{k_t}{q_t^e}\right)\right] = 0.$$

That means, the optimized value of the objective function equals zero, which should not come as a surprise given the fact that f and also c exhibit CRS. This immediately implies that given the optimal choice of k_t, we have that

$$\max_{q_t^e} \left\{\rho_t q_t^e - c(q_t^e, k_t) - \frac{1+\tilde{i}_{\ell,t+1}}{1+\iota_{t+1}} k_t\right\} = 0.$$

This verifies the conjecture that the constraint is slack.

A.3 The Optimality Conditions of the Bank's Problem

The Lagrangian of the bank's problem writes:

$$\begin{aligned}\mathcal{L} = &\theta u\left(\frac{d_t^I}{p_t}\right) + \beta(1-\theta)\phi_{t+1} d_{t+1}^P - \phi_t(m_t^b + a_t^b) \\ &+ \beta\phi_{t+1}\left[\zeta_t(m_t - \theta d_t^I) + \xi_t(m_t + a_t^b(1+i_{t+1}) - \theta d_t^I - (1-\theta)d_{t+1}^P) + (1-\theta)\psi_t(d_{t+1}^P - d_t^I)\right]\end{aligned}$$

The first-order conditions are

$$d_t^I: \quad \frac{u'(q_t)}{p_t} = \beta\phi_{t+1}\left(\zeta_t + \xi_t + \left(\tfrac{1}{\theta} - 1\right)\psi_t\right) \tag{69}$$

$$d_{t+1}^P: \quad \xi_t - \psi_t = 1 \tag{70}$$

$$m_t^b: \quad 1+\iota_t = \zeta_t + \xi_t \tag{71}$$

$$a_t^b: \quad \xi_t = \frac{1+\iota_t}{1+i_{t+1}} \tag{72}$$

where we have used $1+\iota_t \equiv \frac{\phi_t}{\beta\phi_{t+1}}$. From (70)-(72), we get the expressions for the Lagrange multipliers in (27). Inserting the expressions for the multipliers into (69) yields equation (28). Finally, the bank's asset demand schedule (30) follows from the complementary slackness conditions $\zeta_t(m_t^b - \theta d_t^I) = 0$ and $\psi_t(d_{t+1}^P - d_t^I) = 0$.

A.4 Proof that a Stationary Monetary Equilibrium Exists

As shown in the main text, i pins down κ, which in turn pins down ρ. Furthermore, ρ and i together pin down q. To show that an SME always exists, it remains to show that there exists always an $i \in [0, \iota]$ such that condition (37) is fulfilled. Note that $\mathcal{A}, \rho$ and q can all be expressed as continuous functions of i. From (37), we get that an equilibrium with $i \in (0, \iota)$ exists iff:

$$Q(i) = 1 + i - \frac{(1-\theta)(1+\iota)\rho(i)q(i)}{\mathcal{A}(i)} = 0. \tag{73}$$

We also get from (37) that $Q(0) > 0$ and $Q(\iota) < 0$ if neither an equilibrium with $i = 0$ nor one with $i = \iota$ exists. By the intermediate value theorem, an equilibrium with $i \in (0, \iota)$ then exists, which also proofs that an equilibrium with $i \in [0, \iota]$ always exists. As a side note, we have that $Q(0) < 0$ and $Q(\iota) > 0$ if both an equilibrium with $i = 0$ and one with $i = \iota$ exist. It then follows again from the intermediate value theorem that there also exists (at least one) equilibrium with $i \in (0, \iota)$.

A.5 Derivations of the Thresholds for the Penalty on Early Redemptions

A.5.1 Threshold to Stop a Run

Rewriting the condition in (51) gives

$$\eta d_S^I \leq (\eta+\sigma)\underline{d}_R^P(\eta,\sigma). \tag{74}$$

Since the right-hand side of (74) is strictly increasing in σ while the left-hand side does not change in σ, there is a unique threshold, denoted $\underline{\sigma}(\eta)$, such that condition (74) is fulfilled iff $\sigma \geq \underline{\sigma}(\eta)$. Substituting for $\underline{d}_R^P(\eta,\sigma)$ we can rewrite condition (74) as:

$$\begin{aligned} \eta d_S^I &\leq (\eta+\sigma)\frac{(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)}{1-\lambda-\overline{\omega}(\eta)(\eta+\sigma)} \\ \Leftrightarrow \sigma\left[(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)+\overline{\omega}(\eta)\eta d_S^I\right] &\geq (1-\lambda)\eta d_S^I-\eta\left[(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)+\overline{\omega}(\eta)\eta d_S^I\right] \\ \Leftrightarrow \sigma &\geq \left[\frac{d_S^I-m_S^b-(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)}{m_S^b-\lambda d_S^I+(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)}\right]\eta \equiv \underline{\sigma}(\eta) \end{aligned} \tag{75}$$

Notice that $\underline{\sigma}(0)=0$. Also, $\underline{\sigma}(\eta)$ is strictly increasing in η iff $d_S^I \geq (1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)+m_S^b$, which is the condition for banks to be fragile under pure deposit freezes (see (49)).

Furthermore, the fact that $\sigma \leq 1-\eta$ means there is an upper bound on η, denoted $\tilde{\eta}^{max}$, defined as the unique value of η solving $\underline{\sigma}(\eta)=1-\eta$. Note that setting $\eta \leq \tilde{\eta}^{max}$ is a necessary condition to stop a run. Solving for $\tilde{\eta}^{max}$ gives

$$\tilde{\eta}^{max} = \frac{(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)+m_S^b-\lambda d_S^I}{(1-\lambda)d_S^I} = \overline{\eta}+\frac{(1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)}{(1-\lambda)d_S^I} \in (\overline{\eta},1), \tag{76}$$

where $\tilde{\eta}^{max}<1$ follows from $d_S^I \geq (1+i_S)\left(b_S^b+\frac{m_S^b}{d_S^I}\ell_S^b\right)+m_S^b$. It follows that a run is stopped whenever (η,σ) satisfies $\eta \in [\overline{\eta},\tilde{\eta}^{\max}]$ as well as $\sigma \in [\underline{\sigma}(\eta),1-\eta]$. Clearly, the set of (η,σ) that stop a run has positive mass.

A.5.2 Threshold to Satisfy Impatient Depositors' Incentive Constraint

Note first that we can reformulate condition (54) as

$$u(\eta q_S) \geq \frac{\phi}{1+\iota}(\eta+\sigma)\overline{d}_R^P(\eta,\sigma). \tag{77}$$

Since the right-hand side of (77) is strictly increasing in σ while the left-hand side does not change in σ, there exists a unique threshold, denoted $\overline{\sigma}(\eta)$, such that condition (77) is fulfilled iff $\sigma \leq \overline{\sigma}(\eta)$. It will be useful to define:

$$T(\eta) \equiv \frac{m_S^b+(1+i_S)(b_S^b+(\lambda+(1-\lambda)\eta)\ell_S^b)-(\lambda+\theta(1-\lambda)\eta)d_S^I}{1-\lambda}. \tag{78}$$

In words, $T(\eta)$ denotes banks' per capita CM revenue in a run, given that the run is stopped and given that the impatient depositors who were not among the first λ to arrive withdraw fraction η of their deposit; 'per capita' means that the total revenue is divided by the measure of depositors who did not manage to redeem before the bank imposed the redemption penalty. Notice that we have $T'(\eta)=(1+i_S)\ell_S^b-\theta d_S^I<0$.[54]

[54] To see why the derivative is negative, recall that θd_S^I equals the total steady state DM revenue of entrepreneurs while $(1+i_S)\ell^b$ equals their total loan repayment in steady state. Since entrepreneurs keep part of their revenue as compensation for their labor effort, we have $\theta d_S^I > (1+i_S)\ell_S^b$.

Substituting for $\overline{d}_R^P(\eta,\sigma)$, we can rewrite condition (77) as:

$$\begin{aligned} u(\eta q_S) &\geq \frac{(\eta+\sigma)}{(1-\theta(\eta+\sigma))}\frac{\phi}{1+\iota}T(\eta) \\ \Leftrightarrow \sigma\left[\theta u(\eta q_S)+\frac{\phi}{1+\iota}T(\eta)\right] &\leq (1-\theta\eta)u(\eta q_S)-\frac{\phi}{1+\iota}\eta T(\eta) \\ \Leftrightarrow \sigma &\leq \frac{u(\eta q_S)}{\theta u(\eta q_S)+\frac{\phi}{1+\iota}T(\eta)}-\eta\equiv\overline{\sigma}(\eta) \end{aligned} \tag{79}$$

Note that we have $\overline{\sigma}(0)=0$, and

$$\overline{\sigma}'(\eta)=\frac{\frac{\phi}{1+\iota}\big(q_S T(\eta)u'(\eta q_S)-u(\eta q_S)T'(\eta)\big)}{\left(\theta u(\eta q_S)+\frac{\phi}{1+\iota}T(\eta)\right)^2}-1, \tag{80}$$

with $\lim_{\eta\to 0}\overline{\sigma}'(\eta)=\frac{q_S u'(0)}{\frac{\phi}{1+\iota}T(0)}-1=+\infty$, which results from the fact that $u(q)$ satisfies the Inada conditions. Notice that $\overline{\sigma}'(\eta)$ need not be positive everywhere on $\eta\in(0,1)$. The reason is that the return to banks' asset portfolio increases in η due to lower loan defaults. This makes it more costly to give up a given share of the deposit. If this effect is strong enough, the haircut which impatient depositors are willing to pay may decrease in η.

Next, we have $\overline{\sigma}''(\eta)=\frac{f'g-fg'}{g^2}$, where f and g are defined by $\overline{\sigma}'(\eta)\equiv\frac{f}{g}-1$, and f' and g' denote the derivatives of f and g w.r.t. η respectively. We have

$$f'=\overbrace{\frac{\phi}{1+\iota}q_S^2T(\eta)}^{>0}\overbrace{u''(\eta q_S)}^{<0}<0 \tag{81}$$

$$g'=2\underbrace{\left[\theta u(\eta q_S)+\frac{\phi}{1+\iota}T(\eta)\right]}_{>0}\underbrace{\left[\theta q_S u'(\eta q_S)+\frac{\phi}{1+\iota}T'(\eta)\right]}_{>0\text{ for }\eta\in[0,1]}>0. \tag{82}$$

Since $f>0$ and $g>0$, it follows that $\overline{\sigma}''(\eta)<0$, i.e. $\overline{\sigma}(\eta)$ is strictly concave on $\eta\in[0,1]$. Regarding expression (82), note that

$$\begin{aligned} \theta q_S u'(\eta q_S)+\frac{\phi}{1+\iota}T'(\eta)&>0 \\ \Leftrightarrow \theta q_S u'(\eta q_S)&>\frac{\phi}{1+\iota}(\theta d_S^I-(1+i_S)\ell_S^b) \\ \Leftrightarrow \frac{u'(\eta q_S)}{\rho_S}&>1-\frac{\phi}{1+\iota}\frac{(1+i_S)\ell_S^b}{\theta\rho_S q_S} \end{aligned} \tag{83}$$

where we used $T'(\eta)=-(\theta d_S^I-(1+i_S)\ell_S^b)$, $d_S^I=p_Sq_S$ and $\rho_S\equiv\frac{\phi}{1+\iota}p_S$. Since $\frac{u'(q_S)}{\rho_S}>1$ in steady state (see (33) and (35)), we know that condition (83) is fulfilled for any $\eta\in[0,1]$.

A.5.3 Threshold to Prevent a Run

Substituting for $\overline{d}_R^P(\eta,\sigma)$ we can rewrite the condition in (55) as:

$$\begin{aligned} (1-\lambda)(1-\theta(\eta+\sigma))d_S^I &\leq m_S^b+(1+i_S)(b_S^b+(\lambda+(1-\lambda)\eta)\ell_S^b)-(\lambda+\theta(1-\lambda)\eta)d_S^I) \\ \Leftrightarrow \sigma\theta(1-\lambda)d_S^I &\geq (\lambda+(1-\lambda))d_S^I-\left[m_S^b+(1+i_S)b_S^b+(1+i_S)(\lambda+(1-\lambda)\eta)\ell_S^b\right] \\ \Leftrightarrow \sigma &\geq \left[\frac{d_S^I-(m_S^b+(1+i_S)(b_S^b+\lambda\ell_S^b))}{(1-\lambda)\theta d_S^I}\right]-\frac{(1+i_S)\ell_S^b}{\theta d_S^I}\eta\equiv\underline{\hat{\sigma}}(\eta) \end{aligned} \tag{84}$$

Note that we have $\underline{\hat{\sigma}}(0) \in (0,1)$, and $\underline{\hat{\sigma}}'(\eta) = -\frac{(1+i_S)\ell_S^b}{\theta d_S^I} < 0$, with $\left|\underline{\hat{\sigma}}'(\eta)\right| < 1$.[55]

A.6 Proof of Proposition 8

Given that banks set (η, σ) such as to stop the run, DM activity in a run is maximized by setting η as high as possible subject to the relevant constraints. Since we start from the premise that runs cannot be prevented, we can disregard constraint (55). The relevant constraints - besides banks' liquidity constraint - are thus (51) and (54). We are therefore looking for values (η, σ) that maximize η subject to the constraint that $\underline{\sigma}(\eta) \leq \sigma \leq \overline{\sigma}(\eta)$ with $\sigma \in [0, 1-\eta]$, and subject to the liquidity constraint $\eta \leq \overline{\overline{\eta}}$.

Recall from Appendices A.5.1 and A.5.2 that $\underline{\sigma}(0) = \overline{\sigma}(0) = 0$, where $\underline{\sigma}(\eta)$ is linearly increasing in η, while $\overline{\sigma}(\eta)$ is strictly concave on $\eta \in [0,1]$ with $\lim_{\eta \to 0} \overline{\sigma}'(\eta) = +\infty$. This implies that whenever there exist values $\eta \in [0,1]$ for which $\overline{\sigma}(\eta) < \underline{\sigma}(\eta)$, then there exists a unique strictly positive value $\hat{\eta}^{max}$ such that $\overline{\sigma}(\eta) \geq \underline{\sigma}(\eta)$ iff $\eta \leq \hat{\eta}^{max}$. If $\overline{\sigma}(\eta) \geq \underline{\sigma}(\eta)$ for all $\eta \in [0,1]$, we define $\hat{\eta}^{max} = 1$. Next, we will show that $\hat{\eta}^{max} > \overline{\eta}$. We have

$$
\begin{aligned}
& \overline{\sigma}(\overline{\eta}) > \underline{\sigma}(\overline{\eta}) \\
\Leftrightarrow \quad \frac{1}{\theta + \frac{\phi}{1+\iota}\frac{T(\overline{\eta})}{u(\overline{\eta}q_S)}} > & \frac{\overline{\eta}(1-\lambda)d_S^I}{(1+i_S)\left(b_S^b + \frac{m_S^b}{d_S^I}\ell_S^b\right) + m_S^b - \lambda d_S^I} \\
\Leftrightarrow \quad & T(\overline{\eta}) > \frac{\phi}{1+\iota}\frac{\overline{\eta} d_S^I T(\overline{\eta})}{u(\overline{\eta}q_S)} \\
& \frac{u(\overline{\eta}q_S)}{\overline{\eta}q_S} > \rho_S
\end{aligned}
\tag{85}
$$

where we used $d_S^I = p_S q_S \equiv \frac{1+\iota}{\phi}\rho_S q_S$ in the last step. Since $u(q)$ is strictly concave with $u(0) = 0$, we have $\frac{u(q)}{q} > u'(q)$. Since $u'(q_S) > \rho_S$ (see (33) and (35)), condition (85) is fulfilled, from which it follows that $\hat{\eta}^{max} > \overline{\eta}$.

Finally, from Appendix A.5.1, we know that the highest value of η consistent with $\underline{\sigma}(\eta) \leq 1 - \eta$, i.e. stopping a run, equals $\tilde{\eta}^{\max}$, with $\tilde{\eta}^{\max} \in (\overline{\eta}, 1)$. It then follows that constraints (51) and (54) can only be jointly satisfied if $\eta \leq \min\{\tilde{\eta}^{max}, \hat{\eta}^{max}\} \in (\overline{\eta}, 1)$. The fact that banks' liquidity constraint requires additionally that $\eta \leq \overline{\overline{\eta}}$ then leads to the result in Proposition 8.

A.7 Example with $\lambda = 0$ where Runs cannot be Prevented

Figure 2 shows an example where there exists no (η, σ) that prevents runs, despite the fact that banks can react to runs immediately ($\lambda = 0$). The functional forms used here are the same as for the example shown in Figure 1, and the parameter values are given by Table 2. As in the example of Figure 1, parameters imply that the economy is in a zero lower bound equilibrium. We can see graphically that the

Table 2: Parameter values for Figure 2.

α	ν	θ	n	λ	ι	$\mathcal{B}$	q_S	ρ_S	κ_S
0.65	0.95	0.6	0.6	0	0.02	0	$5.725 * 10^{-7}$	1.886	1.251

set of values (η, σ) satisfying constraints (51), (54) and (55) is empty. Notice that the banks' liquidity

[55] See Footnote 54 for why the absolute value of the derivative is less than one.

constraint is fulfilled for any $\eta \in [0,1]$ since banks can react to runs without delay. What is key is that impatient depositors' willingness to pay a redemption penalty is low, which results from the fact that the DM utility function is close to linear. In particular, the redemption penalty cannot be set high enough to deter patient depositors from running, since impatient depositors would not be willing to incur such a penalty. Nevertheless, banks can stop runs by partially freezing deposits and charging a modest penalty on redemptions once a run has started (grey area). DM activity in a run is maximized by setting $(\eta, \sigma) = (\hat{\eta}^{max}, \underline{\sigma}(\hat{\eta}^{max}))$, where $\hat{\eta}^{max}$ corresponds to the highest DM payout consistent with stopping the run.

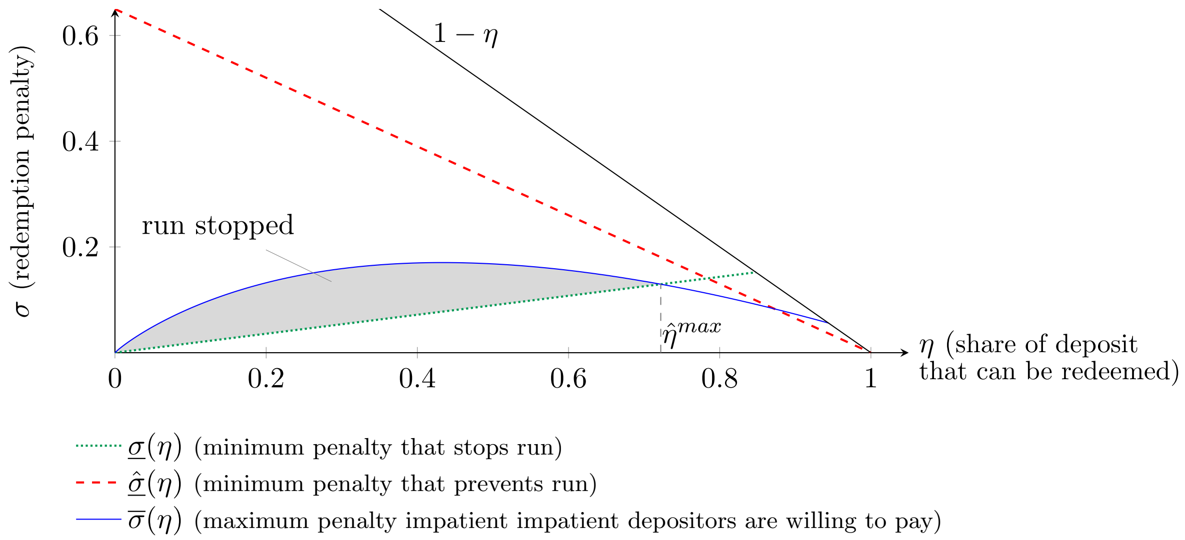

Figure 2: Example where runs cannot be prevented.

A.8 Proof of Proposition 9

From (63), we know that $\pi > \pi_S$ can only occur if $\chi < \delta_\ell$. From (40) and (60), we get that

$$\chi < \delta_\ell \;\Leftrightarrow\; \frac{\alpha(\rho_S)}{\alpha(\rho)} > \frac{\theta + \eta(1-\theta)}{\delta_\ell}. \tag{86}$$

Substituting for η using (59), we get:

$$\frac{\theta + \eta(1-\theta)}{\delta_\ell} = \begin{cases} \dfrac{1}{\delta_\ell} & \text{if } \eta = 1 \\ \dfrac{m_S^b + (1+i_S)(\delta_\ell \ell_S^b + \delta_b b_S^b)}{\delta_\ell d_S^I} & \text{if } \eta < 1 \end{cases}$$

Using the fact that $\delta_b \geq \delta_\ell$ and $d_S^I \leq (1+i_S)(\ell_S^b + b_S^b)$ (which follows from patient depositors' IC constraint, $d_S^I \leq d_S^P$), it is easy to see that

$$\frac{\theta + \eta(1-\theta)}{\delta_\ell} \geq 1, \text{ with strict inequality if } \delta_\ell < 1.$$

Condition (86) thus says that $\chi < \delta_\ell$ must go together with a sufficiently strong decrease in the capital share $\alpha(\rho)$ relative to the steady state. Next, from (43), we get

$$\pi > \pi_S \;\Rightarrow\; \rho\alpha(\rho)q^e(\rho) < \rho_S\alpha(\rho_S)q^e(\rho_S), \tag{87}$$

i.e. an increase in inflation implies a decrease in real loan repayments by active entrepreneurs. From (41), we have $\rho\alpha(\rho)q^e(\rho) = \max_{q^e}\{\rho q^e - c(q^e, k_S)\}$, and hence $\partial[\rho\alpha(\rho)q^e(\rho)]/\partial\rho = q^e > 0$, i.e. active entrepreneurs' real loan repayment increases in the real price of the DM good ρ. Together with (87), this implies

$$\pi > \pi_S \;\Rightarrow\; \rho < \rho_S, \tag{88}$$

i.e. an increase in inflation implies a decrease in the real price of the DM good. Note next that, from (43), we have

$$\frac{\alpha(\rho_S)}{\alpha(\rho)} = \frac{\rho q^e(\rho)}{\rho_S q^e(\rho_S)} \frac{1+\pi}{1+\pi_S}. \tag{89}$$

From (89) and the fact that $q^e(\rho)$ is strictly increasing in ρ (see the proof of Proposition 3), we get that $\pi > \pi_S$ implies

$$\frac{\alpha(\rho_S)}{\alpha(\rho)} < \frac{1+\pi}{1+\pi_S} = 1 + \frac{(1+i_S)\left(\delta_\ell - \frac{\alpha(\rho)}{\alpha(\rho_S)}(\theta+\eta(1-\theta))\right)\tau_\ell \ell_S^b - (1+i_S)(1-\delta_b)\tau_b b_S^b}{M_S + (1+i_S)B_S}, \tag{90}$$

i.e. a decline in the capital share relative to the steady state needs to be accompanied by a corresponding increase in inflation that reduces entrepreneurs' real debt burden. Note that we have used (40), (60) and (63) to substitute for $(1+\pi)/(1+\pi_S)$ in (90). The last step of the proof is to show that (86) and (90) are inconsistent with each other. To do this, we first take the derivative of the RHS in (90) with respect to $\alpha(\rho_S)/\alpha(\rho)$, which gives

$$\frac{\partial[(1+\pi)/(1+\pi_S)]}{\partial[\alpha(\rho_S)/\alpha(\rho)]} = \frac{(1+i_S)\tau_\ell \ell_S^b}{M_S + (1+i_S)B_S}\left(\frac{\alpha(\rho)}{\alpha(\rho_S)}\right)^2 (\theta+\eta(1-\theta)) < 1, \tag{91}$$

where we use the fact that aggregate nominal loan repayments in steady state cannot exceed the steady state money stock, i.e. $(1+i_S)\ell_S^b < M_S$. Expression (91) shows that inflation (relative to the steady state) reacts less than 1:1 to changes in the capital share. Finally, it is easy to see that (90) is violated for $\alpha(\rho_S)/\alpha(\rho) = [\theta+\eta(1-\theta)]/\delta_\ell$, i.e. if (86) holds at equality. Together with (91), this means there exists no $\alpha(\rho_S)/\alpha(\rho)$ that satisfies both (86) and (91). This completes our proof that $\chi \geq \delta_\ell$ and hence $\pi = \pi_S$. Furthermore, if $\delta_\ell < 1$ and we replace the strict inequality in (86) with a weak inequality, we still get a contradiction, such that $\delta_\ell < 1$ implies $\chi > \delta_\ell$.

A.9 Proof of Proposition 10

When loans are real, the actual gross nominal repayment on a loan ℓ becomes $(1+i_S)\ell(1+\pi)/(1+\pi_S)$. We assume that the government remains to value loans at their steady state nominal price, so that equations (58) and (59) remain unchanged. Equation (39) remains valid and Equation (42) becomes $\max_{q^e}\{\rho q^e - c(q^e, k_S)\} \geq \beta \frac{1+i_S}{1+\pi_S} k_S$. In turn, Equation (43) becomes

$$\frac{\rho\alpha(\rho)q^e(\rho)}{\rho_S\alpha(\rho_S)q^e(\rho_S)} \geq 1, \text{ with equality if } \chi < 1. \tag{92}$$

Combining this with (39) and (60) yields that the fraction of active entrepreneurs satisfies:

$$\chi = \min\left\{\frac{[\theta+\eta(1-\theta)]\alpha(\rho)}{\alpha(\rho_S)}\frac{1+\pi_S}{1+\pi}, 1\right\} \tag{93}$$

Equation (61) remains unchanged, whereas equation (62) (nominal liabilities of the government) becomes

$$M_S + [\tau_b\delta_b(1+i_S)b_S^b + \tau_\ell\delta_\ell(1+i_S)\ell_S^b] - \tau_\ell\chi(1+i_S)\ell_S^b\frac{1+\pi}{1+\pi_S} \\ + (1-\tau_b)(1+i_S)b_S^b + (1+i_S)(B_S - b_S^b), \tag{94}$$

so that equation (63) changes to

$$\frac{1+\pi}{1+\pi_S} = \max\left\{1 + \frac{(1+i_S)\left(\delta_\ell - \chi\frac{1+\pi}{1+\pi_S}\right)\tau_\ell \ell_S^b - (1+i_S)(1-\delta_b)\tau_b b_S^b}{M_S + (1+i_S)B_S}, 1\right\}. \quad (95)$$

We now show that $\pi = \pi_S$ by means of a contradiction. Clearly $\pi \neq \pi_S \Leftrightarrow \pi > \pi_S$. In turn, it is immediate from (95) that $\pi > \pi_S$ only if $\chi < 1$. From (92), we get that $\chi < 1$ only if $\rho = \rho_S$. From (93), we thus get that $\chi < 1$ implies $\chi = [\theta + \eta(1-\theta)]\frac{1+\pi_S}{1+\pi}$. Using this in condition (95), we get that $\pi > \pi_S$ only if $\delta_\ell > \theta + \eta(1-\theta)$. We have shown already in the proof of Proposition 9 that $\delta_\ell \leq \theta + \eta(1-\theta)$, so that we obtain a contradiction.

A.10 Proof of Proposition 12

Note that condition (67) can be rewritten as

$$d_S^I \leq m_S^b + (1+i_S)(b_S^b + \chi \ell_S^b). \quad (96)$$

Consider first ZLB and PLI equilibria. Since the IC constraint of patient depositors binds in these equilibria, we have $d_S^I = d_S^P = m_S^b + (1+i_S)(\ell_S^b + b_S^b)$. Condition (96) is thus fulfilled if and only if $\chi = 1$ which, by Corollary 1, is the case if and only if $\delta_\ell = 1$.

Next, consider FLI equilibria. Assuming $\eta < 1$, we get from (59) and (64) that

$$\chi = \theta + \eta(1-\theta) = \frac{m_S^b + (1+i_S)(\delta_b b_S^b + \delta_\ell \ell_S^b)}{d_S^I}. \quad (97)$$

Inserting this into condition (96), and using the fact that $m_S^b = \theta d_S^I$ in FLI equilibria, yields that runs are prevented in an FLI equilibrium iff

$$(1-\theta)d_S^I \leq (1+\iota)b_S^b + (1+\iota)\left(\theta + \frac{(1+\iota)(\delta_b b_S^b + \delta_\ell \ell_S^b)}{d_S^I}\right)\ell_S^b. \quad (98)$$

Substituting $\phi d_S^I = \phi p_S q_S \equiv (1+\iota)\rho_S q_S$, $\phi b_S^b = \mathcal{B}$ and $\phi \ell_S^b = \theta q_S \kappa_S$ in condition (98) (the latter two equations follow from the fact that banks hold all assets in the economy if $\mathcal{A}_S \leq \overline{\mathcal{A}}$) gives

$$(1-\theta)\rho_S q_S \leq \mathcal{B} + \theta^2 q_S \kappa_S + \frac{\theta \kappa_S}{\rho_S}(\delta_b \mathcal{B}_S + \delta_\ell \theta q_S \kappa_S). \quad (99)$$

Substituting $\mathcal{A}_S = \mathcal{B} + \theta q_S \kappa_S$ and using the fact that $\kappa_S = \kappa^*$ in FLI equilibria, condition (99) can then be rewritten as (68). So far we have taken it as given that $\eta < 1$, i.e. the discounts are high enough such that banks impose a redemption penalty in the DM. Inserting $m_S^b = \theta d_S^I$ into condition (65) and substituting for d_S^I, ℓ_S^b and b_S^b in the same manner as above yields that in an FLI equilibrium

$$\eta < 1 \;\Leftrightarrow\; (\delta_b - 1)\mathcal{B} + (\delta_\ell - 1)\theta q_S \kappa^* < -[\mathcal{A}_S - (1-\theta)\rho_S q_S]. \quad (100)$$

We can see immediately that $\eta < 1$ whenever condition (68) is violated, which means that (68) is both necessary and sufficient for emergency liquidity to prevent runs.

A.11 Reinterpreting the Government's Intervention as Secured Lending

Suppose that, instead of purchasing assets outright, the government stands ready to provide emergency credit at a gross interest rate $1/\delta_b$ when the loan is secured by government bonds and at gross interest rate $1/\delta_\ell$ when it is secured by loans, where we continue assuming $\delta_b, \delta_\ell \leq 1$. The amount of collateral banks need to post equals the face value of the loan, i.e. in order to obtain one unit of money in the

DM, banks need to post either government bonds with face value $1/\delta_b$ or loans with face value $1/\delta_\ell$ (or any combination thereof). Analogous to section 8, τ_b and τ_ℓ denote the fraction of bonds and loans that banks pledge as collateral. Emergency loans extended to banks are due in the CM next period, after entrepreneurs repay their loans; if banks fail to repay, the government seizes the collateral.

As before, we assume that the government cannot increase its real indebtedness, i.e. constraint (57) applies. In said constraint, $M + (1 + i_S)B$ now denotes the government's nominal liabilities at the beginning of the next CM, just after entrepreneurs have repaid loans and banks have repaid their emergency credit. The provision of emergency liquidity will then be inflationary whenever it leads to an increase in nominal government liabilities beyond the point at which the assets pledged as collateral mature.

It is easy to see that the total amount of liquidity a bank can access by pledging all its assets as collateral is still given by (58). This implies that the redemption penalty banks need to charge in order to stop a run is still given by (59), and the aggregate cash held by impatient depositors in a run equals (60). Equation (61) remains the same as well, with the RHS now denoting the total money a bank raises by obtaining a secured credit from the government. Finally, nominal liabilities of the government at the beginning of the next CM will never exceed (62). To see this, note that if banks repay their emergency loans, all money injected in the DM will be removed from circulation. If banks default on their loans, the government seizes the collateral, in which case its liabilities will be identical to the case where it purchases the assets outright, i.e. they will be exactly equal to (62). As a result, inflation created by the government's intervention is still given by (63), from which it follows that all further results derived in Section 8 remain the same. Note in particular that a bank's CM payouts - and hence incentives to run for patient depositors - do not depend on whether the bank sells a given amount of assets in the DM or pledges them as collateral.